OXFORD

ARCHITECTURES OF GLOBAL AI GOVERNANCE

FROM **TECHNOLOGICAL CHANGE** TO **HUMAN CHOICE**

MATTHIJS M. MAAS

Architectures of Global AI Governance

Architectures of Global AI Governance

From Technological Change to Human Choice

MATTHIJS M. MAAS

Senior Research Fellow, Institute for Law & AI, Cambridge, MA, USA
Associate Fellow, Leverhulme Centre for the Future of Intelligence,
University of Cambridge, Cambridge, UK

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

To Wilfried Maas (1957–2024)
and to Kai (2022–)
For teaching me the value of choice
when the world changes

Acknowledgements

How do we govern a changing technology, in a changing world, using governance instruments that may themselves be left changed?

One good answer is: with each other.

In writing this book I have had the fortune to spend years exploring an issue I both deeply care about and am unendingly fascinated by. Throughout this process, I have been especially lucky to work with and learn from a large and supportive community. I am deeply grateful to so many people for their support, feedback—and patience.

Stretching over many years, my thinking on AI and its governance has benefited tremendously from work by, and conversations and exchanges with, many people, including Cecil Abungu, Ashwin Acharya, Sumaya Nur Adan, Michael Aird, Marco Almada, Markus Anderljung, Bill Anderson-Samways, Jacob Arbeid, Renan Araujo, Shahar Avin, Joslyn Barnhart, Seth Baum, Haydn Belfield, Lyria Bennett Moses, John Bliss, Emma Bluemke, Ingvild Bode, Jan Brauner, Roger Brownsword, Miles Brundage, Justin Bullock, Duncan Cass-Beggs, Alan Chan, Peter Cihon, Di Cooke, Carla Zoe Cremer, Rebecca Crootof, Allan Dafoe, John Danaher, Ashley Deeks, Jeffrey Ding, Noemi Dreksler, Connor Dunlap, Sophie-Charlotte Fischer, Carrick Flynn, John Fogle, Kevin Frazier, Gideon Futerman, Iason Gabriel, Denise Garcia, Benjamin Garfinkel, Katja Grace, Ross Gruetzemacher, Oliver Guest, Carlos Ignacio Gutierrez, Rose Hadshar, Hiski Haukkala, Keith John Hayward, Seán Ó hÉigeartaigh, Lennart Heim, Jeffrey Heninger, Lewis Ho, Jeroen Hopster, Shin-Shin Hua, Cullen O'Keefe, Luke Kemp, Kyle Kilian, Leonie Koessler, Noam Kolt, Victoria Krakovna, David Krueger, Karin Kuhlemann, Martina Kunz, Kristian Lauta, Harry Law, Gavin Leech, Jade Leung, John-Clark Levin, Gregory Lewis, Lucy Lim, Hin-Yan Liu, Garrison Lovely, Suvradip Maitra, Kayla Matteucci, Andrew Mazibrada, Zvi Mowshowitz, Rune Nyrup, Tobi Olasunkanmi, Michael Page, Kenneth Payne, Carina Prunkl, Neil Renic, Regina Rini, Léonard van Rompaey, Anders Sandberg, Lewin Schmitt, Jonas Schuett, Nathan Sears, Elizabeth Seger, Toby Shevlane, Konrad Seifert, Rumtin Sepasspour, Henry Shevlin, Morgan Simpson, Jacob Livingston Slosser, Maxime Stauffer, Merlin Stein, Charlotte Stix, Andrew Strait, Thomas Streinz, Janna Tay, Sue Anne Teo, Lena Trabucco, Robert Trager, Maaike Verbruggen, José Jaime Villalobos, Karina Vold, Gabriel Weil, Christoph Winter, Misha Yagudin, Baobao Zhang, and many, many others.

For sparking and nourishing my interest in these topics, and for supporting my career in this field from an early stage, I am especially grateful to Seán Ó

Contents

PART II. FACETS OF CHANGE

Tables

Abbreviations

AAAI	Association for the Advancement of Artificial Intelligence
AAJI	Advanced Artificiaţl Judicial Intelligence
ABM	Anti-Ballistic Missile (Treaty)
ADM	Automated Decision-Making (systems)
AGI	Artificial General Intelligence
AI	Artificial Intelligence
AIA	Artificial Intelligence Act *(EU Regulation)*
AIAB	AI Advisory Body (also 'High-Level Advisory Body on AI', HLAB)
AIDA	AI and Data Act *(Canadian regulation)*
AIDD	AI Incident Database
AISI	AI Safety Institute
ASEAN	Association of South East Asian Nations
ATT	Arms Trade Treaty
AUKUS	*[Trilateral security partnership between Australia, UK, US]*
AWS	Autonomous Weapons Systems
BMEWS	Ballistic Missile Early Warning System
CAAI	China Administration of AI *(proposed)*
CAC	Cyberspace Administration of China
CAHAI	Council of Europe Ad Hoc Committee on Artificial Intelligence
CAIS	Comprehensive AI Services / Centre on AI Safety
CASS	Chinese Academy of Social Sciences
CBM	Confidence-Building Measures
CBRN	Chemical, Biological, Radiological and Nuclear
CCGAI	Coordinating Committee for the Governance of AI *(proposed)*
CCTV	Closed-Circuit Television
CCW	Convention on Certain Conventional Weapons
CEB	(UN) Chief Executives Board for Coordination
CERN	European Organization for Nuclear Research
CIL	Customary International Law
CoC	Code of Conduct
CoE	Council of Europe
CS	Computer Science
CSTO	Collective Security Treaty Organization
CTBT	Comprehensive Test Ban Treaty
DARPA	Defense Advanced Research Projects Agency
DART	Dynamic Analysis and Replanning Tool
DETF	Digital Economy Task Force
DUFM	Dual-Use Foundation Models
EOD	Earth Observation Data

EU	European Union
FATF	Financial Action Task Force
FGCS	Fifth Generation Computer Systems
FIGO	Formal Intergovernmental Organisation
FLOPS	Floating Point Operations Per Second
FMF	Frontier Model Forum
FOBS	Fractional Orbital Bombardment System
FSB	Financial Stability Board
G7	Group of Seven
G20	Group of Twenty
GATS	General Agreement on Tariffs in Services
GATT	General Agreement on Tariffs and Trade
GAN	Generative Adversarial Network
GDC	Global Digital Compact
GDPR	General Data Protection Regulation
GGE	Group of Governmental Experts
GIS	Geographic Information Systems
GOHAI	Global Organization for High-Risk AI *(proposed)*
GPAI	Global Partnership on AI
GPAIS	General-Purpose AI System
GPT	General-Purpose Technology/Generative Pretrained Transformer
GPT-*[n]*	Generative Pretrained Transformer- *[n]* (series of OpenAI LLMs)
GPU	Graphics Processing Units
HBP	Human Brain Project
HLAB	High-Level Advisory Body on AI (also 'AI Advisory Body', AIAB)
HLAI	Human-Level AI
HLEG	High-Level Expert Group on Artificial Intelligence
HLMI	High-Level Machine Intelligence
IAEA	International Atomic Energy Agency
IAIA	International Autonomous Incidents Agreement *(proposed)*
IAIO	International Artificial Intelligence Organisation *(proposed)*
ICAO	International Civil Aviation Organization
ICBM	Intercontinental Ballistic Missile
ICC	International Criminal Court
ICJ	International Court of Justice
ICRAC	International Committee for Robot Arms Control
ICTY	International Criminal Tribunal for the Former Yugoslavia
IDI	*Institut de Droit International*
IEC	International Electrotechnical Commission
IEEE	Institute for Electrical and Electronics Engineers
IGF	Internet Governance Forum
IHL	International Humanitarian Law
IHRL	International Human Rights Law
IIGO	Informal Intergovernmental Organization
IL & IR	International Law & International Relations

ILC	International Law Commission
ILO	International Labour Organization
IMO	International Maritime Organization
IMS	International Monitoring System
INF	Intermediate-Range Nuclear Forces (Treaty)
IoT	Internet of Things
IPAI	Intergovernmental Panel on AI (*proposed*)
IPAIS	International Panel on AI Safety (*proposed*)
IPCC	Intergovernmental Panel on Climate Change
ISO	International Organization for Standardization
ISR	Intelligence, Surveillance, and Reconnaissance
ISTAR	Intelligence, Surveillance, Target Acquisition, Reconnaissance
IT	Information Technology
ITER	International Thermonuclear Experimental Reactor
ITS	Intelligent Treaty Systems
ITU	International Telecommunications Union
LAWS	Lethal Autonomous Weapons Systems
LLM	Large Language Model
LTPA	Long-Term Planning Agent
MAGIC	Multinational AGI Consortium (*proposed*)
MAIM	Mutual Assured AI Malfunction (*proposed strategy*)
MAIRI	Multilateral Artificial Intelligence Research Institute (*proposed*)
MAV	Maritime Autonomous Vehicles
MHC	Meaningful Human Control
ML	Machine Learning
MLLM	Multi-Modal Language Model
MTCR	Missile Technology Control Regime
MTOPS	Million Theoretical Operations Per Second
NAT	Normal Accident Theory
NGAD	Next-Generation Air Dominance
NGO	Non-Governmental Organization
NIST	National Institute of Standards and Technology
NLP	Natural Language Processing
NPT	Nuclear Non-Proliferation Treaty
NSG	Nuclear Suppliers Group
OECD	Organisation for Economic Co-operation and Development
OHCHR	Office of the High Commissioner of Human Rights
OoD	Out-of-Distribution [*inputs to neural networks*]
OSET	Office of the Secretary-General's Envoy on Technology
OST	Outer Space Treaty
PAI	Partnership on AI
PGIAI	Partnership for Global Inclusivity on AI
PLA	People's Liberation Army
REAIM	Responsible and Ethical AI in the Military Domain
RFM	Robot Foundation Model

RL	Reinforcement Learning
RLAIF	Reinforcement Learning from AI Feedback
RLHF	Reinforcement Learning from Human Feedback
SALT	Strategic Arms Limitation Talks
SCI	Strategic Computing Initiative
STS	Science and Technology Studies
TAI	Transformative AI
TBT	Agreement on Technical Barriers to Trade
TPU	Tensor Processing Units
UAV	Unmanned Aerial Vehicle
UDHR	Universal Declaration of Human Rights
UN	United Nations
UNCLOS	United Nations Convention on the Law of the Sea
UNGA	United Nations General Assembly
UNEP	United Nations Environment Programme
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNICRI	UN Interregional Crime and Justice Research Institute
UNIDIR	United Nations Institute for Disarmament Research
UNODC	United Nations Office on Drugs and Crime
UNU	United Nations University
VCLT	Vienna Convention on the Law of Treaties
VLM	Vision Language Model
WAMI	Wide-Area Motion Imagery
WHO	World Health Organization
WSIS	World Summit on Information Society
WTO	World Trade Organization

Introduction: AI and Change

Prologue

New technologies emerge in many forms. Many arrive on the scene with a flourish of initial public attention before they settle into the background. Some are normalized and become a regular part of the everyday world for us or for our children. Other technologies fail to live up to their initial promise or potential.

Few technologies fundamentally change the human condition in a way that is both broad and deep. Artificial intelligence (AI) will likely be one such transformative technology. In the past decade AI technology has made remarkable progress; within the next, it will come into its own as a major force of global change, for good or ill.

AI has been described as a 'coming wave,'[1] with its transformative potential as a generally enabling technology compared by some to fire, electricity, the internal combustion engine, nuclear power (or weapons), or computing.[2] If machine intelligence turns out to be even half as impactful as these past technologies, it will still be one of the key developments of this century. Indeed, its stakes could be even greater, with some arguing that AI will soon constitute one of the most important global challenges facing the world[3] and that, if handled poorly, the most severe outcomes could well be irreversible or even catastrophic for human societies worldwide.[4]

---

[1] Mustafa Suleyman and Michael Bhaskar, *The Coming Wave: Technology, Power, and the Twenty-First Century's Greatest Dilemma* (Crown 2023).

[2] Lauren Goode, 'Google CEO Sundar Pichai Says AI Is More Profound than Electricity or Fire' (*The Verge*, 19 January 2018) <https://www.theverge.com/2018/1/19/16911354/google-ceo-sundar-pichai-ai-artificial-intelligence-fire-electricity-jobs-cancer> accessed 12 September 2018; Shana Lynch, 'Andrew Ng: Why AI Is the New Electricity' [2017] *Stanford Graduate School of Business* <https://www.gsb.stanford.edu/insights/andrew-ng-why-ai-new-electricity> accessed 22 October 2018; Michael C Horowitz, 'Artificial Intelligence, International Competition, and the Balance of Power' [2018] *Texas National Security Review* <https://tnsr.org/2018/05/artificial-intelligence-international-competition-and-the-balance-of-power/> accessed 17 May 2018; Greg Allen and Taniel Chan, 'Artificial Intelligence and National Security' (Harvard Kennedy School, Belfer Center for Science and International Affairs 2017) <http://www.belfercenter.org/sites/default/files/files/publication/AI%20NatSec%20-%20final.pdf> accessed 19 July 2017.

[3] Yoshua Bengio and others, 'Managing Extreme AI Risks amid Rapid Progress' (2024) 384 Science 842.

[4] Yoshua Bengio and others, 'International AI Safety Report' (2025) DSIT 2025/001 <https://www.gov.uk/government/publications/international-ai-safety-report-2025>; Tom Davidson, 'The Danger of Runaway AI' (2023) 34 Journal of Democracy 132; Atoosa Kasirzadeh, 'Two Types of AI Existential Risk: Decisive and Accumulative' [2025] Philosophical Studies <http://arxiv.org/abs/2401.07836> accessed 20 February 2024; Karina Vold and Daniel R Harris, 'How Does Artificial Intelligence Pose an

For some, these prospects are grounds for anticipation and excitement; for others, they are cause for concern, caution. or even alarm. For all, the stakes involved should be grounds for making deliberate, informed, and effective choices. To choose well in the face of a transformative technology, and to govern it in a deliberate, informed and effective way, is to reckon with questions of change in technology, in society, and in law and governance.

Effective tools of governance will be key in making these choices well, and in determining whether AI becomes a driver of societal progress or decline. Yet while the challenges of AI are widely viewed as increasingly urgent on the global stage, current governance debates struggle to make progress, especially at the global level. Disciplinary gaps along with a disconnect between the changing pace of the technology and international legal approaches to regulating technology mean that governance efforts in response to AI technology are often reactive, fragmented, and grounded in shallow analogies to past technologies or legacy institutions. Unless these knots are untied, effective AI governance may not be established, in time or at all, to allow the world to meet the profound social changes AI will entail.

Humanity is at a crossroads with AI: there is an urgent need to make choices that ensure that regulators and global governance institutions alike are up to the task of governing it well. Will they—will we—succeed? How can we (re)design our global governance institutions to meet the challenges and promises of this increasingly advanced technology?

Summary of the Book

In this book, I aim to explore these questions. The book argues that global governance for AI can and must do better than it has to date, and that it can do so by drawing on the concepts of *sociotechnical change*, *governance disruption*, and *regime complexity*, drawn from established fields such as technology law and global governance scholarship.

These frameworks, the book argues, help ensure that global governance institutions and norms for AI can be adapted or scaled in response to three vectors of ongoing global change: (1) changes in the *sociotechnical problems* to be addressed by governance; (2) changes in the *tools* of governance; (3) changes in the *environment* of governance. These frameworks, the book argues, ground richer analytical descriptions, explanations, and projections of key trends in the emerging global

Existential Risk?' in C Veliz (ed), *Oxford Handbook of Digital Ethics* (Oxford University Press 2021); Yoshua Bengio, 'AI and Catastrophic Risk' (2023) 34 Journal of Democracy 111; Jan Kulveit and others, 'Gradual Disempowerment: Systemic Existential Risks from Incremental AI Development' (arXiv, 29 January 2025) <http://arxiv.org/abs/2501.16946> accessed 31 January 2025; see also the survey of AI researchers in Katja Grace and others, 'Thousands of AI Authors on the Future of AI' (arXiv, 5 January 2024) <http://arxiv.org/abs/2401.02843> accessed 8 January 2024.

architecture of AI governance under these conditions. Simultaneously, they help AI governance advocates and practitioners give greater weight to a range of new approaches and strategies to ensure that institutions and policies will be more *effective* in responding to AI's diverse societal impacts, more *resilient* to future AI-driven disruption to laws, and more *coherent* in terms of the different institutions and norms in the AI governance regime complex. Through this approach, we can adapt our global governance proposals for advanced AI in a manner that is more robust to ongoing and future changes in AI technology, in law, and in international politics.

As such, this book is not just an investigation into how to govern AI, nor just a study of how AI may change governance. It is an exploration of what AI can tell us about the changing nature of global cooperation, today and into the future. How can we govern this changing technology, in a changing world, while using global governance instruments that may themselves become subject to technology-driven change?

This book is split into three parts: it comprehensively explores the foundations of AI governance (Part I); introduces three conceptual lenses for understanding the challenges facing this architecture (Part II); and synthesizes these in a practical framework for analysing and shaping global AI governance (Part III).

Background: The Advanced AI Governance Challenge

AI is significantly reshaping the world. But what is AI? Many debates over the global governance of AI frequently start from, and rapidly flounder on, attempts to pin down a single definition, a challenge compounded by the widely diverse perspectives, framings, and analogies applied to the technology,[5] as well as by the dizzying range of terms for AI systems—including many concepts which are operationalized in different or inconsistent ways.[6] AI researchers have characterized 'AI' as a scientific field focused on 'making machines intelligent, [where] intelligence is that quality that enables an entity to function appropriately and with foresight in its environment'.[7] Yet, in a practical, technical perspective, 'AI' has served as an umbrella term for a motley range of techniques—ranging from traditional

---

[5] For a survey of framings, see Matthijs Maas, 'AI Is Like . . . A Literature Review of AI Metaphors and Why They Matter for Policy' (Institute for Law & AI 2023) AI Foundations Report 2 <https://law-ai.org/ai-policy-metaphors> accessed 25 February 2025.

[6] Matthijs Maas, 'Concepts in Advanced AI Governance: A Literature Review of Key Terms and Definitions' (Institute for Law & AI 2023) AI Foundations Report 3 < https://law-ai.org/advanced-ai-gov-concepts> accessed 25 February 2025; Elliot Jones, 'Explainer: What Is a Foundation Model?' (Ada Lovelace Institute 2023) <https://www.adalovelaceinstitute.org/resource/foundation-models-explainer/> accessed 31 August 2023. For further discussion of definitions of AI, see also Chapter 1.

[7] Nils J Nilsson, *The Quest for Artificial Intelligence: A History of Ideas and Achievements* (Cambridge University Press 2010) xiii. For a classic overview, see Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach* (4th edn, Pearson 2020).

rule-based symbolic AI approaches to the more modern data-driven machine learning (ML) approaches which have been responsible for the surge in AI progress this last decade, all the way to the contemporary 'foundation models',[8] such as transformer-based large language models (LLMs),[9] which have over the past few years enabled the swift rise of so-called generative AI[10] and general-purpose AI systems (GPAIS).[11]

In this book, however, I will approach this technology in a broader societal perspective, which considers it as a stack of artefacts, processes, and practices at different levels of abstraction, covering: (1) a diverse suite of computational *techniques* which improve the accuracy, speed, and/or scale of machine decision-making across diverse information-processing or decision-making contexts; which produce (2) *capabilities* that can be used to support, substitute for, or improve upon, human performance in tasks in many domains, such as (among others) data classification or generation, pattern recognition, prediction, optimization, anomaly detection, or autonomous decision-making; which enable (3) useful *applications* across domains and industries; which drive (4) diverse *societal impacts*.

At whichever level we consider AI—techniques, capabilities, applications, or societal impacts—it is increasingly clear that recent years are seeing a rate of system capability progress that is surprising, steep, and (so far) sustained. To be clear, many of today's AI systems remain subject to a range of limitations, and face increasingly steep and restrictive resource requirements. Yet even so, they have at long last begun to demonstrate remarkable capabilities in, and across, many key tasks and domains.[12]

---

[8] Rishi Bommasani and others, 'On the Opportunities and Risks of Foundation Models' [2021] arXiv:2108.07258 [cs] <http://arxiv.org/abs/2108.07258> accessed 18 August 2021 (defining 'foundation models' as 'models trained on broad data at scale [. . .] that are adaptable to a wide range of downstream tasks').

[9] Helen Toner, 'What Are Generative AI, Large Language Models, and Foundation Models?' (*Center for Security and Emerging Technology*, 12 May 2023) <https://cset.georgetown.edu/article/what-are-generative-ai-large-language-models-and-foundation-models/> accessed 25 May 2023.

[10] Laura Weidinger and others, 'Sociotechnical Safety Evaluation of Generative AI Systems' (arXiv, 18 October 2023) 6 <http://arxiv.org/abs/2310.11986> accessed 20 October 2023 (defining 'generative AI' as 'models that input and output any combination of image, audio, video, and text. This includes transformer-based systems, such as large language models, diffusion-based systems, and hybrid architectures').

[11] See Carlos I Gutierrez and others, 'A Proposal for a Definition of General Purpose Artificial Intelligence Systems' (2023) 2 Digital Society 36 (defining 'general-purpose AI systems' as '[a]n AI system that can accomplish or be adapted to accomplish a range of distinct tasks, including some for which it was not intentionally and specifically trained').

[12] For overviews of trends in AI system performance and application, see generally Nestor Maslej and others, 'The AI Index 2024 Annual Report' (AI Index Steering Committee, Human-Centered AI Initiative, Stanford University 2024) <https://aiindex.stanford.edu/report/?sf187707917=1> accessed 16 April 2024; Epoch, 'AI Trends' (*Epoch*, 11 April 2023) <https://epochai.org/trends> accessed 10 January 2024. For definitions and explanations of AI in a legal context, see Jacob Turner, *Robot Rules: Regulating Artificial Intelligence* (Springer Berlin Heidelberg 2018) 7–21; David Lehr and Paul Ohm, 'Playing with the Data: What Legal Scholars Should Learn About Machine Learning' (2017) 51 UC Davis Law Review 653; Jonas Schuett, 'Defining the Scope of AI Regulations' (2023) 15 Law, Innovation and Technology 1.

As a result, AI has come into its own as a widely applicable set of technologies, with applications in virtually all sectors.[13] While the performance of today's AI systems is not without problems and limits, fewer and fewer people today doubt that AI's development and proliferation will ultimately come to impact every aspect of human society. Indeed, the emergence of increasingly capable and advanced AI systems are, one could suggest, a landmark moment in history. After all, if many human achievements and successes are the result of our 'intelligent' or adaptive behaviour (broadly and functionally defined), then the creation of technologies that manage to automate—or even merely mimic,[14] in a philosophical perspective—even some parts of this capacity is likely a major development,[15] sufficient to drive significant upheaval in all societies that take up this technology.

The Spectrum of AI's Global Challenges

While many hail the benefits that AI could bring to humanity, there is growing public and expert concern about the technology's diverse impacts and challenges.[16]

---

[13] See Richard Van Noorden and Jeffrey M Perkel, 'AI and Science: What 1,600 Researchers Think' (2023) 621 Nature 672, and previously Maithra Raghu and Eric Schmidt, 'A Survey of Deep Learning for Scientific Discovery' [2020] arXiv:2003.11755 [cs, stat] <http://arxiv.org/abs/2003.11755> accessed 29 June 2020. See also the discussion of various breakthroughs in Chapter 1.

[14] Melanie Mitchell and David C Krakauer, 'The Debate Over Understanding in AI's Large Language Models' (2023) 120 Proceedings of the National Academy of Sciences e2215907120.

[15] Ross Gruetzemacher and Jess Whittlestone, 'The Transformative Potential of Artificial Intelligence' (2022) 135 Futures 102884. However, a counterargument could also be made that AI is in fact far from our first 'intelligence technology', given that past (social) technologies such as markets or bureaucracies have also managed to embed, externalize, and (in some cases) exceed some aspects of human cognition. Allan Dafoe, 'AI Governance: Overview and Theoretical Lenses' in Justin B Bullock and others (eds), *The Oxford Handbook of AI Governance* (Oxford University Press 2023).

[16] For a range of relevant opinion surveys, see also: Noemi Dreksler and others, 'What Does the Public Think About AI? An Overview of the Public's Attitudes towards AI and a Resource for Future Research' (Centre for the Governance of AI 2025) <https://www.governance.ai/research-paper/what-does-the-public-think-about-ai>; Centre for Data Ethics and Innovation, 'International Survey of Public Opinion on AI Safety' (UK Government, Department for Science, Innovation and Technology and Centre for Data Ethics and Innovation 2023) <https://www.gov.uk/government/publications/international-survey-of-public-opinion-on-ai-safety> accessed 11 December 2023; Noemi Dreksler and others, 'Preliminary Survey Results: US and European Publics Overwhelmingly and Increasingly Agree That AI Needs to Be Managed Carefully' (*GovAI Blog*, 17 April 2023) <https://www.governance.ai/post/increasing-consensus-ai-requires-careful-management> accessed 19 April 2023; however, global opinion surveys do show significant regional variation in perceptions and concerns, as seen in Peter John Loewen and others, 'Global Public Opinion on Artificial Intelligence' (Schwartz Reisman Institute for Technology and Society 2024) <https://srinstitute.utoronto.ca/public-opinion-ai> accessed 19 April 2024; other domestic surveys also show rising concerns. For instance, in the US context, see: AI Policy Institute, 'Poll Shows Overwhelming Concern About Risks From AI as New Institute Launches to Understand Public Opinion and Advocate for Responsible AI Policies' (9 August 2023) <https://theaipi.org/poll-shows-overwhelming-concern-about-risks-from-ai-as-new-institute-launches-to-understand-public-opinion-and-advocate-for-responsible-ai-policies/> accessed 10 August 2023; Baobao Zhang and Allan Dafoe, 'U.S. Public Opinion on the Governance of Artificial Intelligence', *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (ACM 2020) <http://dl.acm.org/doi/10.1145/3375627.3375827> accessed 12 February 2020; Ross Gruetzemacher and

Indeed, concerns over these risks may be rapidly growing in recent years.[17] As a technology aimed at automating various human activities either in part or in full, AI, almost by definition, has the potential to reach into virtually any sphere of human activity. The use of AI systems is driving considerable social changes—both for better and for worse—across a staggering array of domains.[18]

Many of these concerns have been exacerbated and scaled up by the ascent of large language models (LLMs), and their application in 'generative AI' tools.[19] Globally, concerns are growing over the contributions AI might make to pervasive economic and social instability,[20] including through large-scale work displacement and possible changes in the nature of work itself.[21] Concurrently, the rise in new military AI systems, which include but are certainly not limited to 'lethal autonomous weapons systems' (LAWS), has raised concerns over these weapons' incompatibility with the laws of war and norms of human dignity; risks of terrorist weaponization; their impacts on the dynamics of interstate crisis instability and even (nuclear) escalation;[22] or their effects on the dynamics of domestic and civil conflict.[23] Furthermore, AI technology is projected to contribute to widespread challenges in political economy, global power relations, surveillance,

---

others, 'Implications for Governance in Public Perceptions of Societal-Scale AI Risks' (arXiv, 10 June 2024) <http://arxiv.org/abs/2406.06199> accessed 26 July 2024.

[17] For example, one 2023 opinion survey of 3,500 experts across 50 countries, and 20,000 members of the general public in 15 countries, found that concerns over AI had jumped to the 4th-highest rank (from 14th-highest in 2022), and that 64% of risk experts and 70% of the general public believed 'there should be a break in research into AI and other disruptive technologies'. AXA, 'Future Risks Report 2023' (AXA 2023) 5 <https://www.axa.com/en/press/publications/future-risks-report-2023-report> accessed 1 August 2024.

[18] Usman Anwar and others, 'Foundational Challenges in Assuring Alignment and Safety of Large Language Models' 73–104 <https://llm-safety-challenges.github.io/challenges_llms.pdf> accessed 25 February 2025; Iason Gabriel and others, 'The Ethics of Advanced AI Assistants' (Google DeepMind 2024) <https://storage.googleapis.com/deepmind-media/DeepMind.com/Blog/ethics-of-advanced-ai-assistants/the-ethics-of-advanced-ai-assistants-2024-i.pdf> accessed 25 February 2025.

[19] Laura Weidinger and others, 'Taxonomy of Risks Posed by Language Models', *2022 ACM Conference on Fairness, Accountability, and Transparency* (ACM 2022) <https://dl.acm.org/doi/10.1145/3531146.3533088> accessed 8 May 2023.

[20] Nick Bostrom, Allan Dafoe, and Carrick Flynn, 'Public Policy and Superintelligent AI: A Vector Field Approach' in SM Liao (ed), *Ethics of Artificial Intelligence* (Oxford University Press 2019).

[21] Mary L Gray and Siddharth Suri, *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass* (Houghton Mifflin Harcourt 2019); Carl Benedikt Frey and Michael A Osborne, 'The Future of Employment: How Susceptible Are Jobs to Computerisation?' (2017) 114 Technological Forecasting and Social Change 254. Though for a more optimistic view, see John Danaher, *Automation and Utopia: Human Flourishing in a World without Work* (Harvard University Press 2019).

[22] Nehal Bhuta and others (eds), *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press 2016); Jürgen Altmann and Frank Sauer, 'Autonomous Weapon Systems and Strategic Stability' (2017) 59 Survival 117; Matthijs Maas, Kayla Lucero-Matteucci and Di Cooke, 'Military Artificial Intelligence as a Contributor to Global Catastrophic Risk', *The Era of Global Risk* (Open Book Publishers 2023) <https://www.openbookpublishers.com/books/10.11647/obp.0336/chapters/10.11647/obp.0336.10> accessed 9 August 2023.

[23] Lance Y Hunter and others, 'The Fourth Industrial Revolution, Artificial Intelligence, and Domestic Conflict' (2023) 37 Global Society 375.

authoritarianism,[24] and cyber-power,[25] among broader changes to the fabric of international relations.[26] Finally, recent years have seen concern over a set of cross-spectrum societal-scale risks that, many expect, will result from the deployment of increasingly capable general-purpose AI systems at the frontier of the field.[27] Such systems—which may acquire a wide range of new capabilities matched with increasing levels of autonomy—could contribute to a range of significant and potentially global catastrophic risks, deriving from their potential misuse, misalignment, or loss of control over autonomous systems, or from their broader systemic impacts.[28] While still emerging, these risks have become of increasing global and national concern to many governments and international organizations alike.[29]

[24] Allan Dafoe, 'AI Governance: A Research Agenda' (Center for the Governance of AI, Future of Humanity Institute 2018) <https://www.fhi.ox.ac.uk/govaiagenda/> accessed 25 February 2025; Dafoe (n 15).

[25] Joe Devanny, 'Artificial Intelligence and Cyber Power: Foreign Policy Implications' (2024) 63 Research Publications <https://digitalcommons.fiu.edu/jgi_research/63> accessed 25 February 2025.

[26] Stephane J Baele and others, 'AI IR: Charting International Relations in the Age of Artificial Intelligence' (2024) 26 International Studies Review viae013.

[27] These have also been described as 'highly capable foundation models' or 'frontier AI'. See Elizabeth Seger and others, 'Open-Sourcing Highly Capable Foundation Models: An Evaluation of Risks, Benefits, and Alternative Methods for Pursuing Open-Source Objectives' (Centre for the Governance of AI 2023) <https://www.governance.ai/research-paper/open-sourcing-highly-capable-foundation-models>; Markus Anderljung and others, 'Frontier AI Regulation: Managing Emerging Risks to Public Safety' (arXiv, 11 July 2023) <http://arxiv.org/abs/2307.03718> accessed 7 August 2023. However, note that there remains some contestation over the term 'frontier AI', and over how it would be best defined or operationalized in the context of regulation. See Jones (n 6); Helen Toner and Shelton Fitch, 'Regulating the AI Frontier: Design Choices and Constraints' (*Center for Security and Emerging Technology*, 26 October 2023) <https://cset.georgetown.edu/article/regulating-the-ai-frontier-design-choices-and-constraints/> accessed 25 November 2023; Maas, 'Concepts in Advanced AI Governance' (n 6) 50–52. See also Chapter 2.

[28] Bengio and others (n 4); Bengio and others (n 3). Michael K Cohen and others, 'Regulating Advanced Artificial Agents' (2024) 384 Science 36. See also Lewis Ho and others, 'International Institutions for Advanced AI' (arXiv, 10 July 2023) <http://arxiv.org/abs/2307.04699> accessed 11 July 2023. ('the potential dangerous capabilities of powerful and general-purpose AI systems create global externalities in their development and deployment'). For other discussions of the risks from advanced, general-purpose AI systems, see also Pegah Maham and Sabrina Küspert, 'Governing General Purpose AI: A Comprehensive Map of Unreliability, Misuse and Systemic Risks' (Stiftung Neue Verantwortung 2023) <https://www.stiftung-nv.de/de/publikation/governing-general-purpose-ai-comprehensive-map-unreliability-misuse-and-systemic-risks>; Yoshua Bengio, 'International Scientific Report on the Safety of Advanced AI: Interim Report' (Department for Science, Innovation & Technology, and AI Safety Institute 2024) 2024/009 <https://www.gov.uk/government/publications/international-scientific-report-on-the-safety-of-advanced-ai> accessed 21 May 2024.

[29] Department for Science, Innovation & Technology, 'New Commitment to Deepen Work on Severe AI Risks Concludes AI Seoul Summit' (*GOV.UK*, 22 May 2024) <https://www.gov.uk/government/news/new-commitmentto-deepen-work-on-severe-ai-risks-concludes-ai-seoul-summit> accessed 24 May 2024; Government Office for Science, 'Future Risks of Frontier AI: Which Capabilities and Risks Could Emerge at the Cutting Edge of AI in the Future?' (UK Government Office for Science 2023) <https://assets.publishing.service.gov.uk/media/653bc393d10f3500139a6ac5/future-risks-of-frontier-ai-annex-a.pdf> accessed 25 February 2025; Jim Mitre and Joel B Predd, 'Artificial General Intelligence's Five Hard National Security Problems' (RAND Corporation 2025) <https://www.rand.org/pubs/perspectives/PEA3691-4.html> accessed 13 February 2025.

Developments in Global AI Governance

Many national and global polls indicate rising public concern over AI development and its potential risks.[30] There have been growing calls from civil society, policymakers, and, indeed, AI companies themselves, demanding the implementation of policies and regulation to adequately address these challenges.[31]

The technology has also made its way to the top of international and multilateral agendas. In 2023, UN Secretary-General António Guterres identified AI as a technology that would have a 'dramatic impact on sustainable development, the world of work, and the social fabric',[32] and a range of global governance efforts and institutional initiatives for the technology now date back almost a decade.

Yet the difficulty of addressing these challenges internationally is compounded by the (perceived) strategic stakes of AI, and by rising tensions around digital governance as well as over the architecture of global cooperation more broadly.[33] AI, it has been argued, will be 'the biggest geopolitical revolution in human history'.[34] It has been suggested that whichever nation leads in AI technology will dominate global politics for decades to come.[35]

---

[30] Loewen and others (n 16); Centre for Data Ethics and Innovation (n 16); Dreksler and others (n 16); Zhang and Dafoe (n 16); AI Policy Institute (n 16).

[31] AI Now Institute, 'Zero Trust AI Governance' (Accountable Tech, AI Now Institute, EPIC 2023) <https://ainowinstitute.org/publication/zero-trust-ai-governance>; David McCabe, 'Microsoft Calls for A.I. Rules to Minimize the Technology's Risks' *The New York Times* (25 May 2023) <https://www.nytimes.com/2023/05/25/technology/microsoft-ai-rules-regulation.html> accessed 10 January 2024; Courtney Rozen, 'AI Leaders Are Calling for More Regulation of the Tech. Here's What That May Mean in the US' *Washington Post* (31 July 2023) <https://www.washingtonpost.com/business/2023/07/27/regulate-ai-here-s-what-that-might-mean-in-the-us/f91462c8-2caa-11ee-a948-a5b8a9b62d84_story.html> accessed 10 January 2024. See also Microsoft, 'Global Governance: Goals and Lessons for AI' (Microsoft 2024) <https://aka.ms/AIGlobalGovernanceBook> accessed 25 February 2025.

[32] António Guterres, 'Secretary-General's Remarks to the Security Council on Artificial Intelligence | United Nations Secretary-General' (*United Nations Secretary-General*, 18 July 2023) <https://www.un.org/sg/en/content/sg/speeches/2023-07-18/secretary-generals-remarks-the-security-council-artificial-intelligence> accessed 18 September 2023.

[33] Thorsten Jelinek, Wendell Wallach, and Danil Kerimi, 'Policy Brief: The Creation of a G20 Coordinating Committee for the Governance of Artificial Intelligence' [2020] AI and Ethics <https://doi.org/10.1007/s43681-020-00019-y> accessed 30 October 2020; Ian Bremmer and Mustafa Suleyman, 'The AI Power Paradox' [2023] *Foreign Affairs* <https://www.foreignaffairs.com/world/artificial-intelligence-power-paradox> accessed 24 August 2023.

[34] Kevin Drum, 'Tech World: Welcome to the Digital Revolution' [2018] *Foreign Affairs* 46.

[35] For instance, Indermit Gill has suggested that 'whoever leads in artificial intelligence in 2030 will rule the world until 2100'. Indermit Gill, 'Whoever Leads in Artificial Intelligence in 2030 Will Rule the World until 2100' (*Brookings*, 17 January 2020) <https://www.brookings.edu/blog/future-development/2020/01/17/whoever-leads-in-artificial-intelligence-in-2030-will-rule-the-world-until-2100/> accessed 22 January 2020. Note, however, that an often-repeated claim by Russian President Vladimir Putin, that 'whoever rules AI rules the world', may have been taken out of context: rather than an official statement of Russian foreign policy, this appears to have been an off-the-cuff comment which Putin made in the context of giving young Russian schoolchildren feedback on their science projects. Interview with Robert Wiblin, Keiran Harris, and Allan Dafoe, 'The Academics Preparing for the Possibility that AI Will Destabilise Global Politics' (18 March 2018) <https://80000hours.org/podcast/episodes/allan-dafoe-politics-of-ai/> accessed 12 August 2020.

As such, it should be no surprise that in just a short few years, dozens of nation-states have articulated national AI strategies,[36] and have begun investing vast sums in AI research, application, and the underlying computing hardware infrastructure. Both the US and China, considered global AI powers, consider the technology a lynchpin of their future strategic dominance.[37] Beyond them, a set of challenger states that are relatively new to the AI scene—among them France, Saudi Arabia, South Korea, Singapore, and the United Arab Emirates—have all begun coordinated efforts to develop their own AI industries and become rising powers in the technology.[38] While not all national AI investment programmes have persisted,[39] many of these states are eager to secure their place among the lead innovators in this technology.

[36] For an overview, see: OECD, 'OECD's Live Repository of AI Strategies & Policies' (*OECD AI Policy Observatory*) <https://oecd.ai/en/dashboards> accessed 8 January 2024. For an analysis, see Fernando Filgueiras, 'Artificial Intelligence Policy Regimes: Comparing Politics and Policy to National Strategies for Artificial Intelligence' (2022) 3 Global Perspectives <https://online.ucpress. edu/gp/article/3/1/32362/119790/Artificial-Intelligence-Policy-Regimes-Comparing> accessed 19 May 2022.

[37] In the US, this was initially articulated (in 2016) under the Obama administration: Office of Science and Technology Policy, 'The National Artificial Intelligence Research and Development Strategic Plan' (National Science and Technology Council 2016) <https://obamawhitehouse.archives. gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/national_ai_rd_strategic_plan.pdf> accessed 26 February 2017. Meanwhile, in 2017 China's State Council issued a plan that anticipated China becoming the world leader in the AI field by 2030: China's State Council, 'A Next Generation Artificial Intelligence Development Plan' (Rogier Creemers and others trs, New America Cybersecurity Initiative 2017) <https://na-production.s3.amazonaws.com/documents/translation-fulltext-8.1.17. pdf> accessed 23 October 2017.

[38] Mark Bergen, 'Microsoft to Invest €4 Billion in French Cloud and AI Services' *Bloomberg.com* (13 May 2024) <https://www.bloomberg.com/news/articles/2024-05-13/microsoft-to-invest-4-bill ion-in-french-cloud-ai-services> accessed 6 June 2024; Commission de l'Intelligence Artificielle, 'Our AI: Our Ambition for France' (2024) <https://www.info.gouv.fr/upload/media/content/0001/ 09/02cbcb40c3541390be391feb3d963a4126b12598.pdf> accessed 10 April 2024; Noah Greene, 'Will France Be the AI Hub of Continental Europe?' (*Tech Policy Press*, 2 April 2024) <https://techpolicy. press/will-france-be-the-ai-hub-of-continental-europe> accessed 6 June 2024; Adam Satariano and Paul Mozur, '"To the Future": Saudi Arabia Spends Big to Become an A.I. Superpower' *The New York Times* (25 April 2024) <https://www.nytimes.com/2024/04/25/technology/saudi-arabia-ai.html> accessed 6 June 2024; Reuters, 'South Korea to Invest $7 Billion in AI in Bid to Retain Edge in Chips' *Reuters* (9 April 2024) <https://www.reuters.com/technology/south-korea-invest-7-bln-ai-bid-retain- edge-chips-2024-04-09/> accessed 6 June 2024; Sheila Chiang, 'Singapore's AI Ambitions Get a Boost with $740 Million Investment Plan' (*CNBC*, 19 February 2024) <https://www.cnbc.com/2024/02/19/ singapores-ai-ambitions-get-a-boost-with-740-million-investment-plan.html> accessed 6 June 2024; Billy Perrigo, 'The UAE Is on a Mission to Become an AI Power' (*TIME*, 20 March 2024) <https://time. com/6958369/artificial-intelligence-united-arab-emirates/> accessed 6 June 2024.

[39] For instance, one notable case of policy reversal came from the United Kingdom. In March 2023, the UK Conservative government under Rishi Sunak announced a national GBP 900 million exascale supercomputer investment dedicated to AI development, along with a GBP 500 million AI Research Resource computing reserve. Dan Milmo and Alex Hern, 'UK to Invest £900m in Supercomputer in Bid to Build Own "BritGPT"' *The Guardian* (15 March 2023) <https://www.theguardian.com/tec hnology/2023/mar/15/uk-to-invest-900m-in-supercomputer-in-bid-to-build-own-britgpt> accessed 8 January 2024. However, less than a year later, the UK's new Labour government reversed course, abandoning the projects. Zoe Kleinman, 'Government Shelves £1.3bn UK Tech and AI Plans' *BBC* (2 August 2024) <https://www.bbc.com/news/articles/cyx5x44vnyeo> accessed 5 August 2024.

Likewise, in the military domain, a range of influential countries have steadily scaled up their AI investments,[40] as well as establishing various strategic partnerships to facilitate the development of AI weapons and military support systems.[41] As a result, some claim we are seeing a new global 'arms race' in all but name, or even an impending 'Cold War' to be fought within the domain of AI,[42] while others anticipate the rise of 'techno-nationalism' around AI.[43]

Even if AI's technological trajectory were not, eventually, to fulfil these ambitious promises, and even if zero-sum framings of global AI development as a race might be misconceived or even hazardous,[44] such depictions increasingly shape global debates, setting the terms for governance efforts.[45]

---

[40] Justin Haner and Denise Garcia, 'The Artificial Intelligence Arms Race: Trends and World Leaders in Autonomous Weapons Development' (2019) 10 Global Policy 331.

[41] Lena Trabucco and Matthijs M Maas, 'Technology Ties: The Rise and Roles of Military AI Strategic Partnerships' (10 November 2023) <https://papers.ssrn.com/abstract=4629283> accessed 10 November 2023; Zoe Stanley-Lockman, 'Military AI Cooperation Toolbox: Modernizing Defense Science and Technology Partnerships for the Digital Age' (Center for Security and Emerging Technology 2021) <https://cset.georgetown.edu/wp-content/uploads/CSET-Military-AI-Cooperation-Toolbox.pdf> accessed 23 August 2021.

[42] Edward Moore Geist, 'It's Already Too Late to Stop the AI Arms Race—We Must Manage It Instead' (2016) 72 Bulletin of the Atomic Scientists 318; Haner and Garcia (n 40); Kai-Fu Lee, *AI Superpowers: China, Silicon Valley, and the New World Order* (Houghton Mifflin Harcourt 2018); Nicholas Thompson and Ian Bremmer, 'The AI Cold War That Threatens Us All' [2018] *Wired* <https://www.wired.com/story/ai-cold-war-china-could-doom-us-all/> accessed 20 November 2018; Michael Auslin, 'Can the Pentagon Win the AI Arms Race?' [2018] *Foreign Affairs* <https://www.foreignaffairs.com/articles/united-states/2018-10-19/can-pentagon-win-ai-arms-race> accessed 20 November 2018; Julian E Barnes and Josh Chin, 'The New Arms Race in AI' *Wall Street Journal* (2 March 2018) <https://www.wsj.com/articles/the-new-arms-race-in-ai-1520009261> accessed 22 November 2018.

[43] Claudio Feijóo and others, 'Harnessing Artificial Intelligence (AI) to Increase Wellbeing for All: The Case for a New Technology Diplomacy' (2020) 44 Telecommunications Policy 101988. See also Ian Hogarth, 'AI Nationalism' (*Ian Hogarth*, 13 June 2018) <https://www.ianhogarth.com/blog/2018/6/13/ai-nationalism> accessed 23 July 2018.

[44] For a range of critiques of the 'arms race' framing, on empirical, conceptual and normative grounds, see for instance Stephen Cave and Seán S ÓhÉigeartaigh, 'An AI Race for Strategic Advantage: Rhetoric and Risks', *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society* (ACM 2018) <https://dl.acm.org/doi/10.1145/3278721.3278780> accessed 12 December 2020; Seán Ó hÉigeartaigh, 'The Most Dangerous Fiction: The Rhetoric and Reality of the AI Race' in Max Rangeley and Nicholas Fairfax (eds), *The Artificial Intelligence Revolution* (Springer Nature 2025) <https://papers.ssrn.com/abstract=5278644> accessed 5 June 2025; and previously: Heather M Roff, 'The Frame Problem: The AI "Arms Race" Isn't One' (2019) 75 Bulletin of the Atomic Scientists 3; Remco Zwetsloot, Helen Toner, and Jeffrey Ding, 'Beyond the AI Arms Race: America, China, and the Dangers of Zero-Sum Thinking' [2018] *Foreign Affairs* <https://www.foreignaffairs.com/reviews/review-essay/2018-11-16/beyond-ai-arms-race> accessed 20 November 2018; Elsa Kania, 'The Pursuit of AI Is More Than an Arms Race' [2018] *Defense One* <https://www.defenseone.com/ideas/2018/04/pursuit-ai-more-arms-race/147579/> accessed 26 April 2018; Arthur Holland Michel, 'Recalibrating Assumptions on AI: Towards an Evidence-Based and Inclusive AI Policy Discourse' (Chatham House 2023) 22–27 <https://www.chathamhouse.org/2023/04/recalibrating-assumptions-ai> accessed 11 June 2024; Kerry McInerney, 'Yellow Techno-Peril: The "Clash of Civilizations" and Anti-Chinese Racial Rhetoric in the US–China AI Arms Race' (2024) 11 Big Data & Society 2.

[45] This should be somewhat nuanced. Already in 2020, a report by the Center for Security and Emerging Technology reviewed coverage of AI in 4,000 English-language articles over a seven-year period, and found that while a growing number of these framed AI development as a competition, these represented a declining proportion of all articles about AI, suggesting narratives around AI are becoming more diverse. Andrew Imbrie and others, 'Mainframes: A Provisional Analysis of Rhetorical Frames in AI' (Center for Security and Emerging Technology 2020) <https://cset.georgetown.edu/research/mainframes-a-provisional-analysis-of-rhetorical-frames-in-ai/> accessed 18 August 2020.

Nonetheless, the stakes of AI have also prompted significant energy and attention for renewed international cooperation in order to govern this technology well, and bend its impacts towards peace, human rights, safety, and welfare. Nonetheless, at the international level, debates around the global governance of AI are relatively novel, having only started in the early 2010s and not picking up sustained momentum until around 2016.[46] However, the pace of such initiatives has increased rapidly since; according to an ongoing inventory of AI regulatory initiatives maintained by the Council of Europe, international organizations (IOs) overtook national authorities as the main source of AI regulation initiatives in 2020, with national authorities launching 170 policy initiatives between 2015 and 2022, and IOs establishing 210 initiatives in the same period.[47]

Much of this activity initially proceeded along parallel tracks for military and (narrow, conventional) civilian AI. Early debates in AI governance focused predominantly on the military domain, and the threats from LAWS.[48] This is perhaps not surprising, given how the spectre of 'killer robots' is one that is both particularly visceral within the popular imagination as well as an issue that is located squarely in the domain of international peace and security, a founding concern of the global legal order.[49] However, there has also been substantial growth in broader global AI governance efforts for conventional, narrow AI systems. This includes an explosive increase in the number of AI ethics codes issued,[50] an accompanying rise in discussions exploring many important questions in the global governance of AI,[51] and a wide range of governance initiatives ranging from private governance

---

[46] For a history of these governance developments, see also Chapter 3.

[47] Council of Europe, 'Datavisualisation of AI Initiatives' (*Artificial Intelligence*, 2023) <https://www.coe.int/en/web/artificial-intelligence/national-initiatives> accessed 8 January 2024; Jonas Tallberg and others, 'The Global Governance of Artificial Intelligence: Next Steps for Empirical and Normative Research' (2023) 25 International Studies Review viad040, 5–6. Note, however, that the OECD AI Policy Observatory has registered more than 1,000 AI policy initiatives, across 69 countries and territories. See OECD, 'OECD's Live Repository of AI Strategies & Policies' (n 36).

[48] Bhuta and others (n 22); Human Rights Watch, *Losing Humanity: The Case against Killer Robots* (Human Rights Watch 2012) <https://www.hrw.org/sites/default/files/reports/arms1112_ForUpload.pdf> accessed 25 February 2025; Thomas Burri, 'International Law and Artificial Intelligence' (2017) 60 German Yearbook of International Law 91.

[49] Colin B Picker, 'A View from 40,000 Feet: International Law and the Invisible Hand of Technology' (2001) 23 Cardozo Law Review 151.

[50] Anna Jobin, Marcello Ienca, and Effy Vayena, 'The Global Landscape of AI Ethics Guidelines' [2019] Nature Machine Intelligence 1; Jessica Fjeld and others, 'Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI' <https://dash.harvard.edu/handle/1/42160420> accessed 16 January 2020; Yi Zeng, Enmeng Lu, and Cunqing Huangfu, 'Linking Artificial Intelligence Principles' [2018] arXiv:1812.04814 [cs] <http://arxiv.org/abs/1812.04814> accessed 30 January 2019.

[51] For overviews, see James Butcher and Irakli Beridze, 'What Is the State of Artificial Intelligence Governance Globally?' (2019) 164 The RUSI Journal 88; Angela Daly and others, 'Artificial Intelligence Governance and Ethics: Global Perspectives' (2019) <https://arxiv.org/ftp/arxiv/papers/1907/1907.03848.pdf> accessed 28 June 2019; Daniel Schiff and others, 'What's Next for AI Ethics, Policy, and Governance? A Global Overview' (ACM 2020) <https://econpapers.repec.org/paper/osfsocarx/8jaz4.htm> accessed 12 January 2020.

initiatives and industry partnerships, standard-setting organizations, various UN initiatives and instruments,[52] and major non-binding principles from the OECD,[53] the G20,[54] and UNESCO,[55] as well as recently concluded negotiations on binding legal frameworks for AI, in the forms of the EU AI Act and the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy, and the Rule of Law.[56] There have also been steps towards the establishment of new organizations and institutions, such as the 2020 Global Partnership on AI (GPAI).[57]

Moreover, the period from 2023 to 2025 and onwards has seen a flurry of multilateral and institutional activity in response to the acute global attention for the potential extreme risks and impacts of increasingly advanced and general-purpose AI systems, including the 2023 UK AI Safety Summit and the Bletchley Declaration,[58] the 2024 Seoul Declaration and Statement of Intent,[59] new provisions on

[52] Eugenio V Garcia, 'Multilateralism and Artificial Intelligence: What Role for the United Nations?' in Maurizio Tinnirello (ed), *The Global Politics of Artificial Intelligence* (CRC Press 2020) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3779866> 25 February 2025.

[53] OECD, 'Recommendation of the Council on Artificial Intelligence' <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449> accessed 28 May 2019.

[54] G20, 'G20 Ministerial Statement on Trade and Digital Economy' <https://www.mofa.go.jp/files/000486596.pdf> accessed 25 May 2020.

[55] UNESCO, 'Recommendation on the Ethics of Artificial Intelligence' <https://unesdoc.unesco.org/ark:/48223/pf0000381137> accessed 6 September 2023.

[56] European Council, 'Artificial Intelligence (AI) Act: Council Gives Final Green Light to the First Worldwide Rules on AI' (21 May 2024) <https://www.consilium.europa.eu/en/press/press-releases/2024/05/21/artificial-intelligence-ai-act-council-gives-final-green-light-to-the-first-worldwide-rules-on-ai/> accessed 22 May 2024; Council of Europe, 'Council of Europe Adopts First International Treaty on Artificial Intelligence' (*Council of Europe, Newsroom*, 17 May 2024) <https://www.coe.int/en/web/portal/-/council-of-europe-adopts-first-international-treaty-on-artificial-intelligence> accessed 22 May 2024.

[57] Global Partnership on Artificial Intelligence, 'Joint Statement from Founding Members of the Global Partnership on Artificial Intelligence' <https://www.gov.uk/government/publications/joint-statement-from-founding-members-of-the-global-partnership-on-artificial-intelligence>.

[58] UK Government, 'The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023' (*GOV.UK*, 1 November 2023) <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> accessed 3 November 2023.

[59] 'Seoul Declaration for Safe, Innovative and Inclusive AI by Participants Attending the Leaders' Session: AI Seoul Summit, 21 May 2024' (*Department for Science, Innovation & Technology, GOV.UK*, 21 May 2024) <https://www.gov.uk/government/publications/seoul-declaration-for-safe-innovative-and-inclusive-ai-ai-seoul-summit-2024/seoul-declaration-for-safe-innovative-and-inclusive-ai-by-participants-attending-the-leaders-session-ai-seoul-summit-21-may-2024> accessed 22 May 2024; 'Seoul Statement of Intent toward International Cooperation on AI Safety Science, AI Seoul Summit 2024 (Annex)' (*Department for Science, Innovation & Technology, GOV.UK*, 21 May 2024) <https://www.gov.uk/government/publications/seoul-declaration-for-safe-innovative-and-inclusive-ai-ai-seoul-summit-2024/seoul-statement-of-intent-toward-international-cooperation-on-ai-safety-science-ai-seoul-summit-2024-annex> accessed 22 May 2024; Department for Science, Innovation & Technology, 'Global Leaders Agree to Launch First International Network of AI Safety Institutes to Boost Cooperation of AI' (*GOV.UK*, 21 May 2024) <https://www.gov.uk/government/news/global-leaders-agree-to-launch-first-international-network-of-ai-safety-institutes-to-boost-understanding-of-ai> accessed 22 May 2024.

general-purpose AI systems in the EU AI Act,[60] the 2023 G7 Hiroshima AI Process,[61] and major new national regulatory initiatives in the US and China.[62]

Many proposals have been made for the way forward. In the field's early years, many focused on how existing international law might be directly applied to AI,[63] or how we might extend or adapt the mandates of existing organizations.[64] However, since 2022 there have been renewed calls for the establishment of entirely new international organizations specifically for AI. These proposals range widely.[65] Among them, we have seen designs for an International Panel on Artificial Intelligence Safety (IPAIS),[66] an AI Observatory,[67] an International AI Organization (IAIO) to certify state jurisdictions compliance with international standards,[68] international consortia for the testing and evaluation of advanced AI systems,[69] or broad-based

---

[60] Benjamin Larsen and Sabrina Küspert, 'Regulating General-Purpose AI: Areas of Convergence and Divergence across the EU and the US' (*Brookings*, 21 May 2024) <https://www.brookings.edu/articles/regulating-general-purpose-ai-areas-of-convergence-and-divergence-across-the-eu-and-the-us/> accessed 22 May 2024; Cornelia Kutterer, 'Regulating Foundation Models in the AI Act: From "High" to "Systemic" Risk' (*MIAI*, 11 January 2024) <https://ai-regulation.com/regulating-foundation-models-in-the-ai-act-from-high-to-systemic-risk/> accessed 31 January 2024.

[61] OECD, 'G7 Hiroshima Process on Generative Artificial Intelligence (AI): Towards a G7 Common Understanding on Generative AI' (OECD 2023) <https://www.oecd-ilibrary.org/science-and-technology/g7-hiroshima-process-on-generative-artificial-intelligence-ai_bf3c0c60-en> accessed 13 September 2023.

[62] Executive Office of the President, 'Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence' (*The White House*, 30 October 2023) <https://bidenwhitehouse.archives.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/> accessed 3 November 2023. Cyberspace Administration of China, 'Interim Measures for the Management of Generative Artificial Intelligence Services' (*China Law Translate*, 13 July 2023) <https://www.chinalawtranslate.com/generative-ai-interim/> accessed 13 January 2024.

[63] Martina Kunz and Seán Ó hÉigeartaigh, 'Artificial Intelligence and Robotization' in Robin Geiß and Nils Melzer (eds), *The Oxford Handbook of the International Law of Global Security* (Oxford University Press 2021); Mark Chinen, *The International Governance of Artificial Intelligence* (Edward Elgar Publishing 2023).

[64] Rumtin Sepasspour, 'A Reality Check and a Way Forward for the Global Governance of Artificial Intelligence' [2023] 79 Bulletin of the Atomic Scientists 304. <https://doi.org/10.1080/00963402.2023.2245249>; Jelinek, Wallach, and Kerimi (n 33).

[65] Matthijs M Maas and José Jaime Villalobos, 'International AI Institutions: A Literature Review of Models, Examples, and Proposals' (Institute for Law & AI 2023) 1 <https://law-ai.org/international-ai-institutions>; Ho and others (n 28).

[66] Mustafa Suleyman and others, 'Proposal for an International Panel on Artificial Intelligence (AI) Safety (IPAIS): Summary' (*Carnegie Endowment for International Peace*, 27 October 2023) <https://carnegieendowment.org/2023/10/27/proposal-for-international-panel-on-artificial-intelligence-ai-safety-ipais-summary-pub-90862> accessed 29 November 2023.

[67] Geoff Mulgan and others, 'The Case for a Global AI Observatory (GAIO)' (*Carnegie Council for Ethics in International Affairs*, 6 June 2023) <https://www.carnegiecouncil.org/media/article/the-case-for-a-global-ai-observatory-gaio-2023> accessed 3 September 2023.

[68] Robert Trager and others, 'International Governance of Civilian AI: A Jurisdictional Certification Approach' (arXiv, 29 August 2023) <http://arxiv.org/abs/2308.15514> accessed 31 August 2023.

[69] Ross Gruetzemacher and others, 'An International Consortium for AI Risk Evaluations' (2023) <https://openreview.net/forum?id=HoIEKQhiRs> accessed 17 December 2023; Jason Hausenloy, Andrea Miotti, and Claire Dennis, 'Multinational AGI Consortium (MAGIC): A Proposal for International Coordination on AI' (arXiv, 13 October 2023) <http://arxiv.org/abs/2310.09217> accessed 3 November 2023; see also more generally Shayne Longpre and others, 'A Safe Harbor for AI

benefit and access-providing organizations.[70] Notably, these proposals often draw an analogy to or are outright modelled on the template of one or another extant international organization,[71] such as the International Atomic Energy Agency (IAEA),[72] the European Organisation for Nuclear Research (CERN),[73] the International Civil Aviation Organisation (ICAO),[74] or the Intergovernmental Panel on Climate Change (IPCC).[75]

Indeed, many of these proposals have gained uptake at a high political level. Global figures ranging from UN Secretary-General António Guterres to OpenAI's CEO Sam Altman, and from former UK Prime Minister Rishi Sunak to a range of former heads of state, have all supported the idea of establishing an AI equivalent of the IAEA.[76] European Commission President Ursula von der Leyen has advocated for a new AI risk-monitoring body modelled on the IPCC,[77] as well as an

Evaluation and Red Teaming' (arXiv, 7 March 2024) <http://arxiv.org/abs/2403.04893> accessed 2 April 2024.

[70]   Ho and others (n 28).
[71]   ibid; Maas and Villalobos (n 65); see also Siegele, 'A Global Agency to Oversee AI Is a Tall Order' [2023] *The Economist* <https://www.economist.com/the-world-ahead/2023/11/13/a-global-agency-to-oversee-ai-is-a-tall-order> accessed 10 January 2024. See also Microsoft (n 31) (reviewing the examples of ICAO, CERN, IAEA, IPCC, and a cluster of organizations active in financial governance, specifically the Bank for International Settlements, Basel, the Financial Stability Board, and the Financial Action Task Force).
[72]   Sam Altman, Greg Brockman, and Ilya Sutskever, 'Governance of Superintelligence' (*OpenAI*, 22 May 2023) <https://openai.com/blog/governance-of-superintelligence> accessed 23 May 2023; Eoghan Stafford and Robert F Trager, 'The IAEA Solution: Knowledge Sharing to Prevent Dangerous Technology Races' (Centre for the Governance of AI 2022) <https://www.governance.ai/research-paper/knowledge-sharing-to-prevent-dangerous-technology-races>; Harry Law and Lewis Ho, 'Can a Dual Mandate Be a Model for the Global Governance of AI?' [2023] Nature Reviews Physics 1; Mark Robinson, 'The Establishment of an International AI Agency: An Applied Solution to Global AI Governance' [2025] *International Affairs* iiaf105.
[73]   Gary Marcus, 'Two Models of AI Oversight—and How Things Could Go Deeply Wrong' [2023] Communications of the ACM <https://cacm.acm.org/blogs/blog-cacm/273791-two-models-of-ai-oversight-and-how-things-could-go-deeply-wrong/fulltext> accessed 19 September 2023; Andrea Miotti, 'We Can Prevent AI Disaster Like We Prevented Nuclear Catastrophe' [2023] *Time* <https://time.com/6314045/prevent-ai-disaster-nuclear-catastrophe/> accessed 18 September 2023. For a discussion of the CERN analogy, and how it has been used in global AI governance debates, see also Kevin Kohler, 'CERN for AI: An Overview' (*AI Analogies*, 7 May 2024) <https://machinocene.substack.com/p/cern-for-ai-an-overview> accessed 23 May 2024.
[74]   Brad Smith, 'Advancing AI Governance in Europe and Internationally' (*Microsoft—EU Policy Blog*, 29 June 2023) <https://blogs.microsoft.com/eupolicy/2023/06/29/advancing-ai-governance-europe-brad-smith/> accessed 10 January 2024; summarizing: Microsoft, 'Governing AI: A Blueprint for the Future' (Microsoft 2023) <https://cdn-dynmedia-1.microsoft.com/is/content/microsoftcorp/micros oft/msc/documents/presentations/CSR/Governing-AI-A-Blueprint-for-the-Future.pdf> accessed 25 February 2025; Trager and others (n 68).
[75]   Joseph Bak-Coleman and others, 'Create an IPCC-like Body to Harness Benefits and Combat Harms of Digital Tech' (2023) 617 Nature 462; Suleyman and others (n 66).
[76]   Altman, Brockman, and Sutskever (n 72); Michelle Nichols, 'UN Chief Backs Idea of Global AI Watchdog like Nuclear Agency' *Reuters* (12 June 2023) <https://www.reuters.com/technology/un-chief-backs-idea-global-ai-watchdog-like-nuclear-agency-2023-06-12/> accessed 10 January 2024; Mary Robinson, 'The Elders Urge Global Co-Operation to Manage Risks and Share Benefits of AI' (31 May 2023) <https://theelders.org/news/elders-urge-global-co-operation-manage-risks-and-share-benefits-ai> accessed 31 May 2023.
[77]   Marine Strauss, 'EU's von Der Leyen Speaks to European Lawmakers' *Reuters* (13 September 2023) <https://www.reuters.com/world/europe/eus-von-der-leyen-speaks-european-lawmakers-2023-09-13/> accessed 10 January 2024.

'European AI Research Council' modelled on CERN, with a potential budget of €100 billion;[78] Before his passing, the late Pope Francis called for a legally binding international treaty on AI,[79] reflecting a concern with the technology that has also been taken up by his successor Pope Leo XIV. Many of these ideas are now avidly pursued by governments and in international bodies.[80]

These ideas are in the air, and—at the time of writing—they are gaining traction within international fora.[81] Others have reacted and pushed back, arguing that new institutions are not at all needed, since AI issues can be adequately governed within existing international initiatives.[82] At the same time, in early 2025, many AI governance initiatives—along with many other areas and institutional structures of international law generally—received significant shocks as a result of pivots in the geopolitical priorities of the US.[83]

Under these challenging technological and political conditions, governments will likely decide within the next few years on whether to establish any global body for AI and—if so—what kind of organization should be pursued, or—if not—how to instead organize the global governance architecture for AI. Are we ready to make such decisions?

[78] Jacob Wulff Wold, 'Von Der Leyen Gives Nod to €100 Billion "CERN for AI" Proposal' *Euractiv* (25 July 2024) <https://www.euractiv.com/section/digital/news/von-der-leyen-gives-nod-to-e100-billion-cern-for-ai-proposal/> accessed 30 July 2024 (noting that this budget—slated over five to seven years—would 'exceed the European Commission's entire seven-year research budget for Horizon Europe, which is €95.5 billion').

[79] Philip Pullella, 'Pope Francis Calls for Binding Global Treaty to Regulate AI' *Reuters* (14 December 2023) <https://www.reuters.com/technology/pope-calls-binding-global-treaty-artificial-intelligence-2023-12-14/> accessed 10 January 2024.

[80] Inter-Agency Working Group on Artificial Intelligence (IAWG-AI), 'United Nations System White Paper on Artificial Intelligence Governance: An Analysis of Current Institutional Models and Related Functions and Existing International Normative Frameworks within the United Nations System That Are Applicable to Artificial Intelligence Governance' (United Nations Chief Executives Board for Coordination 2024) CEB/2024/1/Add.1 <https://unsceb.org/sites/default/files/2024-11/UNSystemWhitePaperAIGovernance.pdf> accessed 14 January 2025.

[81] High-Level Advisory Body on Artificial Intelligence, 'Governing AI for Humanity: Final Report' (United Nations 2024) <https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf>. Inter-Agency Working Group on Artificial Intelligence (IAWG-AI) (n 80). See also Allison Stanger and others, 'Terra Incognita: The Governance of Artificial Intelligence in Global Perspective' <https://www.annualreviews.org/content/journals/10.1146/annurev-polisci-041322-042247> accessed 18 April 2024; Mark Coeckelbergh, 'The Case for Global Governance of AI: Arguments, Counter-Arguments, and Challenges Ahead' [2024] AI & Society <https://doi.org/10.1007/s00146-024-01949-5> accessed 4 May 2024.

[82] Huw Roberts and others, 'Global AI Governance: Barriers and Pathways Forward' [2024] International Affairs iiae073; Cameron F Kerry, 'The Good, the Not-so-Good, and the Ugly of the UN's Blueprint for AI' (*Brookings*, 29 August 2024) <https://www.brookings.edu/articles/the-good-the-not-so-good-and-the-ugly-of-the-uns-blueprint-for-ai/> accessed 2 September 2024; Emma Klein and Stewart Patrick, 'Envisioning a Global Regime Complex to Govern Artificial Intelligence' (Carnegie Endowment for International Peace 2024) <https://carnegieendowment.org/research/2024/03/envisioning-a-global-regime-complex-to-govern-artificial-intelligence?lang=en> accessed 13 June 2024.

[83] Giulia Torchio and Francesco Tasin, 'The Paris Summit: Au Revoir, Global AI Safety?' (*European Policy Centre*, 14 February 2025) <https://epc.eu/en/Publications/The-Paris-Summit-Au-Revoir-global-AI-Safety~61ea68> accessed 19 February 2025; see generally Monica Hakimi and Jacob Katz Cogan, 'The End of the U.S.-Backed International Order and the Future of International Law' [2025] American Journal of International Law 1.

The Hurdles to Global Governance, and a
Window of Opportunity for AI

Clearly, this is an active and exciting time for AI governance. However, the field remains at a vulnerable and uncertain juncture. Many AI governance initiatives are still relatively incipient and fragmented, while AI is developing rapidly, and its adoption proceeds apace. It remains unclear what trajectory existing global governance regimes may take in the coming years, whether or not they will be able to keep pace with the technology, and what is the need for, feasibility of, or optimal design for any new international institutions within that landscape.

All this matters, because in spite of AI's high public profile—or because of it—regulatory success is certainly not guaranteed. To be sure, the idea of multilateralism as dead or irrecoverable may, even today, be premature. As noted by Benvenisti and Downs, '[h]istorically, reports about the death of international law are invariably premature'.[84] While international law has a chequered history, it is one that includes periods where it saw landmark achievements—including shifts in norms which were well in excess of initial expectations—that are often invisible today only because they have become largely normalized.[85] Today, many parts of the international legal architecture, and many global governance regimes, still operate quietly in the background, with most states still meeting most of their binding international commitments,[86] even in high-stakes contexts.[87] Even in the current period of crisis,[88] such systems will continue to be necessary and relevant—even if they may need to evolve in response.

However, guarded optimism should not distract us from the real challenges that face the global governance architecture generally, or the global governance of AI specifically. In spite of the considerable—and often under-appreciated—achievements

---

[84] Eyal Benvenisti and George W Downs, 'Comment on Nico Krisch, "The Decay of Consent: International Law in an Age of Global Public Goods"' (2014) 108 AJIL Unbound 1, 2.

[85] See notably Oona A Hathaway and Scott J Shapiro, *The Internationalists: How a Radical Plan to Outlaw War Remade the World* (Simon & Schuster 2017) (arguing that the 1928 Kellogg-Briand Pact to outlaw war, often remembered critically for failing to avert the Second World War, nonetheless precipitated a significant change in the international legal status of war which, not just immediately but also in the post-War world, drove lasting shifts in State perceptions of—and incentives around—wars of conquest and gunboat diplomacy, significantly affecting the subsequent occurrence of these historically ubiquitous practices and providing more leeway for the use of third-party sanctions while maintaining neutrality and setting the foundations for the modern UN system). See also Oona Hathaway and Scott J Shapiro, 'What Realists Don't Understand About Law' (*Foreign Policy*, 9 October 2017) <https://foreignpolicy.com/2017/10/09/what-realists-dont-understand-about-law/> accessed 17 April 2019 ('this [sceptical] reaction reveals a misunderstanding about how law works. When it is most effective, the law doesn't induce states to act contrary to incentives; it changes those incentives themselves').

[86] This recalls Louis Henkin's observation (though made of a different era) that 'almost all nations observe almost all principles of international law and almost all of their obligations almost all of the time'. Louis Henkin, *How Nations Behave: Law and Foreign Policy* (Columbia University Press 1979) 47; cited in Rebecca Crootof, 'Jurisprudential Space Junk: Treaties and New Technologies' in Chiara Giorgetti and Natalie Klein (eds), *Resolving Conflicts in the Law* (Brill | Nijhoff 2019) 109.

[87] See again Hathaway and Shapiro (n 85).

[88] Hakimi and Cogan (n 83).

of international law in many domains over the past decades and centuries, it is also painfully clear that global governance actors do not always rise to emergent challenges in time—or at all.

For example, even as recent decades have seen a global proliferation in new international institutions and agreements,[89] the coverage of global issues remains partial. Numerous pressing global problems simply remain mostly or entirely ungoverned—under-institutionalized,[90] locked into a 'non-regime' state.[91]

Among such pressing issues that remain in relative international regulatory limbo are challenges such as coral reef degradation,[92] the accumulation of space junk, the recognition of professional qualifications for migrants, and the regulation of potential chemical endocrine disruptors;[93] global plastic pollution;[94] foreign surveillance activities and intelligence collection activities affecting state leaders or foreign private citizens;[95] many forms of cyber operations[96] and cyber-attacks;[97] the security of the internet-critical global submarine cable infrastructure,[98] and the

---

[89] Karen J Alter and Kal Raustiala, 'The Rise of International Regime Complexity' (2018) 14 Annual Review of Law and Social Science 329. See also Chapter 6.

[90] See Jean-Frédéric Morin and others, 'How Informality Can Address Emerging Issues: Making the Most of the G7' (2019) 10 Global Policy 267, 2.

[91] Dimitrov and colleagues have called this a 'nonregime' state. Radoslav S Dimitrov and others, 'International Nonregimes: A Research Agenda' (2007) 9 International Studies Review 230 (defining 'nonregimes' as 'transnational policy arenas characterized by the absence of multilateral agreements for policy coordination among states').

[92] Rado S Dimitrov, 'Confronting Nonregimes: Science and International Coral Reef Policy' (2002) 11 The Journal of Environment & Development 53.

[93] See also Anne van Aaken, 'Is International Law Conducive To Preventing Looming Disasters?' (2016) 7 Global Policy 81, 84 (on the WHO's failure to discuss chemical endocrine disruptors).

[94] Daniel Bodansky, 'The Most Important Negotiation You've (Probably) Never Heard Of' (*EJIL: Talk!*, 21 November 2024) <https://www.ejiltalk.org/the-most-important-negotiation-youve-proba bly-never-heard-of/> accessed 11 December 2024 (reviewing however contemporary progress in the work of the Intergovernmental Negotiating Committee on Plastic Pollution).

[95] Ashley Deeks, 'An International Legal Framework for Surveillance' (2015) 55 Virginia Journal of International Law 291 (however also discussing subsequent developments at UN bodies, judicial tribunals, and beyond to bring such practices under control).

[96] For instance, on disruptive hacking campaigns, see Kristen E Eichensehr, 'Not Illegal: The SolarWinds Incident and International Law' [2022] European Journal of International Law chac060. Though for a counterargument, see Antonio Coco, Talita Dias, and Tsvetelina van Benthem, 'Illegal: The SolarWinds Hack under International Law' (2022) 33 European Journal of International Law 1275 (arguing that this operation was likely illegal under international law rules on sovereignty and non-intervention, as well as under general due diligence duties and international human rights law).

[97] Mette Eilstrup-Sangiovanni, 'Why the World Needs an International Cyberwar Convention' (2018) 31 Philosophy & Technology 379. Note: while many have argued that general norms of existing international law apply to cyberspace—such as, notably, norms of international humanitarian law—it has also been countered that, even where those broad norms exist, they have largely not been effective in shaping the behaviour even of non-state actors, let alone states. See Nori Katagiri, 'Why International Law and Norms Do Little in Preventing Non-State Cyber Attacks' (2021) 7 Journal of Cybersecurity <https://doi.org/10.1093/cybsec/tyab009> accessed 16 March 2021.

[98] Abra Ganz and others, 'Submarine Cables and the Risks to Digital Sovereignty' (12 January 2024) 4–7 <https://papers.ssrn.com/abstract=4693206> accessed 24 June 2024 (noting that although there exists an international forum to discuss relevant technical, legal, and environmental information, in the form of the International Cable Protection Committee (ICPC), the organization can only make non-binding recommendations, and international regulatory frameworks are broadly insufficient). See also Kevin Frazier, 'Policy Proposals for the United States to Protect the Undersea Cable System' (2021) 13 Journal of Law, Technology, & the Internet 12–18 <https://scholarlycommons.law.case.edu/

supervision of cross-border financial institutions under global monetary and financial governance,[99] among others.

This highlights how there are often numerous hurdles that must be overcome before an emergent issue such as AI becomes well governed at the global level. States can be slow to be moved to action on regulating a longstanding issue or a new technology: the 1982 signing of the UN Convention on the Law of the Sea followed half a century of failed attempts to create a treaty regime for the ocean.[100] The same has been true for international responses to new technology. Modern anti-personnel landmines have existed since at least the US Civil War,[101] but the first concerted international legal effort to ban their use did not occur until 1977, and it would take two more decades before there was a measure of success in the form of the 1997 Ottawa Mine Ban Treaty.[102] This highlights that even where there is genuine state interest in multilateral or institutional solutions, negotiating the creation of new instruments is an extremely slow process, which can take many years,[103] and often will stretch into decades.[104]

In other cases, major state efforts to coordinate and consolidate many different rules and regimes into a single integrated and coherent global regime soon floundered, as in the case of the 2017 French initiative to negotiate a new Global Pact for the Environment to unify international environmental law.[105] Despite initially

---

jolti/vol13/iss1/1> accessed 25 February 2025. (reviewing gaps in UNCLOS and other sources of international law, which afford only limited or non-binding protection to undersea cable systems).

[99]   Emilios Avgouleas, 'Effective Governance of Global Financial Markets: An Evolutionary Plan for Reform' (2013) 4 Global Policy 74; Rosa M Lastra, 'Do We Need a World Financial Organization?' (2014) 17 Journal of International Economic Law 787. However, note that the past few years have seen steps towards reforming global financial governance. See Wafa Abedin and others, 'Reforms for a 21st Century Global Financial Architecture: Independent Expert Reflections on the United Nations "Our Common Agenda"' (Global Economy and Development at Brookings 2024) <https://www.brookings.edu/articles/reforms-for-a-21st-century-global-financial-architecture/> accessed 16 April 2024.

[100]   Kirsten Sellars, A 'Constitution for the Oceans': The Long Hard Road to the UN Convention on the Law of the Sea (Cambridge University Press 2025).

[101]   Lorraine Boissoneault, 'The Historic Innovation of Land Mines—And Why We've Struggled to Get Rid of Them' (Smithsonian Magazine, 24 February 2017) <https://www.smithsonianmag.com/innovation/historic-innovation-land-minesand-why-weve-struggled-get-rid-them-180962276/> accessed 7 May 2024.

[102]   Picker (n 49) 184–185.

[103]   For a survey and analysis, see Nicole M Simonelli, 'Bargaining over International Multilateral Agreements: The Duration of Negotiations' (2011) 37 International Interactions 147.

[104]   For another example, one can consider how the international community has taken nearly 20 years to move from initial informal debate to the final negotiation of a new treaty on marine biodiversity, namely in the form of the 2023 Biodiversity Beyond National Jurisdiction (BBNJ) Agreement (also known as the High Seas Treaty). For discussion, see: Rachel Tiller and others, 'Shake It Off: Negotiations Suspended, but Hope Simmering, after a Lack of Consensus at the Fifth Intergovernmental Conference on Biodiversity beyond National Jurisdiction' (2023) 148 Marine Policy 105457; Elizabeth Mendenhall, Rachel Tiller, and Elizabeth Nyman, 'The Ship Has Reached the Shore: The Final Session of the "Biodiversity Beyond National Jurisdiction" Negotiations' (2023) 155 Marine Policy 105686.

[105]   John H Knox, 'The Global Pact for the Environment: At the Crossroads of Human Rights and the Environment' (2019) 28 Review of European, Comparative & International Environmental Law 40; Margaret Young, 'Global Pact for the Environment: Defragging International Law?' (EJIL: Talk!, 29

rapid uptake by the UN General Assembly, the programme quickly ran into dip-lomatic impediments.[106] In yet other cases, states have seemingly delivered on the creation of institutional arrangements, but in fact merely aimed at creating 'face-saving' but ultimately 'empty' institutions: suboptimal, ill-designed institutions with mandates that may deprive them of much or any capacity for effective policy formulation or implementation.[107]

Another risk is for institutional arrangements to leave notable gaps or loopholes in their coverage of all important dimensions of an issue. For instance, while inter-national climate governance has made steps in setting targets for national reduc-tions in greenhouse gas emissions, it has struggled to agree targets for reducing emissions that derive from international aviation and maritime transport activ-ities.[108] Such challenges can be especially acute in fast-changing issue domains: it has been argued that the World Trade Organization-led international trade gov-ernance architecture has been ill-equipped for addressing issues such as the cross-border trade in digital products, or questions of the carbon emissions embodied in traded goods and services.[109]

Even when some issues and domains are indeed effectively governed and insti-tutionalized at the global level, these regimes can face—or even incite—relentless contestation,[110] and persistent governance 'gridlock'.[111] In other domains, such

---

August 2018) <https://www.ejiltalk.org/global-pact-for-the-environment-defragging-international-law/> accessed 25 September 2023.

[106] José Juste Ruiz, 'The Rise and Fall of the Global Pact for the Environment' (2022) 1 Review of International and European Economic Law 28. Lucien Chabason and Elisabeth Hege, 'Failure of the Global Pact for the Environment: A Missed Opportunity or a Bullet Dodged?' (*IDDRI*, 28 May 2019) <https://www.iddri.org/en/publications-and-events/blog-post/failure-global-pact-environm ent-missed-opportunity-or-bullet> accessed 12 December 2023.

[107] On 'empty' institutions (discussing the United Nations Forum on Forests, the Copenhagen Accord on Climate Change, and the UN Commission on Sustainable Development), see Radoslav S Dimitrov, 'Empty Institutions in Global Environmental Politics' (2020) 22 International Studies Review 626. On 'face-saving' institutions (discussing the 1980 UN Convention on Certain Conventional Weapons) see Giovanni Mantilla, 'Deflective Cooperation: Social Pressure and Forum Management in Cold War Conventional Arms Control' (2023) 77 International Organization 564.

[108] Indeed, these types of emissions were excluded by the Kyoto Protocol (under Article 2.2), which instead delegated the negotiation of sector-specific regulations to the International Civil Aviation Organization (ICAO) and the International Maritime Organization (IMO). Progress at these venues since then has been limited, however. See Beatriz Martinez Romera, *Regime Interaction and Climate Change: The Case of International Aviation and Maritime Transport* (Routledge 2017).

[109] Emily Jones and Christopher Adam, 'New Frontiers of Trade and Trade Policy: Digitalization and Climate Change' (2023) 39 Oxford Review of Economic Policy 1.

[110] K Egeland, 'The Road to Prohibition: Nuclear Hierarchy and Disarmament, 1968–2017' (University of Oxford 2017) <https://ora.ox.ac.uk/objects/uuid:b03d68ab-4748-4de7-a2e9-15616 de6a05c> accessed 16 December 2022. That is not to say that such contestation is always counter-productive. For instance, for a discussion of how the NPT along with other global arms control in-stitutions have, as a consequence of such contestation, gradually evolved over time, in ways that have replaced or supplemented old forms of institutional inequality, see Caroline Fehl, 'Unequal Power and the Institutional Design of Global Governance: The Case of Arms Control' (2014) 40 Review of International Studies 505.

[111] Thomas Hale and David Held, *Beyond Gridlock* (Polity Press 2017).

as arms control, past successes and landmark achievements in diplomacy have steadily frayed in recent years because of reinflamed great power rivalry.[112]

Concerningly, global governance for digital technologies is plagued by similar issues. Indeed, global cyberspace governance, intuitively perhaps one of the closest technological analogues for the governance of AI, has seen a number of remarkable multi-stakeholder successes but has also, as mentioned, seen non-regime gaps on issues such as cyber espionage,[113] and has been subject to ongoing contestation, especially in areas of cybersecurity and cyberwarfare.[114]

As if that were not already enough, new governance initiatives today may face even steeper political thresholds and hurdles than those in the past, since many of these issue-specific trends are bracketed by broader shifts and trends in global governance,[115] including a rise in reliance on non-binding international agreements,[116] and growing pressures on multilateralism.

The above discussion clearly suggests there are clear political hurdles that will need to be overcome. Will AI governance initiatives manage to do so? If so, what forms or strategies should policymakers and actors aim for? The world may well have entered a sensitive and fleeting window of opportunity to get it right. The historical record suggests that while institutional decisions or (mis) steps taken at an early stage in governance may eventually be corrected later, these initial choices do create strong path dependencies:[117] they can, and often do, impact the viability, operation, form, and effectiveness of governance on a given issue for decades afterwards. For instance, in 1946 the Baruch Plan famously proposed establishing a centralized global approach to regulating nuclear weapons and power. In spite of remarkable contemporary enthusiasm for the ideals of world federalism, however, the proposal failed, and in doing so marked—or at least accelerated—the beginning of the Cold War.[118] In 1970, the

---

[112] Amy J Nelson, 'Innovation Acceleration, Digitization, and the Arms Control Imperative' (Social Science Research Network 2019) SSRN Scholarly Paper ID 3382956 <https://papers.ssrn.com/abstract=3382956> accessed 29 May 2020.

[113] Eichensehr (n 96).

[114] Daniel W Drezner, 'Technological Change and International Relations' (2019) 33 International Relations 286. See also Eneken Tikk and Mika Kerttunen, 'The Alleged Demise of the UN GGE: An Autopsy and Eulogy' (Cyber Policy Institute 2017) <https://cpi.ee/wp-content/uploads/2017/12/2017-Tikk-Kerttunen-Demise-of-the-UN-GGE-2017-12-17-ET.pdf> accessed 25 February 2025.

[115] See also Chapters 6 and 7.

[116] Curtis Bradley, Jack Landman Goldsmith, and Oona A Hathaway, 'The Rise of Nonbinding International Agreements: An Empirical, Comparative, and Normative Analysis' (2023) 90 The University of Chicago Law Review <https://lawreview.uchicago.edu/print-archive/rise-nonbinding-international-agreements-empirical-comparative-and-normative-analysis> accessed 6 September 2023.

[117] Stephanie C Hofmann and Andrew Yeo, 'Historical Institutionalism and Institutional Design: Divergent Pathways to Regime Complexes in Asia and Europe' (2024) 30 European Journal of International Relations 306 (examining the evolution of security regime complexes in Europe and Asia, depending on the institutional design of the initial focal institutions).

[118] John Simpson, 'The Nuclear Non-Proliferation Regime: Back to the Future?' [2004] Disarmament Forum 12; F Bartel, 'Surviving the Years of Grace: The Atomic Bomb and the Specter of World Government, 1945–1950' (2015) 39 Diplomatic History 275; see also the seminal work by

US diplomat George Kennan proposed the establishment of an 'International Environmental Agency', seeing it as an initial step towards an International Environmental Authority.[119] Since then, the merits of a centralized body for bringing coherence to the fragmented instruments and norms of environmental governance have continued to be a vexing subject in the field of environmental law.[120]

As such, while early governance decisions (or the lack of them) may not be irreversible, they often have long-lasting effects, and the reversal of mistaken decisions can be as arduous a process as their establishment was in the first place. All governance decisions therefore merit careful—and early—consideration. This means that for AI technology, there is a need for thinking about institutional design and regime complex organization at an early stage, lest these choices are later forced by crises, in ways that might lock in poor or suboptimal design decisions. It is important that we get AI governance right, and that we do so soon—as we may not have many future chances to do so.

Towards Advanced AI Governance: The Evolving Debate

In the face of these challenges, and what may be a closing global window of opportunity, how has scholarship responded to the challenges posed by AI?

Given the high-profile visibility of AI in public discourse, the past years have seen a great deal of work on the topic. Since AI technology has applications in so many sectors, and touches on so many abiding philosophical questions, it is not surprising that such work has been undertaken from a wide range of disciplines and perspectives.

Indeed, much early work on AI focused on the conceptual, philosophical, and ethical issues related to 'robots'.[121] This focus gradually expanded, aptly broadening the analysis from 'robots', to 'algorithms', to 'responsible AI', as well as from the final artefact itself to the broader infrastructure of artefacts, actors, and practices involved in the development, dissemination, and deployment of AI systems.[122] The time horizon under analysis also broadened, coming to encompass

---

Joseph Preston Baratta, *The Politics of World Federation: From World Federalism to Global Governance* (Greenwood Publishing Group 2004).

[119] George F Kennan, 'To Prevent a World Wasteland: A Proposal' (1970) 48 *Foreign Affairs* 401.
[120] See for instance Frank Biermann, *A World Environment Organization: Solution or Threat for Effective International Environmental Governance?* (Steffen Bauer ed, 1st edn, Routledge 2005).
[121] Patrick Lin, Keith Abney, and George A Bekey, *Robot Ethics: The Ethical and Social Implications of Robotics* (MIT Press 2011); Patrick Lin, Keith Abney, and Ryan Jenkins (eds), *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence* (Oxford University Press 2017).
[122] Virginia Dignum, *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way* (1st edn, Springer 2019).

both possible 'near-term' impacts of AI as well as 'long-term' impacts that might emerge in the future:[123] for instance, among others, Barrat,[124] Bostrom,[125] Tegmark,[126] Callaghan and others,[127] and Clarke and Whittlestone[128] have considered the longer-term ramifications of advanced AI, considering the transformative and potentially existential risks that could arise if or when such systems were to attain very advanced or even superhuman capabilities.

For their part, legal professionals and scholars have also certainly got down to work on AI-related issues. In response to AI's societal challenges, many legal scholars have started exploring issues around the regulation of AI. Much of this exploration has been nationally focused, aiming to compare the national regulatory strategies for AI used in different jurisdictions.[129] Other research took a law-centric approach:[130] for instance, Turner[131] and Chesterman[132] offer detailed introductions and overviews to many questions in AI regulation, through various questions (such as around AI 'personhood') within domestic legal systems. Likewise, Wishmeyer and Rademacher[133] and Barfield and Pagallo[134] discuss the regulation of AI from a range of legal approaches (e.g. free speech doctrine, criminal law, anti-discrimination law, data protection and privacy, tax law, legal rights, etc.), providing valuable thematic overviews.

However, work on the international or global governance of AI is relatively novel,[135] with major early research agendas only being formulated around 2018.[136] Within a short period, however, recent years have seen the growth and

[123] S Matthew Liao, *Ethics of Artificial Intelligence* (Oxford University Press 2020); David J Gunkel (ed), *Handbook on the Ethics of Artificial Intelligence* (Edward Elgar 2024).

[124] James Barrat, *Our Final Invention: Artificial Intelligence and the End of the Human Era* (9.1.2013 edn, Thomas Dunne Books 2013).

[125] Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford University Press 2014).

[126] Max Tegmark, *Life 3.0: Being Human in the Age of Artificial Intelligence* (1st edn, Allen Lane 2017).

[127] Victor Callaghan and others (eds), *The Technological Singularity - Managing the Journey* (1st edn, Springer-Verlag Berlin Heidelberg 2017).

[128] Sam Clarke and Jess Whittlestone, 'A Survey of the Potential Long-Term Impacts of AI: How AI Could Lead to Long-Term Changes in Science, Cooperation, Power, Epistemics and Values', *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society* (Association for Computing Machinery 2022) <https://doi.org/10.1145/3514094.3534131> accessed 2 August 2022.

[129] Araz Taeihagh, 'Governance of Artificial Intelligence' [2021] Policy and Society 1; Roxana Radu, 'Steering the Governance of Artificial Intelligence: National Strategies in Perspective' (2021) 40 Policy and Society 178; Huw Roberts and others, 'The Chinese Approach to Artificial Intelligence: An Analysis of Policy, Ethics, and Regulation' (2021) 36 AI & Society 59; Huw Roberts and others, 'Achieving a "Good AI Society": Comparing the Aims and Progress of the EU and the US' (2021) 27 Science and Engineering Ethics 68.

[130] Matthijs M Maas, 'Aligning AI Regulation to Sociotechnical Change' in Justin B Bullock and others (eds), *The Oxford Handbook of AI Governance* (Oxford University Press 2022).

[131] Turner (n 12).

[132] Simon Chesterman, *We, the Robots? Regulating Artificial Intelligence and the Limits of the Law* (Cambridge University Press 2021).

[133] Thomas Wischmeyer and Timo Rademacher (eds), *Regulating Artificial Intelligence* (Springer International Publishing 2020).

[134] Woodrow Barfield and Ugo Pagallo (eds), *Research Handbook on the Law of Artificial Intelligence* (Edward Elgar Publishing 2018).

[135] Roberts and others, 'Global AI Governance' (n 82) 3.

[136] Dafoe (n 24).

consolidation of an academic field—and even, potentially, an epistemic community or advocacy network—of 'AI governance'. There are various ways to define this budding field—for instance, as one concerned with

> the study and shaping of local and global governance systems—including norms, policies, laws, processes, and institutions—that affect the research, development, deployment, and use of existing and future AI systems, in ways that help the world choose the role of advanced AI systems in its future, and navigate the transition to that world.[137]

Within a matter of years, this field has produced a growing body of work to chart both existing and emergent global AI governance initiatives,[138] to draw lessons from historical experiences with the regulation of new technologies,[139] and to sketch potential strategies for the road ahead.[140]

Specifically, researchers have examined various governance levers[141] and considering their reach and limits in productively regulating AI's global impacts.[142] These include mechanisms such as the AI ethics principles,[143] technical

---

[137] See also Maas, 'Concepts in Advanced AI Governance' (n 6) 54 (reviewing various technical, policy, and strategy-focused definitions of this field, and on that basis distilling this definition). See also Matthijs Maas, 'Advanced AI Governance: A Literature Review of Problems, Options, and Proposals' (Institute for Law & AI 2023) AI Foundations Report 4 <https://law-ai.org/advanced-ai-gov-litrev>. But see also other definitions, for instance in Jess Whittlestone and Samuel Clarke, 'AI Challenges for Society and Ethics' in Justin B Bullock and others (eds), *The Oxford Handbook of AI Governance* (Oxford University Press 2022) 1 ('the role of AI governance is ultimately to take practical steps to mitigate this risk of harm while enabling the benefits . . .'), or in Dafoe (n 15) 2 ('AI governance refers (1) descriptively to the policies, norms, laws, and institutions that shape how AI is built and deployed, and (2) normatively to the aspiration that these promote good decisions (effective, safe, inclusive, legitimate, adaptive) . . . In one formulation, the field of AI governance studies how humanity can best navigate the transition to advanced AI systems').

[138] Peter Cihon, Matthijs M Maas, and Luke Kemp, 'Fragmentation and the Future: Investigating Architectures for International AI Governance' (2020) 11 Global Policy 545; Lewin Schmitt, 'Mapping Global AI Governance: A Nascent Regime in a Fragmented Landscape' (2022) 2 AI and Ethics 303.

[139] See also the review in Maas, 'Advanced AI Governance' (n 137) 58–85 (providing a review and taxonomy of work that aims to derive lessons from historical cases for understanding AI's development, impacts, and options for its governance).

[140] Stanger and others (n 81); Anwar and others (n 18); Cohen and others (n 28).

[141] See also Maas, 'Advanced AI Governance' (n 137) 102 (defining a 'lever of governance' as 'a tool or intervention that can be used by key actors to shape or affect (1) the primary outcome of advanced AI development; (2) key strategic parameters of advanced AI governance; (3) other key actors' choices or key decisions').

[142] A more detailed discussion of the strengths and drawbacks of various different levers and tools is provided in Chapter 3. For related reviews, see also Chinen (n 63). See also 'Artificial Intelligence and the Challenge for Global Governance: Nine Essays on Achieving Responsible AI' (Chatham House 2024) <https://www.chathamhouse.org/2024/06/artificial-intelligence-and-challenge-global-governance> accessed 12 June 2024.

[143] For overviews, see Jobin, Ienca, and Vayena (n 50); Jessica Fjeld and others, 'Principled Artificial Intelligence: A Map of Ethical and Rights-Based Approaches' (Berkman Klein Center for Internet & Society at Harvard University 2019) <https://ai-hr.cyber.harvard.edu/images/primp-viz.pdf>; Zeng, Lu, and Huangfu (n 50). Mélanie Gornet and others, 'Mapping AI Ethics: A Meso-Scale Analysis of Its Charters and Manifestos', *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery 2024) <https://dl.acm.org/doi/10.1145/3630106.3658545> accessed 13 June 2024.

standards,[144] AI certification schemes,[145] or other forms of soft law[146] that have already sprung up around AI. Other work has focused on incorporating AI into the existing general norms of public international law,[147] such as the existing human rights regime,[148] international security and export control regimes,[149] or international economic and global trade governance,[150] among many others.

[144] Peter Cihon, 'Standards for AI Governance: International Standards to Enable Global Coordination in AI Research & Development' (Center for the Governance of AI, Future of Humanity Institute, University of Oxford 2019) Technical Report <https://www.fhi.ox.ac.uk/wp-content/uplo ads/Standards_-FHI-Technical-Report.pdf> accessed 18 April 2019; Philippe Lorenz, 'AI Governance through Political Fora and Standards Developing Organizations' (Stiftung Neue Verantwortung 2020) <https://www.stiftung-nv.de/de/publikation/ai-governance-through-political-fora-and-standa rds-developing-organizations> accessed 25 February 2025.

[145] Peter Cihon and others, 'AI Certification: Advancing Ethical Practice by Reducing Information Asymmetries' (2021) 2 IEEE Transactions on Technology and Society 200. Certification Working Group, 'Artificial Intelligence Certification: Unlocking the Power of AI through Innovation and Trust' (Schwartz Reisman Institute for Technology and Society, Responsible AI Institute, and World Economic Forum 2024) <https://srinstitute.utoronto.ca/news/ai-certification-ecosystem> accessed 25 February 2025; Philip Matthias Winter and others, 'Trusted Artificial Intelligence: Towards Certification of Machine Learning Applications' [2021] arXiv:2103.16910 [cs, stat] <http://arxiv.org/ abs/2103.16910> accessed 1 October 2021.

[146] Wendell Wallach and Gary E Marchant, 'An Agile Ethical/Legal Model for the International and National Governance of AI and Robotics' 7; Gary Marchant, '"Soft Law" Governance Of Artificial Intelligence' (*AI Pulse*, 25 January 2019) <https://aipulse.org/soft-law-governance-of-artificial-intellige nce/> accessed 26 February 2019.

[147] Kunz and Ó hÉigeartaigh (n 63); Turner (n 12) ch 7. Burri (n 48); Chinen (n 63); Georgios I Zekos, 'AI and International Law' in Georgios I Zekos (ed), *Economics and Law of Artificial Intelligence: Finance, Economic Impacts, Risk Management and Governance* (Springer International Publishing 2021) <https://doi.org/10.1007/978-3-030-64254-9_12> accessed 23 March 2021; Talita de Souza Dias and Rashmin Sagoo, 'AI Governance in the Age of Uncertainty: International Law as a Starting Point' (*Just Security*, 2 January 2024) <https://www.justsecurity.org/90903/ai-governance-in-the-age-of-uncertainty-international-law-as-a-starting-point/> accessed 9 June 2024; JG Castel and Mathew E Castel, 'The Road to Artificial Superintelligence—Has International Law a Role to Play?' (2016) 14 Canadian Journal of Law & Technology <https://ojs.library.dal.ca/CJLT/article/download/ 7211/6256> accessed 25 February 2025.

[148] Eileen Donahoe and Megan MacDuffee Metzger, 'Artificial Intelligence and Human Rights' (2019) 30 Journal of Democracy 115; Lorna McGregor, Daragh Murray, and Vivian Ng, 'International Human Rights Law as a Framework for Algorithmic Accountability' (2019) 68 International & Comparative Law Quarterly 309; Kate Jones, 'AI Governance and Human Rights: Resetting the Relationship' (Chatham House 2023) Research Paper <https://www.chathamhouse.org/2023/01/ai-governance-and-human-rights> accessed 12 June 2024; Silja Vöneky, 'How Should We Regulate AI? Current Rules and Principles as Basis for "Responsible Artificial Intelligence"' in Alexander Bruns and others (eds), *Legal Theory and Interpretation in a Dynamic Society* (1st edn, Nomos Verlagsgesellschaft 2021); Vinodkumar Prabhakaran and others, 'A Human Rights-Based Approach to Responsible AI' (arXiv, 6 October 2022) <http://arxiv.org/abs/2210.02667> accessed 8 August 2023.

[149] Burri (n 48); Kenneth Anderson and Matthew C Waxman, 'Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can' [2013] Law and Ethics for Autonomous Weapon Systems; André Brunel, 'A Proposal for a Semiconductor Export Control Treaty' (2023) 19 Journal of Business & Technology Law <https://digitalcommons.law.umaryland.edu/jbtl/ vol19/iss1/2> 25 February 2025; Reinmar Nindler, 'The United Nation's Capability to Manage Existential Risks with a Focus on Artificial Intelligence' (2019) 21 International Community Law Review 5.

[150] Joshua P Meltzer, 'Toward International Cooperation on Foundational AI Models: An Expanded Role for Trade Agreements and International Economic Policy' [2024] Harvard Data Science Review <https://hdsr.mitpress.mit.edu/pub/14unjde2/release/1> accessed 4 June 2024. Han-Wei Liu and Ching-Fu Lin, 'Artificial Intelligence and Global Trade Governance: A Pluralist Agenda' (2020) 61 Harvard International Law Journal <https://papers.ssrn.com/abstract=3675505> accessed 26 September 2020.

Going beyond these existing architectures, others have begun to propose new initiatives and centralized international agencies, empowered to coordinate national regulatory approaches,[151] license AI systems,[152] jointly develop advanced AI,[153] or enforce caps on the capabilities of advanced systems or (as proxy) on the computational resources used in their training.[154] Yet others have explored new governance innovations within existing informal governance bodies,[155] or even under schemes involving private regulatory markets[156] or regimes founded on natural law.[157]

A different line of work instead examines not instruments, but rather the roles played by various key actors[158] in shaping and constituting the AI governance landscape. Such work has explored how different states[159]—whether traditional 'great power' states[160] or various state coalitions in the West,[161] in the Global South,[162] or

---

[151] Olivia J Erdélyi and Judy Goldsmith, 'Regulating Artificial Intelligence: Proposal for a Global Solution' (2022) 39 Government Information Quarterly 101748; Turner (n 12) ch 6.

[152] Trager and others (n 68).

[153] Hausenloy, Miotti and Dennis (n 69).

[154] Andrea Miotti and Akash Wasil, 'An International Treaty to Implement a Global Compute Cap for Advanced Artificial Intelligence' (30 October 2023) <https://papers.ssrn.com/abstract=4617094> accessed 3 November 2023; Tolga Bilge, 'Treaty on Artificial Intelligence Safety and Cooperation (TAISC)' (2023) <https://taisc.org/taisc> accessed 20 October 2023.

[155] Wallach and Marchant (n 146); See also generally Gary E Marchant and Wendell Wallach, 'Coordinating Technology Governance' (2015) 31 Issues in Science & Technology 43.

[156] Gillian K Hadfield and Jack Clark, 'Regulatory Markets: The Future of AI Governance' <https://arxiv.org/ftp/arxiv/papers/2304/2304.04914.pdf> 25 February 2025. See also Dean W Ball, 'A Framework for the Private Governance of Frontier Artificial Intelligence' (arXiv, 15 April 2025) <http://arxiv.org/abs/2504.11501> accessed 1 June 2025; Philip Moreira Tomei, Rupal Jain and Matija Franklin, 'AI Governance through Markets' (arXiv, 29 January 2025) <http://arxiv.org/abs/2501.17755> accessed 1 June 2025.

[157] Y Weng and T Izumo, 'Natural Law and Its Implications for AI Governance' (2019) 2 Delphi—Interdisciplinary Review of Emerging Technologies 122.

[158] See also Maas, 'Advanced AI Governance' (n 137) 87 (defining 'key actors' as 'actors whose key decisions will have significant impact on shaping the outcomes from advanced AI, either directly (first-order), or by strongly affecting such decisions made by other actors (second-order)').

[159] Some of this has examined states' domestic relations to strategic assets, or 'general-purpose technologies'. See for instance Jeffrey Ding and Allan Dafoe, 'The Logic of Strategic Assets: From Oil to Artificial Intelligence' [2020] arXiv:2001.03246 [cs, econ, q-fin] <http://arxiv.org/abs/2001.03246> accessed 15 January 2020; Jade Leung, 'Who Will Govern Artificial Intelligence? Learning from the History of Strategic Politics in Emerging Technologies' (University of Oxford 2019) <https://ora.ox.ac.uk/objects/uuid:ea3c7cb8-2464-45f1-a47c-c7b568f27665> accessed 25 February 2025.

[160] Eric Schmidt, 'AI, Great Power Competition & National Security' (2022) 151 Daedalus 288; Andrew Imbrie and Elsa B Kania, 'AI Safety, Security, and Stability Among Great Powers: Options, Challenges, and Lessons Learned for Pragmatic Engagement' (Center for Security and Emerging Technology 2019) CSET Policy BRief <https://cset.georgetown.edu/publication/ai-safety-security-and-stability-among-great-powers-options-challenges-and-lessons-learned-for-pragmatic-engagement/ > accessed 25 February 2025; Allison Duettmann and others, 'Artificial General Intelligence: Coordination & Great Powers' [2018] Foresight Institute; Leopold Aschenbrenner, 'Situational Awareness: The Decade Ahead' (2024) <https://situational-awareness.ai/> accessed 25 February 2025.

[161] Andrew Imbrie and others, 'Agile Alliances: How the United States and Its Allies Can Deliver a Democratic Way on AI' (Center for Security and Emerging Technology 2020) <https://cset.georgetown.edu/wp-content/uploads/CSET-Agile-Alliances.pdf> accessed 25 February 2025; Trabucco and Maas (n 41).

[162] Sumaya Nur Adan, 'The Case for Including the Global South in AI Governance Discussions' (*GovAI Blog*, 20 October 2023) <https://www.governance.ai/post/the-case-for-including-the-glo

in Asia[163]—may participate in such governance. It also considers the relative and shifting roles of private (technology) companies,[164] AI researchers and expert epistemic communities,[165] trans-national issue networks of norm entrepreneurs,[166] or various actors across the global AI supply chain[167]—from chip manufacturers[168] to compute cloud providers,[169] and from digital open-source AI model marketplaces[170] to downstream providers and users—and how they could or should participate in shaping and enforcing governance efforts for AI.

To be clear, the field of AI governance is a young, interdisciplinary and methodologically diverse field, one that has a long way to go to fully mature and come into its own. Indeed, there may still be missing conceptual puzzle pieces, 'scattered throughout many disciplinary domains and policy areas'.[171] What are these missing

bal-south-in-ai-governance-conversations> accessed 23 October 2023; Cecil Abungu, Michelle Malonza, and Sumaya Nur Adan, 'Can Apparent Bystanders Distinctively Shape an Outcome? Global South Countries and Global Catastrophic Risk-Focused Governance of Artificial Intelligence' (*arXiv. org*, 7 December 2023) <https://arxiv.org/abs/2312.04616v1> accessed 7 January 2024; Malou Estier, Belinda Cleeland, and Maxime Stauffer, 'Safe and Beneficial Artificial Intelligence for Small-Island Developing States' (Simon Institute for Longterm Governance 2023) <https://www.simoninstitute.ch/blog/post/safe-and-beneficial-artificial-intelligence-for-small-island-developing-states/> accessed 7 November 2023.

[163] Emmie Hine, 'Governing Silicon Valley and Shenzhen: Assessing a New Era of Artificial Intelligence Governance in the United States and China' (2024) 3 Digital Society 50; Sihao Huang, 'Beijing's Vision of Global AI Governance' (*ChinaTalk*, 23 October 2023) <https://www.chinatalk. media/p/beijings-vision-of-global-ai-governance> accessed 21 February 2025.

[164] Jessica Cussins Newman, 'Decision Points in AI Governance' (Center for Long-Term Cybersecurity 2020) 12–29 <https://cltc.berkeley.edu/wp-content/uploads/2020/05/Decision_Points_AI_Governance.pdf> accessed 3 September 2020 (exploring the role of AI ethics advisory committees or shifting AI publication norms).

[165] Haydn Belfield, 'Activism by the AI Community: Analysing Recent Achievements and Future Prospects', *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (ACM 2020) <https://dl.acm.org/doi/10.1145/3375627.3375814> accessed 24 August 2020; Matthijs M Maas, 'How Viable Is International Arms Control for Military Artificial Intelligence? Three Lessons from Nuclear Weapons' (2019) 40 Contemporary Security Policy 285.

[166] Şerif Onur Bahçecik, 'Civil Society Responds to the AWS: Growing Activist Networks and Shifting Frames' (2019) 10 Global Policy 365; Elvira Rosert and Frank Sauer, 'How (Not) to Stop the Killer Robots: A Comparative Analysis of Humanitarian Disarmament Campaign Strategies' (2020) 42 Contemporary Security Policy 4; Belfield (n 165).

[167] Jennifer Cobbe, Michael Veale, and Jatinder Singh, 'Understanding Accountability in Algorithmic Supply Chains', *2023 ACM Conference on Fairness, Accountability, and Transparency* (2023) <http://arxiv.org/abs/2304.14749> accessed 29 September 2023; Girish Sastry and others, 'Computing Power and the Governance of Artificial Intelligence' (2024).

[168] Onni Aarne, Tim Fist, and Caleb Withers, 'Secure, Governable Chips: Using On-Chip Mechanisms to Manage National Security Risks from AI & Advanced Computing' (Center for a New American Security 2024) <https://www.cnas.org/publications/reports/secure-governable-chips> accessed 25 February 2025.

[169] Lennart Heim and others, 'Governing Through the Cloud: The Intermediary Role of Compute Providers in AI Regulation' (Oxford Martin AI Governance Initiative 2024) <https://cdn.governance. ai/Governing-Through-the-Cloud_The-Intermediary-Role-of-Compute-Providers-in-AI-Regulation. pdf> accessed 25 February 2025.

[170] Robert Gorwa and Michael Veale, 'Moderating Model Marketplaces: Platform Governance Puzzles for AI Intermediaries' (SocArXiv 2023) preprint <https://osf.io/6dfk3> accessed 25 November 2023.

[171] Dafoe (n 15) 3.

puzzle pieces? Arguably, there are still important bottlenecks which hinder the development and effective advancement of effective AI governance proposals, while these proposals themselves would gain much from clarification and exploration. As such, this book will focus on three of the most critical hindrances to effective global AI governance.

Three Facets of Change

In governing AI, we encounter a specific challenge that arises in the context of the regulation of any new technology: how can we account for future social, technological, and legal change? Can we scale up or adapt our legal systems, and if so, along what dimensions, and in what ways?

Specifically, we face questions about how we can or should design our AI governance approaches, to ensure that they track and respond to (1) the relation between technological change and *societal* changes, (2) the relation between technological change and changes in our *legal* tools; and (3) the relation between external changes in the governance landscape and changes in the *political conditions* for governance regimes for AI. To make sense of these developments, and to understand our collective understanding of the significance of AI in them, there is much to be gained by developing novel analytical concepts, or transferring such analytical frameworks from other disciplines.[172]

After all, many of these questions are not new, nor are they unique to AI. Indeed, scholars in the field of law, regulation, and technology have long wrestled with them. For instance, Roger Brownsword, Eloise Scotford, and Karen Yeung have previously identified three general themes around the idea of technological 'disruption', namely:

> (1) technology's disruption of legal orders; (2) the wider disruption to regulatory frameworks more generally, often provoking concerns about regulatory legitimacy; and (3) the challenges associated with attempts to construct and preserve regulatory environments that are 'fit for purpose' in a context of rapid technological development and disruption.[173]

However, while such questions have received extensive attention in this field, they remain, with occasional exceptions, under-theorized or under-integrated in the

---

[172] See also Ingvild Bode, 'AI Technologies and International Relations: Do We Need New Analytical Frameworks?' [2024] The RUSI Journal 1.

[173] Roger Brownsword, Eloise Scotford, and Karen Yeung, 'Law, Regulation, and Technology: The Field, Frame, and Focal Questions' in Roger Brownsword, Eloise Scotford, and Karen Yeung (eds), *The Oxford Handbook of Law, Regulation and Technology*, vol 1 (Oxford University Press 2017) 4.

field of global AI governance. This is a problem, because it is particularly in the field of AI that governance may encounter three especially strong drivers of change.

Sociotechnical Change: Aligning Regulation to Societal Changes from AI

In the first place, AI governance proposals should better consider *when, how, and why technological change drives (global) societal changes that warrant particular regulatory interventions*. In general, many recent AI governance proposals focus on specific new applications or visceral incidents involving AI, but elide a comprehensive or explicit account of when, how, or why new capabilities produce or enable societal changes.[174] Ryan Calo has noted how general this problem is: 'even technically-savvy lawyers—and many policy-savvy technologists—consistently struggle with technology as social fact.'[175] As a consequence, policy discussions often get stuck in intractable debates over how to properly define AI, what are the correct 'essential qualities' of the technology,[176] and whether or not these are sufficiently exceptional or new to warrant entirely new laws.[177] Consequently, policies and laws are often formulated in siloed ways that focus on local problems caused by specific use cases of AI (e.g. autonomous drones, facial recognition, autonomous vehicles, chatbots), or from the perspective of particular conventional legal subjects (e.g. privacy law, contract law, the law of armed conflict).[178] As a result they pay less attention to either the interrelation or interaction of these issue areas.

On the one hand, there is some sense in taking a piecemeal approach: national expertise bodies and international institutions alike often have clear, predefined areas of expertise.[179] Yet, on the other hand, this approach limits our ability to understand the sources of AI's symptomatic challenges.[180] Lacking an

---

[174] Among other critiques of prominent governance proposals. See also Neel Guha and others, 'AI Regulation Has Its Own Alignment Problem: The Technical and Institutional Feasibility of Disclosure, Registration, Licensing, and Auditing' (2024) 92 George Washington Law Review 1473.

[175] Ryan Calo, 'The Scale and the Reactor' (9 April 2022) <https://papers.ssrn.com/abstract=4079851> accessed 23 May 2023.

[176] Ryan Calo, 'Robotics and the Lessons of Cyberlaw' (2015) 103 California Law Review 513, 549.

[177] Rebecca Crootof and BJ Ard, 'Structuring Techlaw' (2021) 34 Harvard Journal of Law & Technology 347, 348–349. See also the discussion in Chapter 4.

[178] ibid 1 ('Technological breakthroughs challenge core legal assumptions and generate regulatory debates. Practitioners and scholars usually tackle these questions by examining the impacts of a particular technology within conventional legal subjects—say, by considering how drones should be regulated under privacy law, property law, or the law of armed conflict. While individually useful, these siloed analyses mask the repetitive nature of the underlying questions and necessitate the regular reinvention of the regulatory wheel').

[179] Morin and others (n 90).

[180] See Margot E Kaminski, 'Authorship, Disrupted: AI Authors in Copyright and First Amendment Law' (2017) 51 UC Davis Law Review 589; Léonard van Rompaey, 'Discretionary Robots: Conceptual Challenges in the Legal Regulation of Machine Behaviour' (University of Copenhagen 2020).

understanding of when, or why, AI technology actually translates into sufficiently new or relevantly problematic societal changes renders AI governance efforts vulnerable to capture by particularly visceral or emblematic use cases that may, however, also be edge cases that do not (fully) represent the ways that AI affects society. The costs of this can be key: more widespread but less visible or visceral societal changes remain unaddressed. Alternatively, this approach may focus attention on isolated or demonstrated AI harms, without comparing these harms to the baseline of activities that are already feasible without AI, thereby failing to properly account for the actual 'marginal risk' of these technologies,[181] undercutting meaningful assessments of their novelty as well as their benefits and costs. Finally, this siloed approach creates the risk that governance responses may, at best, duplicate one another, and, at worst, work at cross-purposes.[182]

AI governance proposals should be grounded in a better understanding of the vectors of cross-sector change, of how new capabilities or developments might change the array or portfolio of problems that need addressing, and of how to reckon with the increasingly cross-domain impacts of multi-modal foundation models and the general-purpose AI systems they support. It is valuable therefore to approach AI governance through the lens of '*sociotechnical change*',[183] a concept which focuses on when, how, and why changes in technologies actually expand human capabilities in ways that give rise to new activities, ways of being, or new types of relationships.[184] Or, as I will roughly define the concept:

> *Sociotechnical change*: the processes by which changes in technologies extend the capabilities of some or all actors—allowing them new ways of carrying out old conduct or enabling new types of conduct altogether—in ways that (1) can have unanticipated disruptive impacts on society, which (2) may give rise to one or another *rationale* for regulation; and which (3) shape the particular texture and problem logics of the regulatory *target* for those regulatory interventions.

This framework facilitates a better examination of when AI-enabled behaviour actually creates new problems or negative externalities that may provide precise

---

[181] Sayash Kapoor and others, 'On the Societal Impact of Open Foundation Models' <https://crfm.stanford.edu/open-fms/>; Harry Law, 'The Marginal Risk of AI' (*Learning from Examples*, 30 March 2024) <https://www.learningfromexamples.com/p/the-marginal-risk-of-ai?publication_id=1838544> accessed 7 May 2024; Amanda Askell, 'In AI Ethics, "Bad" Isn't Good Enough' (*Amanda Askell*, 14 December 2020) <https://askell.io/posts/2020/12/bad-isnt-good-enough> accessed 3 May 2024.

[182] Turner (n 12) 218–221.

[183] See Lyria Bennett Moses, 'Recurring Dilemmas: The Law's Race to Keep Up With Technological Change' (2007) 21 University of New South Wales Faculty of Law Research Series <http://www.austlii.edu.au/au/journals/UNSWLRS/2007/21.html> accessed 3 July 2018; Lyria Bennett Moses, 'Regulating in the Face of Sociotechnical Change' in Roger Brownsword, Eloise Scotford and Karen Yeung (eds), *The Oxford Handbook of Law, Regulation, and Technology* (Oxford University Press 2017).

[184] This draws on the account in Lyria Bennett Moses, 'Why Have a Theory of Law and Technological Change?' (2007) 8 Minnesota Journal of Law, Science & Technology 589, 591–592.

rationales for governance—and how that technology is consequently best approached as a target for governance.

Governance Change: How AI Disrupts Global Governance

In the second place, those who study or craft AI governance must consider with greater precision *how AI technology changes the doctrinal applicability, tools, processes and assumptions of (global) governance* itself.

In recent years, numerous governance proposals for AI have been tabled. Many of these, however, are relatively reactive and static, responding to one or another prominent form or use of the technology at a particular moment, without anticipating how future changes may result in renewed legal uncertainty or disruption. In addition, such proposals often fail to factor in technology-driven changes to the processes and tools of governance itself, or even changes to the very conditions for governance. However, again, most proposals for the global governance of AI still reason in relative isolation from actually legally disruptive uses of AI, and their impacts on (international) law,[185] which is a problem because, as noted by Colin Picker, new technological innovations have a long history of driving the 'creation, modification, or destruction of international law, or the derailment of the creation of new international law'.[186] This one-way approach underappreciates both important challenges and opportunities for the global governance of AI.

In particular, AI governance should be based on a better account of three related but distinct dynamics of legal disruption: (1) how technological change (in AI) will create ambiguities or conflicts in existing laws or regimes, in ways that demand the ongoing *development* of international law and governance to remain clear and effective; (2) how new (AI) technologies can support the automation of various processes within international law and governance, resulting in (beneficial or adverse) *displacement* of older legal instruments or norms; and (3) how new (AI) technologies may drive shifts in the political foundations for international law, resulting in the erosion or even the *destruction* of regimes.

To be sure, independent bodies of work exist, within fields such as technology law ('TechLaw')[187] that have explored these questions in turn. After all, the idea that new technologies are not just rationales, objects, and targets for regulation, but can also change the operation and processes of existing legal systems (and even

---

[185] Though for an exception, see Allan Dafoe and others, 'Cooperative AI: Machines Must Learn to Find Common Ground' (2021) 593 Nature 33.

[186] Picker (n 49) 156.

[187] BJ Ard and Rebecca Crootof, 'The Case for "Technology Law"' (*Nebraska Governance & Technology Center*, 16 December 2020) <https://ngtc.unl.edu/news/case-technology-law/> accessed 16 March 2021; Crootof and Ard (n 177).

the goals of regulators), is hardly new. Reflections on what a given new technology reveals about the changing face of law have extensive precedent. For instance, in early debates in the field of cyberlaw, Lawrence Lessig famously examined legal questions involving the new technology of cyberspace, not merely to discuss the relative efficacy of different approaches to regulating certain issues (e.g. content zoning or copyright) on the internet, but rather to ground and illustrate his broader reflections on the changing nature and workings of the different 'regulatory modalities' of laws, norms, markets, and architectures ('code').[188] Likewise, Roger Brownsword has used studies of behaviour-shaping technologies and geo-engineering in order to reflect, respectively, upon the rising role of non-normative 'technological management' and on 'regulatory responsibilities' for the core global commons.[189] AI technologies fit naturally within this tradition: after all, the internet 'merely' led to the informatization of infrastructure and public space, in ways that have already significantly altered how regulation operates. In turn, AI systems may enable the increasing 'intelligentization' or 'cognitization' of these infrastructures,[190] suggesting that this technology may in time have an impact on the practices and dynamics of law and governance that is even more far-reaching than any seen before.

As noted, the current field of AI governance has not yet considered in depth the dynamics of how technological change can disrupt existing legal systems.[191] Rather, the focus has still largely been on establishing institutions or regimes in the first place. While some AI policy proposals have begun to signal a greater awareness of the importance of making regulation more resilient or future-proof,[192]

---

[188] Lawrence Lessig, 'The Law of the Horse: What Cyberlaw Might Teach' (1999) 113 Harvard Law Review 501; Lawrence Lessig, *Code: And Other Laws of Cyberspace, Version 2.0* (2nd rev. edn, Basic Books 2006) <http://codev2.cc/download+remix/Lessig-Codev2.pdf> accessed 25 February 2025.

[189] Roger Brownsword, *Law, Technology and Society: Re-Imagining the Regulatory Environment* (1st edn, Routledge 2019); Roger Brownsword, 'Law and Technology: Two Modes of Disruption, Three Legal Mind-Sets, and the Big Picture of Regulatory Responsibilities' (2018) 14 Indian Journal of Law and Technology 1.

[190] This plays on terminology used by the Chinese PLA. See Elsa Kania, 'Chinese Military Innovation in Artificial Intelligence' (Testimony before the U.S.-China Economic and Security Review Commission Hearing on Trade, Technology, and Military-Civil Fusion, 7 June 2019) <https://www.uscc.gov/sites/default/files/June%207%20Hearing_Panel%201_Elsa%20Kania_Chinese%20Military%20Innovation%20in%20Artificial%20Intelligence_0.pdf> accessed 1 June 2025.

[191] Previously set out in Matthijs M Maas, 'International Law Does Not Compute: Artificial Intelligence and the Development, Displacement or Destruction of the Global Legal Order' (2019) 20 Melbourne Journal of International Law 29. A version of this model is also developed at greater length in Hin-Yan Liu and others, 'Artificial Intelligence and Legal Disruption: A New Model for Analysis' (2020) 12 Law, Innovation and Technology 205.

[192] Anson Ho, 'Future-Proof: Monitoring the Development, Deployment, and Impacts of Artificial Intelligence' (2023) 22 Journal of Science Policy & Governance <http://www.sciencepolicyjournal.org/article_1038126_jspg220305.html> accessed 6 November 2023; Helen Toner and others, 'Skating to Where the Puck Is Going: Anticipating and Managing Risks from Frontier AI Systems' (Center for Security and Emerging Technology 2023) <https://cset.georgetown.edu/publication/skating-to-where-the-puck-is-going/> accessed 1 December 2023; Guillem Bas and others, 'The EU AI Act: A Pioneering Effort to Regulate Frontier AI?' (2024) 27 Inteligencia Artificial 55; See generally Maxime

these proposals often lack a strong account of how future technological changes will create uncertainty for law and governance, or of the different avenues by which legal regimes might respond.

Second, AI governance needs to better reckon with the ways in which AI tools may be integrated into international legal lawmaking processes. In other fields, there has also been plenty of work exploring the uses and risks of legal automation—such as the use of AI and other digital technologies in legal systems,[193] in domestic administrative law,[194] and in the general technological management of citizen behaviour.[195] A small but growing body of work even has recently begun to explore this phenomenon at the level of international law,[196] including work on the use of AI tools in support of international criminal courts,[197] or arms control monitoring and verification.[198] However, such work continues to be somewhat underexplored within proposals for the governance of AI technology.

Finally, the AI governance field needs to account for the ways in which new uses of AI systems may change basic geopolitical conditions—affecting the balance and distribution of global power[199]—in ways that may lead to the erosion or even destruction of the political conditions for governance more generally. Recent work suggests that digital technologies may have a range of disruptive effects on the

---

Stauffer and others, 'The FAIR Framework—A Future-Proofing Methodology' (*Simon Institute for Longterm Governance*, 26 April 2023) <https://www.simoninstitute.ch/blog/post/the-fair-framework-a-future-proofing-methodology/> accessed 3 November 2023.

[193] This literature is extensive. But see for instance Roger Brownsword, 'In the Year 2061: From Law to Technological Management' (2015) 7 Law, Innovation and Technology 1; Karen Yeung, '"Hypernudge": Big Data as a Mode of Regulation by Design' (2017) 20 Information, Communication & Society 118; Benjamin Alarie, 'The Path of the Law: Towards Legal Singularity' (2016) 66 University of Toronto Law Journal 443; Anthony J Casey and Anthony Niblett, 'Self-Driving Laws' (2016) 66 University of Toronto Law Journal 429; Christopher Markou and Simon Deakin, 'Is Law Computable? From Rule of Law to Legal Singularity' in Christopher Markou and Simon Deakin (eds), *Is Law Computable? Critical Perspectives on Law + Artificial Intelligence* (Hart 2020) <https://papers.ssrn.com/abstract=3589184> accessed 15 May 2020; Brian Sheppard, 'Warming Up to Inscrutability: How Technology Could Challenge Our Concept of Law' (2018) 68 University of Toronto Law Journal 36; Christoph Winter, Nicholas Hollman, and David Manheim, 'Value Alignment for Advanced Artificial Judicial Intelligence' (2023) 60 American Philosophical Quarterly 187.
[194] Karen Yeung and Martin Lodge (eds), *Algorithmic Regulation* (Oxford University Press 2019).
[195] Brownsword, *Law, Technology and Society: Re-Imagining the Regulatory Environment* (n 189).
[196] Burri (n 48); Ashley Deeks, 'High-Tech International Law' (2020) 88 George Washington Law Review 575. Thomas F McInerney, 'The Emergence of Intelligent Treaty Systems and the Future of International Law' [2022] University of Illinois Journal of Law, Technology & Policy 259. See also the more detailed discussion of this work in Chapter 5.
[197] Gabrielle McIntyre and Nicholas Vialle, 'The Use of AI at the ICC: Should We Have Concerns? Part I' (*Opinio Juris*, 11 October 2023) <https://opiniojuris.org/2023/10/11/the-use-of-ai-at-the-icc-should-we-have-concerns-part-i/> accessed 15 November 2023.
[198] Niklas Schörnig, 'AI for Arms Control: How Artificial Intelligence Can Foster Verification and Support Arms Control' (Peace Research Institute Frankfurt 2022) <https://www.hsfk.de/publikationen/publikationssuche/publikation/ai-for-arms-control> accessed 23 August 2022; Thomas Reinhold and Niklas Schörnig, *Armament, Arms Control and Artificial Intelligence: The Janus-Faced Nature of Machine Learning in the Military Realm* (Springer Nature 2022).
[199] Schmidt (n 160).

continued viability of legal regimes, in domains such as international security governance,[200] arms control,[201] or the environment.[202]

AI technology is likely to affect not just the substance of various international regimes, but also their processes, and potentially even their political scaffolding. It is unlikely that these disruptive effects will plague the governance of other domains without affecting the global regulation of AI. Rather, they will also affect the base conditions and actual efficacy of the very governance instruments being proposed to regulate AI, in ways that will render particular strategies more or less viable or attractive. To boot, in this book I will approach the concept of governance disruption in a way that operationalizes it as:

> *Governance disruption*: a form of sociotechnical change by which new technologies and their resulting changes in artefacts, actors or activities challenge or transform existing systems of law or governance, at the level of (1) their substance, doctrine or assumptions (requiring the urgent development of those laws), (2) their processes of law-creation, adjudication, monitoring, or enforcement (resulting in the potential displacement of those practices through automation or wholesale substitution), or (3) their political foundations (resulting in their potential erosion or even destruction).

Understanding how and where AI can disrupt systems of governance, what kinds of problems or uncertainties this creates, and how legislators can or should adequately respond are all key challenges to address.[203] Scholars and legislators alike will need to reckon with these dynamics. This is not just because they matter for how to design appropriate AI regulations and institutions, but also because they reveal interesting and important lessons about the changing nature of contemporary technology governance specifically, and of twenty-first-century international law in general.

Regime Complexity: The Changing Environment for AI Governance

Third, this field must better consider *how AI governance regimes are shaped by underlying changes in the broader global governance environment*. At present, many

---

[200] Amandeep S Gill, 'The Changing Role of Multilateral Forums in Regulating Armed Conflict in the Digital Age' (2020) 102 International Review of the Red Cross 261.

[201] Nelson (n 112); Giacomo Persi Paoli and others, 'Modernizing Arms Control: Exploring Responses to the Use of AI in Military Decision-Making' (UNIDIR 2020) <https://unidir.org/publication/modernizing-arms-control> 25 February 2025. See generally, Richard Danzig, 'An Irresistible Force Meets a Moveable Object: The Technology Tsunami and the Liberal World Order' (2017) 5 Lawfare Research Paper Series <https://assets.documentcloud.org/documents/3982439/Danzig-LRPS1.pdf> accessed 1 September 2017.

[202] Peter Dauvergne, 'The Globalization of Artificial Intelligence: Consequences for the Politics of Environmentalism' [2020] Globalizations 1.

[203] Crootof and Ard (n 177).

proposals for AI governance take their cue from past multilateral treaties, or previously established international institutions. For instance, recent years have seen calls for a 'Digital Geneva Convention';[204] diverse calls for an international ban on LAWS inspired by global bans on blinding lasers and anti-personnel mines;[205] and a plethora of recent calls for new international institutions for AI that have frequently invoked the IPCC, the IAEA, or other models such as the ICAO, IMO, or FATF.[206] Yet beyond the question of which of these designs would be most substantively appropriate for the diverse bucket of challenges created by AI, it is also unclear how well-aligned such designs are with the broader context and trajectory of the global governance system, which is itself undergoing considerable historical change.

Indeed, in many other domains, international institutions have exhibited a diverse ecology and patterns of evolution over time, as organizational missions shift and as other fora emerge to address distinct aspects of a given governance challenge. As such, while it is certainly valuable to understand and draw up 'ideal' global governance blueprints for AI,[207] in practice the establishment and operation of any AI governance processes will not occur in a vacuum. Indeed, Karen Alter and Kal Raustiala have argued that new cooperative efforts in global governance necessarily occur within a complex and dense global governance architecture. As such, they argue that all too often,

> [g]lobal governance solutions . . . must take one of two approaches: (a) International actors can attempt to create an encompassing regime that can address all dimensions of the problem, or (b) international actors can accept that policy solutions will be crafted, coordinated, and implemented within a larger regime complex . . . although the first option might be more efficient and effective, it is rarely the solution adopted.[208]

---

[204] Brad Smith, 'The Need for a Digital Geneva Convention' (*Microsoft on the Issues*, 14 February 2017) <https://blogs.microsoft.com/on-the-issues/2017/02/14/need-digital-geneva-convention/> accessed 17 April 2019.

[205] Human Rights Watch, 'Precedent for Preemption: The Ban on Blinding Lasers as a Model for a Killer Robots Prohibition: Memorandum to Convention on Conventional Weapons Delegates' (Human Rights Watch 2015) <https://www.hrw.org/news/2015/11/08/precedent-preemption-ban-blinding-lasers-model-killer-robots-prohibition> accessed 28 April 2017. But for critiques of the accuracy of this comparison, see Rosert and Sauer (n 166); Rebecca Crootof, 'Why the Prohibition on Permanently Blinding Lasers Is Poor Precedent for a Ban on Autonomous Weapon Systems' (*Lawfare*, 24 November 2015) <https://www.lawfaremedia.org/article/why-prohibition-permanently-blinding-lasers-poor-precedent-ban-autonomous-weapon-systems> accessed 7 September 2020; Rebecca Crootof, 'The Killer Robots Are Here: Legal and Policy Implications' (2015) 36 Cardozo Law Review 1837.

[206] Maas and Villalobos (n 65); Ho and others (n 28).

[207] Dafoe (n 24) 48–51 (emphasizing the need for research into 'values and principles', 'institutions and mechanisms', and 'positive visions').

[208] Alter and Raustiala (n 89) 337.

To bear this out, the past few decades have been marked by significant changes in the global institutional landscape. These include patterns of institutional proliferation;[209] the ongoing fragmentation of international law, resulting in complex inter-regime impacts and externalities;[210] and growing patterns of contested multilateralism.[211] Others have identified trends of legal stagnation and have argued that, in areas such as international environmental governance or cyberspace,[212] global governance is increasingly marked by a shift towards informality. Among other things, this underscores that fact that the landscape in which AI governance finds itself today differs in important ways from the landscape of the past, when previous landmark institutions or regimes were established.

All of this is not to suggest that historical lessons can no longer be valuable,[213] nor that governance for AI must do away with everything old and blithely adopt the new. As discussed above, it is certainly not the case that the role (or rule) of traditional international law in global affairs has been eclipsed.[214] Nonetheless, many traditional, formal international institutions might be increasingly ill-suited, or uneasily positioned, to address complex cross-sectoral issues such as AI.[215] As a result, it is important to consider how well any proposed AI governance mechanisms, institutions, or arrangements might slot into, and interact with, the complex, evolving architecture of global governance.[216]

To study how AI governance will be shaped by the trends, trajectories, and changes in the global governance architecture, this book draws on the extensive scholarship in global governance studies at the intersection of international law

---

[209] Kal Raustiala, 'Institutional Proliferation and the International Legal Order' in Jeffrey L Dunoff and Mark A Editors Pollack (eds), *Interdisciplinary Perspectives on International Law and International Relations: The State of the Art* (Cambridge University Press 2012).

[210] Martti Koskenniemi and Study Group of the International Law Commission, 'Fragmentation of International Law: Difficulties Arising from the Diversification and Expansion of International Law' (United Nations—General Assembly 2006) A/CN.4/L.682 <http://legal.un.org/ilc/documentation/english/a_cn4_l682.pdf> 25 February 2025; Frank Biermann and others, 'The Fragmentation of Global Governance Architectures: A Framework for Analysis' (2009) 9 Global Environmental Politics 14.

[211] Julia C Morse and Robert O Keohane, 'Contested Multilateralism' (2014) 9 The Review of International Organizations 385; Amitav Acharya, 'The Future of Global Governance: Fragmentation May Be Inevitable and Creative Global Forum' [2016] Global Governance 453; Michael Zürn, 'Contested Global Governance' (2018) 9 Global Policy 138.

[212] J Pauwelyn, RA Wessel, and J Wouters, 'When Structures Become Shackles: Stagnation and Dynamics in International Lawmaking' (2014) 25 European Journal of International Law 733. On cyberspace governance, see also Joseph S Nye, 'The Regime Complex for Managing Global Cyber Activities' (Global Commission on Internet Governance 2014) 1 <https://dash.harvard.edu/bitstream/handle/1/12308565/Nye-GlobalCommission.pdf> accessed 3 September 2019.

[213] Indeed, they can at least provide precedent about the plausibility of achieving governance even on contested, high-stakes technologies. See Maas, 'How Viable Is International Arms Control for Military Artificial Intelligence?' (n 165).

[214] See Karen J Alter, 'The Future of International Law' (2017) 101 iCourts Working Paper Series <https://papers.ssrn.com/abstract=3015177> accessed 11 June 2020; Benvenisti and Downs (n 84).

[215] Morin and others (n 90).

[216] Tallberg and others (n 47).

and international relations.[217] In the context of this book's approach to global AI governance, I will in particular focus on the concept of 'regime complexity',[218] which, drawing on existing articulations, I define as:

> *Regime Complexity*: the governance architecture on a given issue area, consisting of (1) at least three international or transnational institutions or actors that (2) jointly address a specific issue area, (3) whose mandates, functions, and/or memberships overlap (4) in ways that may result in various beneficial or problematic interactions; this complex consequently (5) has significant implications for the efficacy, resilience or coherence of governance functions by its actors, as well as the prospects for establishing new institutions.

Theoretically, the regime complexity lens enables us to engage in a more refined discussion of the development and prospects of AI governance regimes, in terms of the specific origins of individual institutions, the topology of the regime complex, potential drivers of regime architecture towards fragmentation or integration, the consequences of either trajectory, and strategies to mitigate adverse consequences and improve the efficacy, coherence, and resilience of the regime.

For this, the framework of regime complexity provides a valuable tool for debates around AI governance. At present, it remains unclear whether AI will be effectively governed, or if recent legal initiatives might fizzle out, leaving the topic in a non-regime state of limbo or gridlock. If some lasting governance system does emerge in the coming years, it is unclear whether this will continue to take the shape of decentralized or fragmented regimes, or whether an effective centralized treaty or institution may emerge in time. We will need better ways not only of understanding the relative likelihood of these scenarios, but also of how best to conduct ourselves within each. In turn, such a study also provides a valuable examination to the literature on regime complexity, which has mostly focused on retroactive studies of already developed and mature regime complexes—by allowing the analysis of the emergence and early evolution of an emerging regime complex.[219]

On This Book

This book explores the conceptual and practical implications of technological, legal and political change for global AI governance. In it, I argue that to effectively

---

[217] See generally Wayne Sandholtz and Christopher A Whytock, *Handbook on the Politics of International Law* (Edward Elgar Publishing 2017); Jeffrey L Dunoff and Mark A Pollack, 'Reviewing Two Decades of IL/IR Scholarship' in Jeffrey L Dunoff and Mark A Pollack (eds), *Interdisciplinary Perspectives on International Law and International Relations* (Cambridge University Press 2012).

[218] Alter and Raustiala (n 89).

[219] Tallberg and others (n 47); Johannes Geith, Magnus Lundgren and Jonas Tallberg, 'The Emerging Regime Complex for Artificial Intelligence' [2025 forthcoming] *Global Studies Quarterly*.

adapt or 'scale' law and governance for the global challenges of increasingly advanced AI in ways that account for these structural changes, it will be necessary to draw on a broader analytical toolkit.

The Core Argument

The core argument of this book is that, in approaching the global governance of AI, we need to think about three facets of structural change, including:

(1) Changes in AI capabilities which drive *sociotechnical changes* in global society, and how these provide distinct regulatory rationales and targets for global governance instruments;

(2) Changes in AI capabilities which *drive disruption* (doctrinal, substantive, procedural, or political) in the instruments of international law and global governance;

(3) Changes in the *broader governance* architecture that encompasses any specific AI governance regime.

Individually, any one of these three topics warrants close consideration in any studies or proposals for global AI governance. Taken together, they provide a set of foundational considerations with which AI governance scholars and policymakers alike must engage, if they are to reckon effectively with this technology. By providing broader frameworks within which to consider specific AI issues for (international) law, we can gain a deeper understanding of the real issues at stake, the likely effectiveness of different responses, and shared themes across regimes dealing with nominally different issue areas.[220]

In developing this argument, I follow a broad 'pragmatic, analytically eclectic, tool-kit approach',[221] one that draws on a range of theories, models, and cases from adjacent disciplines and fields. I draw connections between three existing bodies of work—on law and sociotechnical change; on technology-driven legal disruption, and on regime complexity theory—and apply these to pressing problems in the field of advanced AI governance.[222] In discussing sociotechnical change and

---

[220] See also Crootof and Ard (n 177) 354 ('being able to situate a particular techlaw question within a broader framework helps us better understand the fundamental issues, better evaluate the likely effectiveness of different legal responses, and better conduct tech-specific and subject-specific analyses').

[221] Dunoff and Pollack (n 217) 653; cf. Andrea Bianchi, *International Law Theories: An Inquiry into Different Ways of Thinking* (Oxford University Press 2016) 125. On the role of 'eclectic theory' in the study of international relations, see generally Peter Katzenstein and Rudra Sil, 'Eclectic Theorizing in the Study and Practice of International Relations' in Christian Reus-Smit and Rudra Sil (eds), *The Oxford Handbook of International Relations* (2008).

[222] Having outlined these contributions, it is also important to clarify what this book does not aim to do. First, while the three conceptual lenses have many contact points and are commensurable, the aim is not to provide a comprehensive or unified framework that must necessarily be adopted in its entirety.

governance disruption, I draw on recent scholarship in various strands of technology law.[223] In discussing and applying the concept of regime complexity, this book gains from the extensive work of global governance and international regime scholars working at the intersection of international law and international relations.

As such, this book diverges from the usual case-study approach. Rather than analysing a few cases in exhaustive detail, I draw instead on a broad review of current work on AI and its global governance, along with case studies from adjacent fields in technology law and global governance for various issues. The aim is not to understand a few examples in depth, but rather to understand the phenomena of AI governance in a wide perspective: to investigate broad patterns in how AI technologies challenge, disrupt, or affect global governance.

The Aims and Contributions

This book has three aims. I seek: (1) to contribute to the practice of global AI policymaking; (2) to contribute to a deeper and more informed academic study of AI governance going forward; and (3) to contribute in turn to scholarship on technology law and global governance.

My first aim in this book, accordingly, is to contribute to improved decision-making around the global governance of advanced AI, by helping inform debates over the benefits or drawbacks of currently proposed governance solutions. Whether or not the present work leads directly to conclusive outcomes, it promises at least to enable better conversations among different approaches and policy proposals. Further conceptual clarity will help AI governance researchers and practitioners consider and design governance architectures for AI which are more effective, coherent, legitimate, and resilient. Rather than necessarily entailing abstract theoretical debate of little practical utility, exercises of conceptual clarification and cross-fertilization have, at their best, been fundamental to much past work at the intersection of law, regulation, and technology.[224] For instance, Colin

---

Rather, the goal is to provide a set of exploratory perspectives and tools on AI governance which, taken individually, can help inform more nuanced, targeted and resilient governance strategies for AI. One need not accept or adopt all three lenses at the same time but can simply focus productive analysis on one among them. Second, this is not meant to be a detailed or exhaustive application of each of these lenses. Rather, it is intended as an initial exploration to highlight their promises and limits, as a demonstration of their promise, and as an indication that they warrant further exploration and application.

[223] Among others, see Roger Brownsword, Eloise Scotford, and Karen Yeung (eds), *The Oxford Handbook of Law, Regulation and Technology* (Oxford University Press 2017); Bennett Moses, 'Why Have a Theory of Law and Technological Change?' (n 184); Crootof and Ard (n 177).

[224] For instance, in reflecting on these topics, Roger Brownsword and others have noted that 'debates over these terms, and about the conceptualization of the field or some parts of it, can significantly contribute to our understanding'. Brownsword, Scotford, and Yeung (n 173) 6.

Picker has noted that a high-level view of the strategic dynamics between technology and international law can be critical for policymakers:

> [e]ven though policy makers must be closely concerned with the 'nitty gritty' of their international regimes and negotiations, . . . [they] have much to gain from taking a macro or holistic view of the issues raised by technology. Macro-examinations can provide larger theoretical understandings and can reveal previously hidden characteristics that are simply not discernible from the 'trenches'. Viewing technology from '40,000 feet up' reveals certain patterns, pitfalls, and lessons for policy.[225]

Following Picker, and other scholars, this book aims to link a high-level view of (AI) technology and global governance with practical, actionable questions. It sets out the theories and concepts used in technology governance, illustrates them with historical and contemporary cases, and finally elicits their most relevant lessons for current AI governance debates. Ideally, this exercise will contribute to greater strategic clarity and coherence in the field, improving the quality of interventions and refining policy debates on these important and urgent issues.[226]

To be clear, the analysis in this book is not meant to provide strong predictions about the direction of governance, but rather to sketch a series of possible scenarios and avenues. Present pervasive uncertainty over the future of AI governance only renders it all the more important that we explore various scenarios and trajectories as well as their strategic and normative implications. Even if one considers it unlikely that a centralized governance regime will emerge for AI, should circumstances shift in the coming years such that such a model suddenly becomes viable, then it will be of particular importance to ensure that this regime is well designed and fit for purpose, since poor institutional design could lock in catastrophic outcomes. Conversely, in a fragmented AI regime complex consisting of many parallel and overlapping institutions, one should expect to see clusters and various (conflictive or cooperative) interactions between different institutions. The intent of my analysis is to help navigate such choppy questions, by exploring dependencies and implications today.

---

[225] Picker (n 49) 151–152.

[226] In doing so, it is important to distinguish between three potential aims or goals for the field and epistemic community: (1) 'strategic clarity', achieving a sensible and grounded theory of change on the basis of a detailed and robust model of both the technical landscape and the policy world around advanced AI, with a resulting roadmap for how to select, evaluate, or prioritize present-day or near-term interventions; (2) 'strategic consensus', where (almost) everyone in a given epistemic community shares this same perspective or judgment; and (3) 'strategic coherence', when policy interventions or initiatives by different individuals or subcommunities in the field do not interfere with, counter, or erode one another (even if there remains underlying disagreement). Notably, while basic strategic clarity is invaluable for formulating robustly beneficial policies for advanced AI, it is unclear whether outright strategic consensus is always necessary or desirable, as a portfolio approach of many actors with different views (i.e. coherence, but lacking consensus) may be preferable.

Second, this book aims to contribute to a deeper and more informed academic study of future AI governance. By making systematic use of concepts, theories, and insights from other areas of scholarship in technology law and global governance, we can gain a better understanding of key dimensions of global governance for AI.[227] Such conceptual clarity will help improve our analysis. In doing so, it will help us answer previously neglected questions and highlight potentially fruitful avenues for further research. For instance, it can complement descriptive research on the global governance of AI with more explanatory insights.[228] My intent is to consolidate and disentangle some of this field, and to promote greater clarity about underexplored questions or conspicuous gaps in the literature. In addition, by providing a detailed and structured account of developments and problems in AI governance, it should also provide an accessible starting point for those entering this diverse field.

Third, this book aims to contribute to its own underpinning fields of scholarship—in technology law, and international law and international relations—by applying and testing the fit and limits of their frameworks within the new sociotechnical contexts of AI development and governance. Doing so can offer a great opportunity to further develop these concepts and theories in turn.[229] For instance, to technology law scholarship, it contributes a more detailed exploration of how dynamics of technological governance disruption play out not just at the domestic level but also at the international level. It also contributes to the scholarship on regime complexity by applying these frameworks to a new issue domain, AI, that has not been widely studied in that field.[230] In doing so, it also helps in the exploration of dynamics of regime complex creation. After all, while there is extensive research on mature regime complexes, there is to date only limited research on the creation and evolution of regime complexes as these processes unfold at an early stage: the case of AI governance offers a valuable opportunity to track the evolution of such a regime complex from an early stage.[231] Moreover, by considering the phenomenon of AI-driven governance disruption, this book also aims to contribute new insight into long-running theoretical debates over the relative role of rationalist (interest-based), constructivist (norms-based), and a set of

---

[227] Tallberg and others (n 47) 3 (arguing that 'making more systematic use of social science concepts and theories will help us to gain a better understanding of various dimensions of the global governance of AI').

[228] ibid 9 ('Thus far, existing research on the global governance of AI has been primarily concerned with descriptive tasks and largely fallen short in engaging with explanatory questions').

[229] ibid 3 ('exploring AI as a regulatory issue offers a critical opportunity to further develop concepts and theories of global governance, as they confront the particularities of regulatory dynamics in this important area').

[230] See also Geith, Lundgren, and Tallberg (n 219).

[231] Tallberg and others (n 47). Another example of such a study is found in studies of the emerging regime complex for marine-biodiversity governance. Arne Langlet and Alice Vadrot, 'Negotiating Regime Complexity: Following a Regime Complex in the Making' (2024) 50 Review of International Studies 231.

re-emerging 'material' (that is, artefactual or architectural) factors in shaping the constitution and maintenance of global governance systems.[232]

Structure of the Book

This book is composed of seven chapters, divided across three parts.

Part I (Chapters 1–3) provides a comprehensive overview and discussion of the foundations for advanced AI governance.

Chapter 1 discusses recent developments in AI technology and their stakes. To understand why and how to regulate AI, we first need to understand the how we got here—and where these trends may go. It draws on recent technical research into the strengths, limits, and use preconditions of AI. It argues that while this technology is today still immature in important ways, it is already functional enough for wide adoption, which will necessarily entail societal impacts. Rather than fall into dichotomous debates over whether the promises of AI are exaggerated, the chapter argues that AI is best considered a high-variance technology, with some applications that may still fall short of promises, but where a small share of peak systems is increasingly positioned to drive far-reaching societal impact. I then explore arguments and uncertainties around how this importance could develop in the coming years, considering whether, how, or under what conditions we could expect change (in technological capabilities, applications, and societal impacts) to derive from either further progress in frontier AI systems, or from straightforward algorithmic proliferation. On this basis, I conclude that, even under excessively conservative assumptions about further fundamental progress, AI's global impact will likely be extreme before the end of the decade.

Chapter 2 explores foundational questions for global AI governance. It first argues that global governance is needed for many of these issues. It sketches the range of AI's challenges, distinguishing between the three focus areas of conventional AI, military AI, and advanced AI governance. It then draws on typologies of global public goods to argue that many of these issues may gain from, or even require, some form of global cooperation. It then argues that effective global regulation is indeed feasible. Drawing on models of transnational regulation of digital markets, as well as novel accounts of compute governance, it makes the case that enforceable policy levers already exist by which to regulate AI. This implies that if global agreements to regulate AI were reached, they would be enforceable. The

---

[232] On the interrelation of these three accounts, and their different historical roles and influence within the field of International Relations, see also broadly Daniel Deudney, 'Turbo Change: Accelerating Technological Disruption, Planetary Geopolitics, and Architectonic Metaphors' (2018) 20 International Studies Review 223. See also Nathan Alexander Sears, 'International Politics in the Age of Existential Threats' [2020] Journal of Global Security Studies 1. See also Chapter 4.

key hurdle instead may be achieving adequate global cooperation and agreement about such regulations in the first place.

Chapter 3 surveys the history of AI governance initiatives over the past decade and a half, reviewing how global governance for military, conventional, and advanced AI systems has proceeded in distinct tracks for each. Specifically, it reviews various influential moments, landmark initiatives and instruments, and discusses the strengths and limits of the resulting AI governance architecture. It then considers a variety of proposals for the way forward, from several perspectives. It concludes that in spite of promising steps, AI governance remains incipient and fragmented today, and there is pervasive uncertainty about the best paths forward.

Throughout Part II (Chapters 4–6), the book explores three facets of change in the AI governance architecture, with each chapter introducing, applying, and exploring lessons from a distinct conceptual lens.

Chapter 4 argues that AI governance should focus more on sociotechnical change, to explore how and when governance should relate to the constantly changing array of capabilities and uses of AI. It distinguishes six situations in which AI-enabled behaviours become relevant *rationales* for global governance. It then discusses how such societal impacts of AI can be best approached as governance *targets*, by distinguishing six types of 'problem logics' that have similar regulatory features across distinct AI architectures or application domains.

Chapter 5 provides an account of governance disruption, addressing how and why AI can produce intended or unintended change in the global governance system itself. It explores three main varieties of potential governance disruption by AI: (1) Under governance development, the sociotechnical changes produced by AI result in situations of substantive legal uncertainty or ambiguity, requiring change in governance to address the new challenges. (2) Under governance displacement, AI technologies can support—or even substitute for—key processes or practices of the global legal order, from rule creation or adjudication to compliance monitoring. (3) Under scenarios of governance destruction, the use of AI can erode the effectiveness, legitimacy, or coherence of the global legal order as such.

Chapter 6 adopts the lens of regime complexity to better understand how overarching changes in the global institutional ecology—whether exogenous or AI-induced—will affect AI governance. The chapter provides a background on regime theory, while previewing some of the debates around the consequences and desirability of the fragmentation or centralisation of a regime complex. Finally, it proposes that the regime complexity lens facilitates exploration of a governance regime at five key levels or stages of analysis.

Part III, the final section of the book, draws these three lenses together to set out frameworks for choice in the face of technological change. Chapter 7 integrates the three lenses to explore key conceptual, strategic, and political questions in AI governance. It argues that these lenses offer insight into five questions that can be asked of diverse existing, emerging, or proposed AI governance systems, in terms

of (1) origins of an AI regime, (2) normative and institutional topology, (3) evolution over time, (4) legal, political, and institutional consequences to fragmentation or integration, and (5) strategies to better ensure the efficacy, resilience, and coherence of the AI regime complex in the face of these conditions.

Finally, in the Conclusion, I briefly review the argument, reflect on the strengths and limits of the frameworks presented and the arguments made, and work out implications of this for pressing questions in AI global governance. I reflect on what this tells us about how to govern AI, how AI changes governance, and what we can learn about the changing nature of global cooperation and about the urgency of human choice in the face of technological change. Let us begin.