Introduction to *The International Governance of Artificial Intelligence*

An autonomous tanker runs aground, spilling its oil and causing severe environmental damage. Authoritarian regimes use facial recognition and massive databases of photographs, videos, and geolocational data to surveil and target minority ethnic groups. Such measures earn widespread condemnation from the international community. Autonomous machines and systems accelerate the pace of automation, causing unemployment to skyrocket among unskilled laborers, in turn creating unrest and destabilizing governments. Some countries develop autonomous weapons systems to bolster their military capability, forcing other nations to follow suit.

Artificial intelligence (AI) plays a role in all these scenarios. AI refers at a minimum to algorithms and techniques, often dependent on large amounts of data, that perform relatively sophisticated tasks without human intervention. The applications and technologies made possible by AI stand to positively impact every domain of human life; the concern is that they could also cause great harm, with many of the effects international in scope. There is, thus, a need to ensure that transformative technologies like AI are developed and deployed to enhance human life and to minimize harm. Accordingly, formal and informal rules and practices concerning artificial intelligence are emerging through the interactions of various stakeholders. These efforts address the concerns raised above: safety, privacy, the environment, employment, and international peace and security—as well as the technical architecture on which AI development depends.

The purpose of this book is to describe a nascent international ‘law’ of artificial intelligence and to understand some of the dynamics that create it. Such law is a combination of hard and soft law that is emerging from the complex interactions of several sources of that law and of the actors who propound them: private ordering by companies and associations, often dictated by the market; artificial developers and academics; national laws and regulations promulgated by states; the corpus of international law itself; international organizations; and civil society. The regulatory tools for AI governance fall along a spectrum from private ordering to formal regulation, shaping an already-existing normative landscape within which stakeholders in artificial intelligence operate and that stakeholders try to mold.
The book is organized into four parts. Part I discusses the rise of artificial intelligence applications and provides a framework for their governance. Chapter 1 gives some technical background about artificial intelligence and shows how AI applications stand to have profound effects on all aspects of human life, some of which are becoming international in scope. The chapter reports that although there are good reasons to steer the development and use of AI applications, the governance of technology has been fraught at every level of governance. Chapter 2 introduces the work of Laurence Boulle and others who approach international governance by exploring the interactions of soft and hard norms and the actors who form and are affected by those norms. The chapter also draws from the work of Iris Chiu, who explains how various tools for governance fall along the spectrum discussed above, with varying degrees of effectiveness over those who are being governed. Chapter 2 concludes by discussing some of the standard tools for governing AI applications that are being proposed and used by stakeholders in AI governance.

Part II forms the bulk of this study. This part examines each of the major actors involved in artificial intelligence development and governance. It can be argued that each actor is informed to some extent by internal logics that drive the norms they are apt to develop and support. At the same time, each actor must contend with the norms that already exist and with other actors who occupy the same space. In addition, at the margins, the constituents of organizational actors can overlap. As a result, the interactions of actors and sources sometimes blur the lines between them.

Chapter 3 discusses the role that the market and technology firms play in the governance of artificial intelligence. The chapter examines how market imperatives and existing legal doctrines such as shareholder primacy inform the decisions that technology firms make about their business models and about the AI-driven products and services they seek to develop and market. The ethos of corporate social responsibility counters those dynamics to some extent, and responsible technology companies have adopted internal measures and industry practices to govern the development and use of AI technologies. But the strong incentive to monetize AI capacities and to take advantage of large amounts of data collected by their products and services, combined with the enormous power of businesses individually and through industry associations to shape the rules that govern them, suggest that other sources of governance are needed.

Chapter 4 considers how artificial intelligence developers, professional associations, and academic institutions also participate in the development of AI norms. Like representatives from industry, they provide expertise to policymakers at all levels of government and participate to some extent in the deliberations of international organizations. These actors often have close relationships with businesses as employees, consultants, or recipients.
of funding. At the same time, the case can be made that they enjoy a meaningful degree of independence. Programmers and researchers are developing technical methods to mitigate risks posed by AI applications. Perhaps more importantly, a distinctive contribution of these actors is the norm of openness in AI development. At the international level this is evidenced by collaborations between AI researchers from different countries despite national security concerns and by the open software movement, the ideals of which have been incorporated into the practices of technology companies, despite initial opposition, and endorsed by international organizations.

Chapter 5 turns to the all-important role of the state. States are uniquely poised to respond to the issues raised by artificial intelligence applications. Some have chosen to play a coordinating role in articulating and facilitating strategies for the development and use of AI while others have taken a largely hands-off approach, leaving it to already-existing law and the market to address AI applications. The chapter examines AI legislation that has been enacted by states to respond to specific policy issues. However, to frame that examination, it first discusses legislation of the European Union. Because of the path-driven nature of governance, the size of the EU market, and the laws’ extraterritorial reach, the EU in some respects has become the de facto regulator of AI applications. Through legislation, states and regional entities such as the EU express choices about privacy, the effect of algorithmic decisions, and the many ways AI applications affect areas of concern to the state and its citizens. At the same time, such decisions have international consequences as states exercising their prerogatives affect the interests of other states, private firms, and civil society. The chapter also argues that the power of states is not limitless. States struggle to govern their own populations. The demands of governance are leading states to use AI applications to assist in performing public functions. Depending on how deeply integrated and how autonomous these applications are, such use challenges traditional understandings of who the governed are and what governance means.

Chapter 6 turns more directly to international norms and examines how international law shapes and limits the prospects for AI governance at the international level. Despite developments over the past 50 years, states remain the primary actor under international law, and have almost exclusive power over their territories and those within them, as well as power to make binding international rules. As an example, the chapter weighs how these principles affect current debates about data sovereignty. At the same time, states have used treaties to arrange their relations with one another and to create regimes to govern matters of international interest and concern. The chapter examines how such a regime has framed discussions about international intellectual property protection for AI-generated inventions and works. Chapter 6 continues by discussing how international trade and
investment law doctrines, such as the national treatment requirement, might constrain states’ ability to govern cross-border digital services and data transfers, necessary for the provision of services with artificial intelligence components and the training of artificial intelligence models. At the same time, that international law applies only indirectly to multinational enterprises shows the limitations of international law as a tool for AI governance.

Chapter 7 considers the role of international organizations in the formation of international AI norms. Several of these institutions are engaged in crafting rules for artificial intelligence applications out of their own legal frameworks and competencies. International organizations provide forums where international actors debate the formation of international norms for AI. They facilitate agreement among states, sometimes by virtue of their substantive expertise and sometimes because states use them strategically to take policy decisions that would not be feasible on the domestic level. At the same time, international organizations have their limitations. Like other forms of centralized governance, they are vulnerable to ossification, infighting, incompetence, and sometimes industry capture. Even more than their domestic counterparts, international institutions are unaccountable to the public. These shortcomings threaten the validity of the policy decisions that emerge from these institutions.

Chapter 8 concludes Part II by examining the activities of social movements and civic organizations at the international level. Civil society participation in the formation of international norms occurs within a broader acceptance of a role for activism, social movements, and non-governmental organizations in the cycle of international norm formation. Protests and grassroots movements have educated their participants, followers, and other groups about AI issues of importance to them, they have resisted certain AI applications, such as facial recognition for surveillance, and have advocated policies in furtherance of their missions. Civic organizations concerned about artificial intelligence are participating in agenda-setting, rule formation, implementation and enforcement of norms that implicate AI. Here, too, they bring issues to the attention of international decision-makers and inform their constituents of international developments. As stakeholders, they help to lend legitimacy to international decisions. Yet, there are valid arguments that civil organizations are at best proxies for a much broader civil society that is impossible to be heard at the international level in any meaningful way. Thus, these organizations must contribute to other aspects of legitimacy, such as effective governance. Finally, they cannot escape what is true with all such organizations—under the present international legal structure, states, at least the more powerful of them, have the final word on what norms will eventually be adopted.
Part III takes a broader perspective on artificial intelligence governance. Chapter 9 investigates how the development of artificial intelligence has come to be associated with geopolitical dynamics between states, which in turn is affecting how AI is being governed. States increasingly see AI as enhancing their economic and political power, hence affecting their influence in international affairs. The chapter discusses the rivalry between the United States and China and how that rivalry is increasingly being understood in realist terms. This dynamic leads to conceiving AI development as a race between nations and regions, with direct effects on policy choices to allow that development to proceed, either through the top–down approach chosen by China or the hands–off approach chosen by the United States. The dynamic also colors how foreign policy experts welcome or view with suspicion state economic and foreign policy measures such as China’s Digital Silk Road initiative. At the same time, not all aspects of AI development and AI applications are part of a zero-sum game between geopolitical rivals. The chapter will examine the distinction between non-zero-sum and zero-sum games by comparing international efforts to develop common safety standards for autonomous vehicles on the one hand and efforts to regulate lethal autonomous weapons on the other. These examples suggest that because artificial intelligence is being used in many domains and because of an international climate that seems averse to cooperation, most governance of these technologies should take place within those domains.

Chapter 10 concludes this study. For purposes of governance, especially on the international level, I join other commentators in arguing that international human rights serve, first, as the best articulation the international community has made of its ultimate goals, and second, as the source of rules and standards by which AI norms at all levels of governance and in all domains are to be evaluated. The chapter first weighs arguments that human rights are weak candidates for this role. International human rights have been criticized for their provenance and for their ineffectiveness. It might be well and good to adopt them as abstract, non-binding principles, but as legal rights, the record of state compliance with them is mixed, and as a matter of law they do not reach the businesses that are developing, marketing, and using AI applications. However, at the very least, human rights provide a common language to conceptualize and evaluate the impacts of artificial intelligence on people. Virtually all countries are signatories to the major human rights conventions and have agreed that human rights are universal in application. Human rights have come to represent a set of widely accepted values. Although they do not address directly all areas relevant to AI such as technical standards, they inform what those standards are to accomplish. Finally, under the current state of international law, private firms have a responsibility to respect international human rights, and no matter what
an actor’s technical status, human rights exert their own normative force. International actors will likely argue about the meaning and scope of these rights or whether specific applications even raise human rights concerns. However, no other framework with international reach provides better terms for vigorous debate and eventual consensus. Part IV concludes.