1. The need for the international governance of AI

As discussed in the introduction, artificial intelligence stands to impact every domain of human life, and several of these impacts will be international in scope. This raises the question whether it is desirable or possible to govern AI applications at the international level. This chapter lays the groundwork for that inquiry. It begins with basic concepts that will be used throughout the book. It then surveys possible international impacts of AI in various domains and concludes by placing debates about the governance of artificial intelligence within larger problems with the governance of technology in general.

BASIC CONCEPTS

There are no universally agreed definitions for terms such as artificial intelligence, AI systems, autonomy, or autonomous systems, and they are often used interchangeably. “Autonomy” has been defined as “the quality or state of being self-governing” or “self-directing freedom and especially moral independence.”¹ Some scholars have argued that any self-executing system that can ‘sense’ and ‘respond’ to external stimuli can be considered autonomous. On this view, a mousetrap qualifies.² For most researchers, however, a device or system must also be able to retain the experiences of its interactions with the environment and then to adapt its responses to perform better the next time around. It must be thus capable of learning. Other commentators set a still higher standard: the machine or system must be able to determine its own goals and the means to achieve them. Murray Shanahan associates general intelligence with having a common-sense view of one’s surroundings and a measure of creativity in responding to new situations.³ There is the famous Turing test

² Patrick Chisan Hew, Artificial Moral Agents are Infeasible with Foreseeable Technologies, 16 ETHICS & INFO. TECH. 197, 198 (2014).
in which the human being becomes the measure of intelligence: a computer will have achieved intelligence when we are unable to tell if we are interacting with a human being or a computer. Others require machines to have some form of self-awareness or consciousness—the debates surrounding John Searle’s Chinese room argument go to what we mean by cognition and understanding, and whether it is possible to devise machines that can achieve them.

This book will not discuss the conceptual and philosophical issues that arise when we apply more stringent definitions of intelligence and autonomy to AI. Rather, it will take a largely pragmatic approach to autonomy and intelligence and focus on the capabilities of artificial intelligence, without asking whether AI shares human attributes such as consciousness or a sense of self. Further, although advances in large language models such as GPT-4 have caused some scholars to revisit these issues, this book focuses primarily on systems and machines that fall short of possessing general intelligence, because these are already on the scene and are raising issues of their own.

Artificial intelligence is of course a kind of technology. Technology has been defined as the “application of scientific knowledge to the practical aims of human life” and “a capability given by the practical application of knowledge.” These definitions suggest that technology straddles science and the ordinary aspects of human experience. Various taxonomies of technologies have been proposed; Susan Brenner, for example, distinguishes tool technologies, machine technologies and smart technologies. Tool technologies such as drills are “more or less complex implements/processes an individual use[s] to intelligence and cognition); Shane Legg and Marcus Hutter, *A Collection of Definitions of Intelligence, in Advances in Artificial General Intelligence: Concepts, Architectures and Algorithms* 17 (Ben Goertzel and Pei Wang eds, 2007) (discussing 70 definitions of intelligence).

4. See, e.g., Blaise Agüera y Arcas, *Do Large Language Models Understand Us?* 151 *Daedalus* 183 (2022) (making the case that large language models do approach intelligence if intelligence is understood and assessed through the interactions with others); Gabriel Grand et al., *Semantic Projection Recovers Rich Human Knowledge of Multiple Object Features from Word Embeddings*, *Nature Hum. Behavior*, Apr. 14, 2022 (presenting a method by which artificial intelligence can emulate human methods of understanding the meaning of words as opposed to the standard approach of measuring the proximity of one word to another).


carry out a physical task.” Such technologies are “extensions and extrapolations of the earlier, more primitive tools that were once used to carry out simpler version of the same tasks.” Machine technologies, like the bicycle, are those that to a large extent replace human effort. Artificial intelligence applications surpass machines and join other smart technologies like blockchain and nanotechnology, but like less complex technologies, intelligent systems and machines involve the application of science to aspects of human life and expand human capacities. This identification of artificial intelligence with science is important because of the instrumental and symbolic roles science has come to play in society.

Machine Learning, Deep Learning

To better appreciate some of the issues posed by artificial intelligence applications, it is helpful to discuss current approaches to machine learning because such learning drives much of what is referred to as artificial intelligence. Jeremy Howard and Sylvain Gugger define machine learning as “[t]he training of programs developed by allowing a computer to learn from its experience, rather than through manually coding the individual steps.” Machine learning comprises several programming methods for creating what are termed models from data. Instead of writing step-by-step instructions to have a computer classify an image as a dog or a cat, the goal is to have the computer create its own program, i.e., a model for doing so by itself, based on data that it has been given. Computers accomplish this via algorithms: sets of automated instructions designed to perform certain tasks such as classification; regression, including prediction; and clustering data into groups.

---

7 Susan W. Brenner, Law in an Era of “Smart” Technology 17 (2007).
8 Id.
9 Id., at 25.
13 Microsoft, How to Select Algorithms for Azure Machine Learning, Mar. 1, 2022, https://docs.microsoft.com/en-us/azure/machine-learning/how-to-select-algorithms; “A regression model is one that attempts to predict one or more numeric quantities, such as a temperature or a location.” 62
Deep learning is a type of machine learning technique:

it is a computer technique to extract and transform data— with use cases ranging from human speech recognition to animal imagery classification— by using multiple layers of neural networks. Each of these layers takes its inputs from previous layers and progressively refines them. The layers are trained by algorithms that minimize their errors and improve their accuracy. In this way, the network learns to perform a specified task.\(^1\)

Consider using deep learning to create a model that will distinguish between images of dogs and cats. Because computers can process only numbers, what a computer ‘sees’ when presented with an image is a set of numbers, perhaps the degree of intensity of pixels. The first layer of the model might simply try to predict whether a given pixel is part of an edge, which will be represented by a number. The next layer will receive the results from the previous layer and perhaps predict whether an edge corresponds to a particular part of the body, again represented by numbers, and so on through various layers until the model makes a final prediction whether or not the image it has been shown is a cat or a dog. At each step of the process, the program instructs each input to be multiplied by a weight, an estimate of how important the input, such as pixel intensity in the first layer, is for the next layer in the analysis. The process involves several layers, and each layer involves receiving an input, performing a calculation (multiplying by the weight), and then forwarding the results of that calculation to the next layer. These three functions: receiving an input, performing a calculation, and then forwarding the results, resemble the functions performed by neurons in the brain, hence the term neural network.

This technique ‘learns’ in the following way. In supervised learning, the computer program described above is trained by being ‘shown’ many pictures of cats and dogs that have been labeled as such. After the computer program makes its predictions as to whether the images are cats or dogs, it compares its results with how the images were labeled. The difference between the program’s prediction and the labeled results is known as the loss. The program is designed to readjust the weights until the loss is minimized. The final combination of layers and weights becomes the model. The model is then validated by being asked to classify new images. In another approach, unsupervised learning, the model is not trained with a labeled data set; rather the model is asked to cluster inputs and find hidden patterns in a data set on its own, in effect labeling on its own the data with which it is presented.\(^2\) More sophisticated

\(^{1}\) Id. at 22.

\(^{2}\) Osvaldo Simeone, *A Very Brief Introduction to Machine Learning with Applications to Communications Systems*, 4 IEEE Trans. on Cognitive Comm. &
natural language processing applications use a combination of unsupervised and supervised training.¹⁶ In another approach called reinforcement learning, a model—for example one for playing chess—generates a particular move and then tests that move in the game. The results are then used to readjust the model.¹⁷

These examples suggest that in the technical sense, artificial intelligence refers to a collection of computer techniques that are designed to perform descriptive, diagnostic, predictive, and prescriptive tasks. As a result, in general, to regulate artificial intelligence is to regulate applications of these techniques.¹⁸ Consistent with the literature on artificial intelligence governance, this book will refer to ‘algorithmic decision-making,’ ‘algorithmic predictions,’ and ‘algorithmic fairness.’ But the results of an artificial intelligence application, such as the identification of an individual by means of their photograph, is the product of more than an algorithm. It would be more accurate to say that such identification results from a model: a computer program that has been developed in part by an algorithm. The example also shows why data is so important to artificial intelligence applications. Techniques are being developed to reduce the need for data, but at present, data are necessary for both supervised and unsupervised learning techniques and will likely remain so for some time. The accuracy of a model therefore depends in large part on the amount and quality of data to which a model is exposed. Finally, the result of an AI application is probabilistic: a classification system for example can only assign a probability, albeit sometimes a very high one, that the image it is being asked to identify is a cat. Perhaps this reflects how human beings really perceive the world, but this runs counter to popular belief that the decisions of artificial systems are based on certainties.

INTERNATIONAL IMPACTS OF AI SYSTEMS

Even though artificial intelligence is circumscribed by the techniques described above, it is increasing human capabilities to describe states of the world; to

---

¹⁷ Naeem, Rizvi, and Coronado, *supra* note 15.
The need for the international governance of AI

diagnose problems; to predict human behavior and other outcomes; and to prescribe courses of action to achieve defined goals. These capabilities are being combined with more sophisticated activators and sensors in devices and machines with increasing levels of autonomy. This is happening in fits and starts, but artificial intelligence applications stand to become ubiquitous within major domains of contemporary human life: transportation, home and service robots, health care, education, services to underserved communities, public safety, employment and the workplace, and entertainment. They also will become ubiquitous in human daily experience. This book will discuss specific applications of artificial intelligence within those domains, but the potential impacts on the international level are better understood by locating artificial intelligence within the web of international activities that straddle different international domains and legal regimes. Such global impacts can be grouped under the three forms of international economic activity: international trade in goods and services, international finance, and international labor flows. This includes the international infrastructure that enables AI technologies. Moreover, there is the potential impact of artificial intelligence on international peace and security, the environment and climate change, and international public health. Each of these areas is discussed below.

International Trade in Goods and Services

According to the United Nations Conference on Trade and Development (UNCTAD), despite the negative impacts of the pandemic, global trade in goods and services was valued at US$28.5 trillion in 2021. The World Trade Organization (WTO) predicts that digital technologies, of which artificial intelligence is a component, will affect the trade of goods and services in several ways. Such effects include lowered international trade costs; a change in trade patterns, with an increase in trade in digitally deliverable

---

19 A second way to assess potential impact is to anticipate how artificial intelligence will affect a person or community’s daily lived experience. It can be argued that a person’s subjective daily experience of AI is not as germane to international governance, but not without some relevance, particularly as it raises concerns of international human rights, understood in the strict legal sense and in the ‘values’ that underlie those rights. I will discuss the role of international human rights in governing artificial intelligence in Chapter 10.

20 UNCTAD, Global Trade Update, Feb. 2022, at 2. Trade values were expected to remain roughly the same in 2022. Id.

services;\textsuperscript{22} and changes to the composition of trade in goods. Lower trade costs will result in part because of lower transportation costs. In this area, artificial intelligence is being deployed in all modes of transport, such as automobiles, trucks, trains, and ships, the operations and routing management of such modalities, and in their supporting infrastructure.\textsuperscript{23} Of particular interest is the development of autonomous ships\textsuperscript{24} and cargo aircraft.\textsuperscript{25} These applications are of international interest because ships, drones and, to a lesser extent, vehicles cross international borders, or in the case of ships, operate on the high seas and traverse home waters.

Concerning the composition of international trade, the WTO anticipates that trade in information technology goods such as computers and semiconductors will increase. Other technologies such as blockchain will also facilitate trade in time-sensitive goods; goods subject to certification and labeling requirements; goods whose sale is governed by groups of contracts, for example, the bills of lading, carriage contracts, and the like used in the international sale of goods; and goods subject to mass customization.\textsuperscript{26} At the same time, trade in digitizable goods such as CDs is likely to continue to fall. Further, the development of 3D printing could lead to a decrease in trade in certain goods as more of them can be produced in-country,\textsuperscript{27} and the development of business models such as ride-sharing could decrease demand for durable goods such as automobiles.\textsuperscript{28}

The WTO further predicts that digital technologies will enhance the importance of intellectual property rights, as more digital products are licensed instead of sold.\textsuperscript{29} Perhaps most importantly, a growing digital economy will affect a country’s comparative advantage, the conventional justification for why countries trade in the first place. The WTO suggests to the extent digital

\textsuperscript{22} Id., at 81–88.
\textsuperscript{24} International Maritime Organization, Visions of the Future as Maritime Safety Committee Celebrates 100th Session, IMO Maritime Safety Committee 100th Session, June 12, 2018, http://www.imo.org/en/MediaCentre/PressBriefings/Pages/22-MSC-100-special-session.aspx
\textsuperscript{25} Conde and Twinn, supra note 23.
\textsuperscript{26} World Trade Report 2018, supra note 21, at 88–92.
\textsuperscript{27} Id., at 92–95.
\textsuperscript{28} Id., at 95.
\textsuperscript{29} Id., at 97.
technologies make physical infrastructure less important to an economy, this will enable less developed countries to benefit from their comparative advantage in labor. But such technologies also require high-skilled labor, and to the extent that they substitute labor, the industries that support them will become more capital intensive. Further, since artificial intelligence relies heavily on data, firms in countries with large populations and markets that serve as data sources will have a competitive advantage over firms from smaller markets.

International Financial Transactions

According to one report, the global financial services market was expected to grow from US$20.4 trillion in 2021 to US$22.5 trillion in 2022. International finance refers to clusters of several interrelated financial activities. Hal Scott and Anna Gelpen organize the area into five groups. The first concerns international aspects of domestic financial markets. Domestic capital markets and banking and banking products such as savings accounts fall within this cluster. Second are activities that form the infrastructure for international financial activities. These include the Basel framework for capital adequacy, currency regimes and foreign exchange markets, and payment and clearing systems. International transactions in financial instruments and offshore markets such as the Eurocurrency market, asset securitization, equity, futures and options, swaps, and investment funds form the third cluster. Fourth, in emerging markets, project finance and emerging market debt are important types of financial arrangements.

---

30 Id., at 98.
31 Id., at 100.
34 HAL S. SCOTT AND ANNA GELPERN, INTERNATIONAL FINANCE, TRANSACTIONS, POLICY, AND REGULATION xi-xxiv (23rd ed. 2020). The fifth cluster comprises rules and norms within the financial community designed to prevent terrorism. Id.
AI applications in finance are mostly being implemented at the domestic level. Some applications, however, affect systemic aspects of finance that can have international significance. Researchers are exploring how artificial intelligence can be used in the infrastructure that underlies financial transactions. This includes using AI techniques to assess bank compliance with capital adequacy standards. A recent white paper suggests how AI might be used for aspects of the payment system for payments that take place between financial intermediaries. AI learning techniques can be used to predict international capital flows. Models are also being developed to assist in trading and investment decisions, risk assessment and management.
The need for the international governance of AI

Foreign Direct Investment

Foreign direct investment overlaps in part with international finance because the purchase of foreign equity and other securities is often considered a form of investment. The previous section has already discussed how artificial intelligence is being used in financial management and trading decisions. The decision to invest in foreign projects, whether through divisions, joint ventures, or through standalone entities, also involves formal and informal risk assessments and valuation forecasts that would be enhanced by artificial intelligence techniques. It is likely that artificial intelligence applications will assist in these pre-investment decisions, as well as in the ongoing operations of the investment once it is made.40

International Labor Flows

Many observers have raised concerns about the displacement of labor by artificial intelligence. This is not a new fear; there have always been worries that automation deprives people of employment. The counterargument has been that new technologies enhance human labor without making it redundant,41 or cause firms to enter industries made more productive by technology,42 thus leading to higher demand for labor.43 Others argue that every occupation involves several tasks, some of which are amenable to automation while others are not. Thus, although some occupations that involve low skills are vulnerable to redundancy, overall, this is not the case with most occupations.44 The distinctive issue posed by artificial intelligence is that AI applications appear to have capacities thought to be exclusive to human beings, so that occupations

40 Wendy Duong suggests that the increasing use of artificial intelligence could also impact patterns of foreign direct investment, as decreasing production costs made possible by AI applications reduce incentives to send work abroad. Wendy Duong, Effect of Artificial Intelligence on the Pattern of Foreign Direct Investment in the Third World: A Possible Reversal of Trend, 36 DENV. J. INT’L L. & POL’Y 325 (2008).
43 Melanie Arntz, Terry Gregory, and Ulrich Zierahn argue that in addition to these factors, the adoption of technology is often slow, thereby giving workers time to adjust to new technology. Melanie Arntz, Terry Gregory, and Ulrich Zierahn, The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis, OECD Social, Employment and Migration Working Papers No. 189, 22 (2016).
44 Id. at 23.
The international governance of artificial intelligence believed to require more skill and knowledge could fall within the reach of AI systems. For the international community, this raises questions about how artificial intelligence technologies will affect international labor flows and migration for work.

Platforms and Technical Infrastructure

Applications with artificial intelligence features do not always fall neatly within categories, and the drive to regulate them stems from a mixture of social, security, and competitive concerns. Social media platforms such as Facebook, Twitter, TikTok, and WeChat provide a broad range of what might be categorized as services and have become enormously popular. Facebook had 2.7 billion monthly active users as of June 30, 2020. TikTok, owned by the Chinese company ByteDance, Ltd., is a social media application used to share short videos on mobile devices. In May 2020, the application was the most downloaded app in the United States, and observers noted that the creators of TikTok were using artificial intelligence techniques to encourage more use of the service. WeChat, owned by Tencent Holdings Ltd, is a messaging, social media, and payments application. Originally designed for the Chinese market, Tencent reported that worldwide there were 1.17 billion monthly active users of the application. Because of their popularity, these platforms are used for sharing information and are sources of large amounts of data. The concern is that either can be misused or serve purposes that are contrary to domestic policy.

In addition to social media, infrastructure that will enhance the use of artificial applications also has international reach. Recently adopted 5G technology has two advantages over previous networks. 5G technology does not rely on fiber optics and is thus able to be built up more quickly. Further, 5G networks

47 Facebook, Inc., Quarterly Report, Form 10-Q, as of June 30, 2020, at 27.
The need for the international governance of AI

can transmit much more information while reducing latency to almost zero. The coordination of processing and transmitting information from decentralized locations is made possible by artificial intelligence. At the same time, the vast information flows over 5G networks enable artificial intelligence applications that rely on near-instantaneous data transmission. Governments have already identified possible risks associated with the technology. The United States and EU have expressed concerns that components manufactured by untrusted companies could expose users to malicious software and hardware, create new vulnerabilities within the networks, and give bad actors more vectors to attack the system and the data collected by it. Like social media apps, 5G network technology manufactured by Chinese companies, such as Huawei Technologies Co. and ZTE Corp., are subject to executive orders and regulation in the United States. The infrastructure that makes 5G possible is already subject to governance by international organizations whose activities are discussed in Chapters 6, 7, and 9.

International Peace and Security

Research and development continue in developing artificial intelligence and autonomous systems for warfare. The goal is for AI to become “a true assistant that aids in understanding the [military] operational environment while supporting the operations process.” This involves the use of autonomous and semi-autonomous systems with learning capability, as well as applications for “indications and warnings, counter-messaging and cyber defense, among other uses.” Present applications are steps towards that goal. They include using


53 James J. Mingus and David Dilly, On Warfare and Watson: Invest Now to Win with Artificial Intelligence, ARMY, Sept. 2017, at 32. See also Courtney Crosby, Operationalizing Artificial Intelligence for Algorithmic Warfare, 100 Mil. Rev. 42, 43 (2020) (“Operationalization hinges on the understanding that AI is not an end state but rather one way of achieving a military advantage.”)

54 Mingus and Dilly, supra note 53, at 33.
computer vision systems to help analysts recognize potential targets, predictive analytics to project equipment maintenance needs, and deep learning to detect heretofore unseen threats.\(^5^5\)

The use of artificial intelligence for military applications raises at least two concerns for international peace and security. As discussed above, military strategists see artificial intelligence as enhancing all aspects of military strategy and operations, but as is well known, most attention has been paid to lethal autonomous weapons systems, or LAWS. Most researchers believe that killer robots are still far in the future. Nevertheless, they worry that autonomous weapons will usher in a revolution in warfare that will allow war to be waged faster and at a greater scale. Such weapons could be used as weapons of terror and thus make asymmetric warfare more deadly and a ready option for sub-state and non-state actors.\(^5^6\) Second, recent cyberattacks on government agencies, businesses and the like have been of increasing concern, in part because they damage national security interests without rising to the level of armed conflict. Artificial intelligence stands to make less violent forms of cyber-warfare more effective, thus contributing to international instability.\(^5^7\)

These issues will be discussed in more detail in Chapter 9.

Because artificial intelligence will increase efficiencies in all military activities, some commentators warn that governments will feel compelled to develop and deploy AI applications, at the very least to maintain their relative strategic positions. James Johnson argues further that AI could have a destabilizing effect on older security strategies. In his view, although artificial intelligence applications will not in themselves be destabilizing, they will enhance other military technologies such as swarm technologies and hypersonic missiles that


\(^{57}\) *See* Guyonneau and Le Dez, *supra* note 55.
The need for the international governance of AI

do threaten the status quo. Those weapons systems represent larger trends in
“the acceleration of modern warfare, the shortening of the decision-making
time frame, and the commingling of military systems.” Johnson worries
that artificial intelligence will hasten these trends, which in turn will increase
the risk among other things of destabilizing nuclear deterrence. But there is
debate on this issue. Jessica Cox and Heather Williams argue that artificial
intelligence could indeed enhance nuclear stability, for example, by improving
early warning systems and by preventing false positives from nuclear attack
detection systems.

Of course, international security involves more than the use of force. Miles
Brundage and his co-authors consider how artificial intelligence can be used
for malicious purposes. The authors point out that AI systems have several
features that are useful to individuals who wish to do harm. AI systems often
have dual uses and can thus be used for detrimental and beneficial purposes.
They are efficient and scalable. As discussed, such systems can exceed human
capacities. AI systems can increase anonymity and the psychological dis-
tance between perpetrators and their victims. They can be rapidly diffused.
Because of these enabling features of AI, the authors expect that AI will allow
more actors to carry out attacks, increase their rate, and expand the number of
potential targets. Moreover, AI will create threats that could not be posed by
humans alone, such as malware and swarms. The authors anticipate that the
character of these attacks will also change as they will become more regular,
more finely targeted, and harder to trace. AI systems also have their own
unique vulnerabilities due to their dependence on data and the way in which
models are trained and operate, thus creating new types of harm. The authors
believe accordingly that attacks on those vulnerabilities will increase.

62 Id., at 16–17.
64 Brundage et al., supra note 61, at 19–22.
The Environment and Climate Change

Artificial intelligence is being used to address concerns about the environment. This includes efforts to monitor climate change and to model it, to predict future climate scenarios and their impacts, to reduce the emission of greenhouse gases and switch to more sustainable energy sources, and to adapt to and mitigate the effects of climate change. However, Benedetta Brevini points out that the infrastructure that supports artificial intelligence itself has negative impacts on the environment. She refers to studies that indicate that training artificial intelligence can require large amounts of energy. The devices and equipment driven by AI create their own energy consumption and electronic waste problems. Further, large centers where data is physically stored for cloud computing and other applications have high demands for energy as well as water for cooling. At the same time, these concerns should be put in perspective. One study calculates that in 2018, data servers accounted for about 1 percent of global energy use. In addition, efficiencies in server processing and server virtualization have slowed the increase in energy use for computation.

65 See, e.g., Chris Huntingford et. al, Machine Learning and Artificial Intelligence to Aid Climate Change Research and Preparedness, 14 ENV. RES. LETTERS 124007 (2019), at 3–7 (discussing how machine learning can be used in climate modeling).

66 Vito Alberto Pizzuli, Vito Talesca, and Gabriela Covatariu, Analysis of Correlation between Climate Change and Human Health Based on a Machine Learning Approach, 9 HEALTHCARE 86 (2121) (using machine learning to correlate anthropogenic climate change with human health).


70 Id., at 3, citing Emma Strubell, Ananya Ganesh, and Andrew McCallum, Energy and Policy Considerations for Deep Learning in NLP (estimating the CO₂ emissions of training for natural language processing).

The need for the international governance of AI

International Public Health

In the area of global health care, the hope is that artificial intelligence can be used to meet health care needs in low- and middle-income countries. Several observers believe that artificial intelligence applications have the potential to be used for population health, individual health, health systems, and pharmaceuticals and medical technology. All observers argue that governance is needed to ensure that such technology is designed and deployed equitably and appropriately. The use of artificial intelligence to respond to the COVID-19 pandemic is discussed in Chapter 5.

AI GOVERNANCE AND THE GOVERNANCE OF TECHNOLOGY

It should be no surprise that the developments discussed above have been accompanied by calls for their governance. For Allan Dafoe, the governance of artificial intelligence “explores how humanity can best navigate the transition to advanced AI systems.” Governance can mean “to control and direct the public business of a country, city, group of people, etc.” More generally, it can refer to having “a controlling influence on something.” The governance of artificial intelligence involves both the public aspects of governance and control over the technology itself. The intention is to develop AI applications in a way that provides the benefits promised by the technology and avoids their negative consequences. However, controlling the development and use of artificial intelligence is part of a fraught history of earlier attempts to regulate...
technology, with disputes about how technology is to be conceived, when it is appropriate to govern technology, and whether such governance is possible at all.

**Conceptions of Technology**

Technology and its relationship to society have been conceptualized in several ways, with different implications for its regulation. Technology has been viewed as neutral: technology can be used for good or bad purposes, but there is nothing objectional in technology itself. This understanding informs, for example, arguments that facial recognition should not be prohibited, rather its uses should be regulated. A second conception of technology is that it is inexorable. This is a deterministic understanding of technology in which the quest for knowledge or the demands of the market or of international relations requires new technologies.\(^76\) This view arises in arguments against bans on advanced technologies. Some observers believe for example that humankind is not yet prepared for general artificial intelligence and therefore should not develop it, but others counter that someone in the world will do so, if only to see if it can be done, or, more likely, to gain an economic or strategic advantage.

Other writers embed technology in the context of broader social and political arrangements. In a famous essay, Langdon Winner suggests that technology has its own politics. Technology, including artificial intelligence, has been used explicitly and implicitly in social ordering.\(^77\) Moreover, Winner argues that adopting a technological system “unavoidably brings with it conditions for human relationships that have a distinctive political cast.”\(^78\) He wonders as a result whether certain technologies require particular types of social structures, or at a minimum structures that are compatible with those technologies.\(^79\) Daniel Sarewitz lists what in his view are four political realities about technological development.

1. The pace and direction of advancing knowledge and applications is determined by human choice.
2. The specific directions that technoscience is steered, and the pace of its advance, reflect who is taking the decisions—their interests, values, motives, perspectives.

---


\(^78\) *Id.*, at 128.

\(^79\) *Id.*, at 132.
3. The decisions that are made are determined within a complex social setting that encompasses a range of socioeconomic, cultural, and political components.
4. The complex social setting interacts with the results of technoscientific advance to yield social outcomes. The setting, the science, and the outcomes mutually evolve over time.

Viewed in this way, the ‘realities’ of technology described in propositions 1 and 2 oppose the belief that technology is inexorable: humans decide what will be developed and the pace at which this will occur. At the same time, a sense of technology’s inevitability persists. If Sarewitz is correct that there is a co-evolving relationship between social setting, science and outcomes, determinism begins to reemerge. This determinism is not attributable to technology alone, but rather to the larger social environment that the technology has played a role in creating or enabling. This viewpoint is associated with the science and technology studies discipline, and one of its implications is that far from being inexorable, technology is a product of political decisions and thus can be subject to further political decisions about how it is to be governed.

Collingridge’s Dilemma

David Collingridge is recognized as best articulating an important problem in the regulation of technology that tends to support a view that technology is ultimately unstoppable. He argues that technology poses a dilemma to a society seeking to control it:

"[A]ttempting to control a technology is difficult, and not rarely impossible, because during its early stages, when it can be controlled, not enough can be known about its harmful social consequences to warrant controlling its development; but by the time these consequences are apparent, control has become costly and slow."
With respect to the first horn of the dilemma, Collingridge is informed by decision theory. When technology is at an early stage of development, he believes there is often not enough factual information to engage in a risk analysis because, according to decision theory, risk analysis requires that one knows the probability of various outcomes. Collingridge also contends that at that early stage, it is not possible to engage in Bayesian analysis, where one imagines a distribution of probable outcomes and then optimizes their sum. For Collingridge, that the information needed for those two types of forecasting is unavailable means that we act under ignorance when we try to govern technology, a term he uses to describe situations when one cannot undertake a risk analysis or optimize over a set of various criteria. Under conditions of ignorance, any decision made about an emerging technology will by definition be arbitrary.

Collingridge’s analysis thus calls into question the usefulness of forecasting potential impacts of technology, including artificial intelligence. This view is consistent with the tenets of complexity theory. In complex adaptive systems, the interactions of various agents can give rise to nonlinear phenomena that cannot be predicted from the actions of the agents themselves. Within such systems, such as a society, long-term forecasting is not possible. Under complexity theory, however, not only are the long-term effects of a technology impossible to predict, but the same is true with the long-term effects of any policy measures taken to address that technology. Neither Collingridge nor complexity theorists argue that one should abandon attempts to predict outcomes, but their work suggests that there are legitimate grounds for skepticism when trying to adopt global solutions over the long term. As will be discussed throughout the book, forecasting methods such as risk analysis and cost–benefit analysis are being used by all the actors involved in AI governance. However, the tension between what can and cannot be foreseen with respect to technology underlies much of the debate surrounding the regulation of artificial intelligence and informs its rhetoric.
An illustration of the challenges posed by forecasting comes from Saswat Sarangi and Pankaj Sharma. Saranji and Sharma assess the possible impact of artificial intelligence on employment, mentioned earlier in this chapter, and lay out six possible scenarios.\(^86\) One possibility is that AI will be responsible for significant job losses of 25–50 percent within the next 20 years, with no government intervention. In that scenario, the savings in labor costs would be retained by large firms with ordinary people left worse off. In another scenario, AI will cause significant job losses, but government intervention will spread cost savings to the larger population. A third outcome is that artificial intelligence will have a lower impact on jobs, resulting in a reduction in employment of less than 10 percent. However, this would likely be a reduction of lower-skilled jobs, leading to greater income inequality. In a fourth scenario, AI could have a low impact on employment that is spread across all sectors. Fifth, artificial intelligence might have the net effect of increasing employment. Finally, as it did in an earlier part of its history, artificial intelligence might encounter theoretical or technical roadblocks that halt its development, leading to no impact on employment.\(^87\) Obviously, each of these scenarios would warrant different private and public responses.\(^88\) One can guess which of the six outcomes are likely to unfold and where, but no one can predict this with certainty.

Because of these difficulties in forecasting, Collingridge and others turn to the other horn of the dilemma: that once technology has become sufficiently developed and adopted, it is too late to regulate it. To reduce this entrenchment, scholars argue for governance systems nimble enough to respond as soon as information about the impacts of a technology becomes available. Moreover, such systems require anticipating, as opposed to forecasting, potential areas of impact\(^89\) early in the technological development process. To obtain this information, stakeholders who stand to be affected by an emerging technology

---

87 Id.
89 See Sarewitz, supra note 80, at 100 (describing “an anticipatory approach, not in the futile sense of first predict, then take action, but in the sense of building institutional capacities to reflect on contexts and choices,” in which “[t]he goal is to build a capacity for reflexiveness—social learning that expands the realm of conscious and available choice—into science and technology institutions and decision processes themselves.”).
should be allowed to participate in all stages of the development process. As will be discussed throughout this book, this ‘solution’ to the dilemma in part informs what is arguably becoming a norm of stakeholder participation in the governance of AI.

The Pace of Regulation and Technological Change

Collingridge’s dilemma is closely related to another theme in the debate about regulating technology. Policymakers and commentators often argue that technological development outpaces regulation. Thus, AI governance is not just hindered by the lack of information about possible impacts, but also by the slowness of regulation to respond to technological change. The United Nations Conference on Trade and Development articulates this theme:

[T]he proliferation of new technologies threatens to outpace the ability of societies and policymakers to adapt to the changes they create. The rate of turnover of technology platforms can reportedly be as short as 5–7 years—half the 10–15 years it may take society and regulatory measures to adapt. This has created widespread anxiety, causing ambivalence to or rejection of technological advances such as … deep learning in artificial intelligence.

Such a view supports the belief that technology is inexorable and, if true, has obvious implications about the role law and other forms of regulation can play in its governance.

UNCTAD cites several reasons for this increase in the pace of development. Technologies build upon each other and enable the development of other technologies. Emerging technologies are tied closely to computing power so that Moore’s law, the prediction that processing power doubles every 18–24 months, has a multiplying effect. Combinations of technologies such as big data, cloud computing, and geolocation systems are making possible the simultaneous emergence of potentially disruptive technologies. The costs of digital devices and products can approach zero. Two international platforms, the internet and global positioning systems, have allowed sub-platforms to be


91 UNITED NATIONS CONFERENCE ON TRADE AND DEVELOPMENT, TECHNOLOGY AND INNOVATION REPORT 2018: HARNESSING FRONTIER TECHNOLOGIES FOR SUSTAINABLE DEVELOPMENT 3 (2018).
The need for the international governance of AI built, upon which businesses can operate. In turn, those platforms reduce entry costs for science and innovation and for new businesses.\footnote{Id., at 4–7.}

A person living 100 years ago might also have believed with some justification that the rate of technological change in their time was accelerating, so it is important to ask in what sense we are observing a new phenomenon. The assessment of innovation and change is its own discipline and thus this section can sketch only a few contours of the field. A large literature is based on a proposal by Robert Solow that the effect of technology can be measured by removing other factors of production, thus leaving technology as the final explanatory variable for growth.\footnote{Aurélien Saïdi, How Salient is the Solow Residual? Debating Real Business Cycles in the 1980s and 1990s, 51 HISTORY OF POL. ECON. 579 (2019) (discussing Solow’s work and subsequent assessments of the Solow residual).} However, for this study, a useful starting point is provided by the OECD and Eurostat in a series of manuals designed to standardize the collection of data about innovation.\footnote{The most recent edition of the manual is OECD AND EUROSTAT, OSLO MANUAL 2018: GUIDELINES FOR COLLECTING, REPORTING AND USING DATA ON INNOVATION: THE MEASUREMENT OF SCIENTIFIC, TECHNOLOGICAL AND INNOVATION ACTIVITIES 60 (4th ed. 2018) [hereinafter Oslo Manual 2018].} The manuals are concerned with the measurement of innovation in general, not strictly technological change, but the concepts used in the manuals are familiar and provide a framework for regulation.

The Oslo Manual 2018 defines an innovation as:

\begin{quote}

a new or improved product or process (or combination thereof) that differs significantly from the unit’s previous products or processes and that has been made available to potential users (product) or brought into use by the unit (process).
\end{quote}

The definition reflects several conceptual decisions. Its generality reflects a belief that innovation takes place in all sectors of economic activity.\footnote{Id., at 60.} The objects of innovation are products and processes. In businesses, products are the goods and services they offer.\footnote{Id., at 60–62.} Business processes refer to various functions necessary for a business to exist, such as production, distribution and logistics, marketing and sales, etc.\footnote{Id., at 70–71.} The term, ‘significantly different’ requires an innovation to have some degree of novelty or disruptiveness (an echo of the Schumpeterian concept of creative destruction), at least to the users within

\begin{verbatim}
92 Id., at 4–7.
95 Id., at 60.
96 Id., at 60–62.
97 Id., at 70–71.
98 Id., at 73.
\end{verbatim}
a particular industry or location.\textsuperscript{99} That the improved product must be made available to users or that the improved process must be put into use by the unit itself to constitute an innovation distinguishes innovation from the research and development that might have contributed to an innovation. Finally, that the definition focuses on innovations by units, such as companies and households, implies that the discipline of measurement aspires to be quite specific about individual behaviors around innovation. The focus on the unit also implies that assessing larger and broader impacts of innovation must be made in connection with other approaches or perspectives.

These concepts suggest a way to organize how one might assess technological change and the rate of that change. As discussed above, technological change and innovation are closely related but not synonymous. Strictly speaking, innovative changes are a subset of changes in technology itself, so that measuring innovative changes would leave out those that are not innovative. This might have implications for governance because it is possible for technology to impact society even if that technology is no longer novel. Nevertheless, most popular attention is being paid to new technologies that are innovative under the Oslo Manual’s definition, and in general, the goods and services and processes with artificial intelligence features, at least at this point in their development and use, most likely fall under that definition.

Other scholarship on rates of technological change is reflected in the UNCTAD report. Some measures go to aspects of the technology itself and the way it is produced. For example, the UNCTAD report refers to Moore’s law. Researchers have found that Moore’s law (generalized to suggest that the cost of a technology decreases exponentially over time) obtains in other industries such as beer production and offshore gas pipelines, albeit at lower rates.\textsuperscript{100} Other ‘laws’ have been forwarded. For example, Wright’s law, first postulated for aircraft production, states that the costs of production decrease at a rate related to cumulative production (thus informing the UNCTAD report).\textsuperscript{101} In a 2013 paper, Béla Nagay and his co-authors found that Moore’s law, Wright’s law and three others are predictive of costs across 62 different technologies in different areas, with some laws performing better than in predicting improvements in production over the short and long term.\textsuperscript{102} To the extent that these laws obtain in technology production, exponentially decreasing production costs would likely impact rates of change, or at least in the distribution of technological goods. If, for example, the cost of photovoltaics can be expected

\textsuperscript{99} Id., at 77–78.
\textsuperscript{100} Béla Nagy et al., \textit{Statistical Basis for Predicting Technological Progress}, 8 PLOS ONE e52669(2013) at 1.
\textsuperscript{101} Id., at 2.
\textsuperscript{102} Id., at 6.
to continue to decrease over time, this could accelerate the rate of solar energy adoption with obvious implications for carbon-based energy.

The quotation from the UNCTAD report also refers to technological life-cycles. Technology can be understood as having a lifespan that begins with defining a product’s requirements, then proceeds to prototype development and testing, early adoption by some users, widespread use of the technology and finally ends with obsolescence.103 If the lifecycle of emerging technologies is in fact shortening, this would suggest an accelerating rate of technological change. Joern Huenteler and his co-authors point out there are two general models of a technology’s lifespan. The first model examines cycles of product and process innovation. Under this model, in the early years of an industry, after a disruptive technology emerges, firms try to exploit the competitive potential of the new technology by producing competing product designs. Eventually a dominant design wins out and core components become standardized. This signals a switch from product innovation, which decreases over the lifespan of the technology, to process innovation, as firms make incremental changes in production processes.104 The second model posits that with some complex products and industries, product innovation never decreases, so that there is no major shift in emphasis to process innovation. Instead, “[a]fter the emergence of a dominant design (constituted by a common product architecture and standardized core sub-systems), innovation along the technological trajectory is focused on individual sub-systems and components.”105

From a governance perspective, lifecycle analysis has been used for forecasting the development of a particular technology, thus determining when intervention for governance purposes might be required,106 and in the systems that make such technology possible.107 It has also been adopted by the

---

103 Lifecycles can be described as having many more stages and different endpoints. For example, Daswin De Silva and Damminda Alahakoon set out a 19-stage lifecycle for the development of an AI product. This process includes risk analysis and stakeholder participation discussed above and elsewhere in this book. Daswin De Silva and Damminda Alahakoon, An Artificial Life Cycle: From Conception to Production, PATTERNS, June 10, 2022.

104 Joern Huenteler et al., Technology Life-Cycles in the Energy Sector—Technological Characteristics and the Role of Deployment for Innovation, 104 TECH. FORECASTING & SOC. CHANGE 102, 103 (2016).

105 Id.


107 Jochen Markard, The Life Cycle of Technological Innovation Systems, 153 TECH. FORECASTING & SOC. CHANGE 119407 (2020) (determining the lifecycle of networks of
OECD in its approach to AI governance. What then do these models mean for assessing the pace of technological change? As might be expected, they are ideal types, so no industry is completely described by either model. Huenteler and his co-authors see the two models as ends of a spectrum along which a particular industry might lie.\(^{108}\) The challenge is determining where along the spectrum a particular industry falls, then identifying the lifecycle stages of this part of the spectrum. Once one can articulate those stages, there must be some way of locating at which stage a technology is now to measure the rate by which it is moving through all the stages. (This is over and above the need to define the technology or industry whose rate of development one is trying to measure.) Researchers have accordingly tried to assess stages of development of an industry by examining levels and concentrations of capital investment in a given industry,\(^{109}\) adoption of capital goods,\(^{110}\) the number and rate of patents taken out in a field,\(^{111}\) changes in the demand and use of products,\(^{112}\) and other microeconomics measures.\(^{113}\)

Under these criteria, it is difficult to estimate the lifecycle of an artificial intelligence application because often not all stages of the cycle are considered, and different terminology is used. There are some indications that, at least in smaller firms, the time between the conception and the deployment of a machine learning model into production can be very short, on the order of a little over a month or less.\(^{114}\) Production, however, refers to making “the model’s predictions available to users, developers or systems....”\(^{115}\)
The need for the international governance of AI

suggests that the short timeframe between development to deployment can be somewhat misleading because an artificial intelligence model is often just one component of a much larger software package or system which is also undergoing its own development process. Moreover, the cycle differs if one is examining an organization that is considering adopting AI applications as part of its business model or mission or operations as opposed to developing them. However, there is a basis for arguing that the lifecycle of these applications can be much shorter than the time it takes to adopt more formal means to govern them.

CONCLUSIONS

If the foregoing discussion shows anything, it is that the issue of the governance of artificial intelligence is fraught. There are good reasons why we would like to steer the development of AI applications, but it has been difficult to govern technology, particularly through harder forms of governance such as laws and regulations, even on the domestic level, let alone on the international plane. The themes that have emerged from the history of the governance of technology—its neutrality, inevitability, political construction and meaning, rates of development, and Collingridge’s dilemma—set the stage for how debates about the regulation of AI, even at the international level, are occurring. At the same time, one should not conclude that the development and deployment of artificial intelligence is taking place in a vacuum. AI is already governed by existing law and other norms that empower and constrain technology companies, developers, researchers, governments, and civil society—the major actors in AI development. Such forms of hard and soft law have not remained static; they are evolving as new technology and the enhanced capacities they create interface with those norms. The next chapter sets out two frameworks for better understanding this dynamic.