

## 2. AI actors and the landscape of AI regulation

---

Chapter 1 made the case for the international governance of artificial intelligence, as AI is poised to have effects in every domain that is of international interest: international economic activity, climate change, peace and security, and public health. At the same time, the chapter set out challenges posed by the governance of that technology.

The governance of any complex phenomena is difficult, particularly on the international level. In prior work in this area, I have found helpful the work of two scholars. The first is Laurence Boulle, who explores various sources of the hard and soft ‘law’ of economic globalization.<sup>1</sup> Boulle’s framework is useful in seeing how various actors and sources of law interact to form the hard and soft law of artificial intelligence. The second framework for governance is presented by Iris Chiu. Chiu focuses on the regulation of business firms and financial institutions, but her work applies directly to the regulation of technology. In a 2010 paper, Chiu describes how regulation extends along a spectrum from private ordering via contract to formal regulation by regulatory agencies.<sup>2</sup> She assesses the strengths and weaknesses of each form of regulation in the contexts in which they are used. Boulle and Chiu are not alone in viewing governance and regulation in these ways,<sup>3</sup> but their presentations are useful starting points for this study.

Before exploring these frameworks in more detail, some remarks on hard and soft forms of law are in order. Like Chiu, Barnali Choudhury explains that

---

<sup>1</sup> LAURENCE BOULLE, *THE LAW OF GLOBALIZATION: AN INTRODUCTION* 363 (2009).

<sup>2</sup> Iris H-Y Chiu, *Enhancing Responsibility in Financial Regulation—Critically Examining the Future of Public—Private Governance: Part I*, 4 *LAW & FIN. MARKETS REV.* 170 (2010).

<sup>3</sup> Olivia Erdélyi and Judy Goldsmith see a very similar landscape for a nascent law of artificial intelligence in Olivia J. Erdélyi and Judy Goldsmith, *Regulating Artificial Intelligence: Proposal for a Global Solution*, 39 *GOVT. INFO. Q.* art. no. 101748 (2022).

hard and soft law fall along a continuum of enforceability.<sup>4</sup> Where a law lands on that spectrum depends on several criteria:

[W]hether obligations are binding, the degree of precision associated with the obligations and the legal consequences of performance and breach, whether there has been a delegation of the interpretations of the obligations to a third party, and the level of authority of the drafter...<sup>5</sup>

Choudhury explains that soft and hard law have their respective advantages and disadvantages. Soft law does not require formal enactment and yet can influence behavior. Authorities can thus be nimbler in responding to changing circumstances with this form of law. Sometimes it is politically unfeasible to reach agreement on harder law, so soft law can be a compromise when non-binding regulation is better than no regulation at all. Soft law can condition those being regulated to the behaviors required by those norms and pave the way for more stringent regulation later. Such law can lower contracting costs because, as just discussed, the procedures for achieving agreement on them are less formal and the stakes for imposing them are less high to all actors. Finally, soft law can complement hard law by filling in gaps left by hard law and by directing how hard law is to be followed.<sup>6</sup>

For Choudhury, hard law has its own strengths. A nation state can signal the credibility of its commitments by agreeing to be bound by hard law. It is often accompanied by mechanisms for its interpretation and application and so too can fill in gaps in legislation. Hard law can prevent states from engaging in opportunistic behavior or reneging on their commitments, thereby enhancing the legitimacy of the norms involved. It facilitates enforcement, particularly if enforcement is delegated to a third party and prevents self-interpretation of an agreement. Finally, hard law as such can enhance the legitimacy of an obligation, which in turn contributes to compliance with the norm.<sup>7</sup>

## THE ‘LAW’ OF ARTIFICIAL INTELLIGENCE

### The Creators of Hard and Soft Law and the Cycle of Regulation

Boulle identifies four sources of soft and hard law that comprise the law of globalization: the law of the market, the law of nation states, international

---

<sup>4</sup> Barnali Choudhury, *Balancing Soft and Hard Law for Business and Human Rights*, 67 BRIT. J. INT’L & COMP. L 961, 963–64 (2018).

<sup>5</sup> *Id.*, at 967.

<sup>6</sup> *Id.*, at 970–71.

<sup>7</sup> *Id.*, at 971.

organizations, and international law proper. It is an easy step from there to identify the actors from which such laws emerge. Boulle thus joins others in placing norms within the relationships between those who have important stakes in their development and application. Jochen Markard combines this approach with the lifecycle of technical innovation systems.<sup>8</sup> Such a system is “a set of networks of actors and institutions that jointly interact in a specific technological field and contribute to the generation, diffusion and utilization of variants of a technology and/or product.”<sup>9</sup> Markard uses LED technology for lighting as an example of a technology that resulted in a technical innovation system that centered around light-emitting diodes.<sup>10</sup> He first identifies the web of actors who contributed to the generation, diffusion, and use of that technology. These actors included “technology manufacturers, suppliers, vendors, research institutes, associations, public authorities, NGOs etc.”<sup>11</sup> Further, like Boulle, Markard is aware of more abstract “institutions,” meaning formal and informal norms that impact the development of innovation systems. Such institutions “comprise of formal structures such as regulations, standards or public policies as well as informal structures such as collective expectations, cognitive frames, user practices, social norms or culture.”<sup>12</sup>

Kenneth Abbott and Duncan Snidal give one of best accounts of how various actors contribute to the development of norms at the international level. They describe a triangle of international governance, with three possible actors in each of the three corners of the triangle: first, states and groups of states; second, private firms; and third, non-governmental actors.<sup>13</sup> Each of

---

<sup>8</sup> Jochen Markard, *The Life Cycle of Technological Innovation Systems*, 153 *TECH. FORECASTING & SOC. CHANGE* art. no. 119407 (2020).

<sup>9</sup> *Id.*, at 2 (citations omitted).

<sup>10</sup> *Id.*, at 12.

<sup>11</sup> *Id.*, at 2.

<sup>12</sup> *Id.* See also, Roxana Radu et al., *Normfare: Norm Entrepreneurship in Internet Governance*, 45 *TELECOMMUNICATIONS POL'Y* art. no. 102148 (2021) (describing how different actors are developing various types of norms for internet governance). Volker Schneider frames this approach in terms of policy networks, a field within political science that emerged in the 1990s. Although there are several branches of this approach, the underlying idea of each is that “policy-making is not simply the design by a single, monolithic actor (the state). It emerges from the interaction of multiple actors participating in the formulation of public (and sometimes private) policies.” Volker Schneider, *Business in Policy Networks: Estimating the Relative Importance of Corporate Direct Lobbying and Representation by Trade Associations*, in *BUSINESS AND GOVERNMENT: METHODS AND PRACTICE* 121, 123 (David Coen and Wyn Grant eds., 2006).

<sup>13</sup> Kenneth W. Abbott and Duncan Snidal, *The Governance Triangle: Regulatory Standards Institutions and the Shadow of the State*, in *THE POLITICS OF GLOBAL REGULATION* 44 (Walter Mattli and Ngaire Woods eds., 2009).

the three actors have various competencies necessary for any type of regulation: independence, representativeness, expertise, and operational capacity.<sup>14</sup> Independence and representativeness refer to a sense of fairness and participation in norm development, while expertise and operational capacity are related to specific knowledge about a regulatory subject area and the ability to govern.<sup>15</sup> These capacities in turn affect an actor's capacity to participate in five stages of regulation: agenda-setting, that is, identifying and bringing to discussion a matter for regulation; negotiation; implementation of the norms that are eventually adopted; monitoring compliance with those norms; and finally, enforcement.<sup>16</sup>

As discussed, the states, firms, and non-governmental organizations that make up the vertices of Abbott and Snidal's triangle have different degrees of competence in governance and hence have different strengths and weaknesses at each stage of the regulatory process. In the authors' view, although states have strong capacities at the domestic level, this is less so at the transnational level.<sup>17</sup> As will be explained in Chapter 6, all states enjoy independence as a matter of international law, thus any one state has less operational capacity at the international level vis-à-vis other states. States also lack regulatory independence in relation to other states because they tend to promote their own national interests.<sup>18</sup> International organizations comprising states can be said to have greater independence, operational capacity, and expertise, but less representativeness in relation to individuals. Further, their operational capacity can be limited because states often circumscribe their remit.<sup>19</sup>

Business firms, particularly major ones, have large operational capacities, access to decision-making forums, and managerial structures. Firms have expertise in their areas of business, thus enabling them to design standards at the negotiation, implementation, and monitoring stages of norm formation.<sup>20</sup> However, as will be discussed in more detail in the next chapter, they lack independence because of their primary goal of maximizing profit, thus impacting their ability to monitor compliance with norms. They also represent primarily their shareholders. As a result, "firms are unlikely to produce regulatory

---

<sup>14</sup> *Id.*, at 46. The authors, however, are primarily concerned with "regulatory standard-setting," voluntary standards that nevertheless address concerns that go beyond technical coordination and emphasize to some extent monitoring and enforcement. *Id.*

<sup>15</sup> *Id.*, at 46, 58–59.

<sup>16</sup> *Id.*

<sup>17</sup> *Id.*, at 61.

<sup>18</sup> *Id.*

<sup>19</sup> *Id.*, at 61. We will return to these themes for international organizations in Chapter 7.

<sup>20</sup> *Id.*

standards and programs that serve common interests, and may lack legitimacy and credibility in the eyes of the public—and certainly those of activists—even when they are sincere about regulation.”<sup>21</sup>

For Abbott and Snidal, NGOs vary in their competences in agenda-setting, monitoring, and enforcement. This includes the “ability to raise the saliency of issues, standards, and violations and to generate public responses.”<sup>22</sup> They are often viewed as independent and thus often have greater credibility.<sup>23</sup> NGOs also have expertise in particular areas and have a strong commitment to norms.<sup>24</sup> At the same time, Abbott and Snidal argue that an NGO’s very commitment to its goals could undermine to some extent the perception of independence.<sup>25</sup> Further, although NGOs purport to further public goals, the degree of representation and accountability varies.<sup>26</sup>

Because no actor has all the competencies needed to govern, it would be expected that the actors would form alliances to pool their respective strengths. As will be discussed in Chapter 8, sometimes this happens. However, Abbott and Snidal observe that in general, regulatory schemes are usually proposed and promulgated by single actors.<sup>27</sup> The authors suggest that this is due to the “regulatory standards bargaining game” that occurs among states, firms, and NGOs. Such bargaining takes place expressly and implicitly. The authors believe that each of the three actors have their own preferences and seek to meet them; this is the case even when each finds it in their interest to collaborate. When actors explicitly bargain with each other, “actors from different groups are likely to differ sharply over the structure and governance of schemes and the scope and content of their standards and procedures.”<sup>28</sup> Moreover, such collaboration carries risks such as losing credibility among constituents if the actors compromise with other actors. Further, such collaboration is costly, particularly when the actors are suspicious of each other.<sup>29</sup> Some actors have greater power to go it alone and create their own regulatory schemes (or not act at all),<sup>30</sup> so given the costs and risks of collaboration, actors who can do so will create their own regulatory structures.<sup>31</sup> These parallel schemes result in implicit bargaining among the actors, as each vie to have their individual

---

<sup>21</sup> *Id.*

<sup>22</sup> *Id.*

<sup>23</sup> *Id.*

<sup>24</sup> *Id.*

<sup>25</sup> *Id.*, at 62

<sup>26</sup> *Id.*

<sup>27</sup> *Id.*, at 46.

<sup>28</sup> *Id.*, at 63.

<sup>29</sup> *Id.*

<sup>30</sup> *Id.*, at 64.

<sup>31</sup> *Id.*, at 67.

scheme take precedence over others. This competition can lead to either higher standards as different actors seek greater legitimacy and support, or lower standards, say, for example, if actors with less monitoring and enforcement ability feel compelled to relax standards to persuade others to join their scheme and to comply with it.<sup>32</sup>

### Institutional Logics

As discussed in the last section, Abbott and Snidal contend that states, firms and non-governmental organizations have their own preferences, which influences whether they will cooperate with each other in the regulatory process. Each actor might be informed and motivated by institutional logics that overlap in some respects and differ in others. Institutional logics are “patterns or practices, assumptions, values, beliefs and rules by which individuals organize their efforts and provide meaning to their reality.”<sup>33</sup> More than one set of logics can exist in the same organization as well as vis-à-vis other organizations. Thus, although a private company might be primarily informed by a need to maximize returns for its investors, or a university by its mission to produce knowledge, the same organizations might be motivated by other logics as well.<sup>34</sup> Similarly, within a given domain such as health care, the “field is shaped by the institutional logics of the market, the logic of the democratic

---

<sup>32</sup> *Id.*, at 67–68.

<sup>33</sup> Mary Kay Gugerty, George E. Mitchell, and Francisco J. Santamarina, *Discourses of Evaluation: Institutional Logics and Organizational Practices among International Development Agencies*, 146 *WORLD DEV.* art. no. 105596, at 2 (2021), citing Roger Friedland and Robert R. Alford, *Bringing Society Back In: Symbols, Practices and Institutional Contradictions*, in *THE NEW INSTITUTIONALISM IN ORGANIZATIONAL ANALYSIS* 232 (Walter W. Powell and Paul J. DiMaggio eds., 1991); Patricia H. Thornton and William Ocasio, *Institutional Logics*, in *THE SAGE HANDBOOK OF ORGANIZATIONAL INSTITUTIONALISM* 99 (Royston Greenwood, Christine Oliver, and Thomas Lawrence eds., 2008).

<sup>34</sup> William Ocasio and Nevena Radoynovska, *Strategy and Commitments to Institutional Logics: Organizational Heterogeneity in Business Models and Governance*, 14 *STRAT. ORG.* 287, 289 (2016).

state, and the professional logic of medical care.”<sup>35</sup> Such logics can also inform the choices of those who work within a given organization.<sup>36</sup>

Institutional logics could well be at work within the actors in this study. Daniel Schiff and his co-authors reviewed 88 statements of ethics for artificial intelligence produced by the public sector, private sector, and NGOs.<sup>37</sup> Such statements are perhaps the primary device expressly targeted toward artificial intelligence. The authors found in general that non-governmental and public sector documents were produced by more participatory processes than statements from the private sector.<sup>38</sup> In their judgment, government and NGO statements tended to have “more ethical breadth and depth ... and are generally more similar to each other than to private sector documents.”<sup>39</sup> Only public sector documents engaged with topics of law and regulation, while the private and non-government sector ethics statements did not.<sup>40</sup> Private sector statements emphasized “ethical issues with ostensible technical fixes, such as algorithmic bias and transparency.”<sup>41</sup> The NGO sector addressed a wider set of topics, such as accountability and misinformation, explainability, misuse and hostile use of artificial intelligence applications than the other two sectors, while public sector AI ethics statements raised issues of unemployment and economic growth.<sup>42</sup>

That an organization and the broader institutions of which they are a part can be informed by patterns or practices, assumptions, values, beliefs and rules might help us understand the motivations of the various actors who are participating in the formation of AI norms. Moreover, these logics might shape how various actors choose the norms they do and how they understand these norms

---

<sup>35</sup> Thornton and Ocasio, *supra* note 33, at 104. Chris Skelcher and Steven Smith apply an institutional logics approach to hybrid governance organizations such as public–private partnerships and quasi-autonomous agencies. Chris Skelcher and Steven Rathgeb Smith, *Theorizing Hybridity: Institutional Logics, Complex Organizations, and Actor Identities: The Case of Nonprofits*, 93 PUB. ADMIN. 433 (2015). Carlos Closa uses an institutional logic approach to explain what he views as the European Union’s reluctance to sanction certain members’ failure to respect the rule of law. Carlos Closa, *Institutional Logics and the EU’s Limited Sanctioning Capacity under Article 7 TEU*, 42 INT’L POL. SCI. REV. 501 (2021).

<sup>36</sup> Benard Ngoye, Vincenta Sierra, and Tamkyo Ysa, *Different Shades of Gray: A Priming Experimental Study of How Institutional Logics Influence Organizational Actor Judgment*, 79 PUB. ADMIN. REV. 256 (2018).

<sup>37</sup> Daniel Schiff et al., *AI Ethics in the Public, Private, and NGO Sectors: A Review of a Global Document Collection*, 2 IEEE TRANS. ON TECH. & SOC’Y 31 (2021).

<sup>38</sup> *Id.*, at 32, 36.

<sup>39</sup> *Id.*, at 32, 37–38.

<sup>40</sup> *Id.*, at 36.

<sup>41</sup> *Id.*, at 32, 38.

<sup>42</sup> *Id.*

are to work. At the same time, if a single logic comprises patterns or practices, assumptions, values, beliefs and rules, and an actor can be informed by several logics, it will be hard to parse which logic predominates when specific decisions are being made. This becomes more complex when organizations such as professional associations allow other actors to become members. These circumstances open the door to other theories of individual and organizational decision-making, such as Herbert Simon's satisficing, to shed light on how norms are chosen and promoted by these actors.

## Norm Entrepreneurs

Each of the actors that will be examined in this study can be considered a norm entrepreneur. The term, according to Cass Sunstein, refers to people or groups "who are interested in changing social norms."<sup>43</sup> Sunstein argues that norm entrepreneurs are sometimes able to change social conditions because those conditions depend on social norms to which people have little private allegiance, simply paying lip service to them.<sup>44</sup> If dissatisfaction with those norms is widespread, norm entrepreneurs can take advantage of this and try to change them. They can do so by engaging in a number of strategies, such as "(a) signal[ing] their own commitment to change, (b) creating coalitions, (c) making defiance of the norms seem or be less costly, and (d) making compliance with new norms seem or be more beneficial."<sup>45</sup> According to Sunstein, if norm entrepreneurs are successful, their efforts can lead to norm bandwagons and norm cascades. Norm bandwagons happen "when the lowered cost of expressing new norms encourages an ever-increasing number of people who reject previously popular norms, to 'tipping point' where it is adherence to the old norms that produces social disapproval."<sup>46</sup> Norm cascades are "rapid shifts in norms."<sup>47</sup>

The norm entrepreneur has several consequences for this study. First, as discussed earlier, private or public individuals or entities can be norm entrepreneurs. Martha Finnemore and Kathryn Sikkink discuss the important role that norm entrepreneurs played in the formation of international norms on women's suffrage and the laws of war.<sup>48</sup> Other studies have explored how various actors

---

<sup>43</sup> Cass R. Sunstein, *Social Norms and Social Roles*, 96 COLUM. L. REV. 903, 909 (1996).

<sup>44</sup> *Id.*, at 909, 929.

<sup>45</sup> *Id.*, at 930.

<sup>46</sup> *Id.*, at 912 (footnote omitted).

<sup>47</sup> *Id.* (footnote omitted).

<sup>48</sup> Martha Finnemore and Kathryn Sikkink, *International Norm Dynamics and Political Change*, 52 INT'L ORG. 887, 895–99 (1998).



have acted in this capacity, sometimes in tandem.<sup>49</sup> We will see how such actors are trying to shape norms governing artificial intelligence applications. Norm entrepreneurship presumes of course that actors have a degree of agency in any given environment. That any number of organizations, firms and entities can be entrepreneurs means that they might vie with one another with competing sets of norms. This means that differences in power can impact which sets of norms will emerge. We will also see this dynamic at work throughout this study.

### The 'Laws' of the Market

The chapter is now ready to return to Boule's description of the four sources of 'law.' The first is the 'law' of the market. For Boule, these refer to principles of economics. This includes the laws of supply and demand and other micro- and macro-economic principles.<sup>50</sup> Of crucial importance is the market itself. A 'free' market provides information on how resources should be directed and allocated. An ideal market assumes several conditions, among other things: utility-maximizing individuals, complete knowledge, relative ease in entering markets, multiple players who do not affect price, and competition, among other things. The loosening of these conditions can lead to market failure. Supply and demand operate on the international level also and perform their allocational role. The most influential economic principle in international trade is comparative advantage, the idea discussed in Chapter 1 that every country has a relative advantage vis-à-vis other countries in the production of certain goods or services and a corresponding disadvantage in the production of others. It is therefore mutually beneficial for countries to produce goods and services in which they have a comparative advantage and traded for goods and services in which they are disadvantaged.

The market as such can be said to have its own organizing principles, and through this lens, government intervention is viewed as distorting the market.

---

<sup>49</sup> See, e.g., Daisuke Madokoro, *International Commissions as Norm Entrepreneurs: Creating the Normative Idea of the Responsibility to Protect*, 45 REV. INT'L STUD. 100 (2019); Daniëlle Flonk, *Emerging Illiberal Norms: Russian and China as promoters of internet content control*, 97 FOR. AFF. 1925 (2021); ANNEGRET FLOHR ET AL., THE ROLE OF BUSINESS IN GLOBAL GOVERNANCE: CORPORATIONS AS NORM-ENTREPRENEURS (2010); Balázs Szent-Iványi and Pēteris F. Timofejevs, *Selective Norm Promotion in International Development Assistance: The Drivers of Naming and Shaming Advocacy among European Non-Governmental Development Organizations*, 35 INT'L REL. 23 (2021); Steven A. Hatcher, *Norm Proselytizers Create A Privacy Entitlement in Cyberspace*, 16 BERKELEY TECH. L. J. (2001); Carol M. Glen, *Norm Entrepreneurship in Global Cybersecurity*, 49 POL. & POL'Y 1121 (2021).

<sup>50</sup> Boule, *supra* note 1, at 70.

However, Boule notes that much of formal law supports the market and is constitutive of it. For example, “[c]ontract law has a close affiliation with the market because it nominally allows the parties to choose and determine the substantive and procedural norms and dispute resolution systems that will inform and regulate their commercial relationship.”<sup>51</sup> Property law also supports the market: “[s]ystems of property provide a framework of certainty and reliability in relation to security of title and rights ... in all forms of property, such as land, personal property, patents, shares, and profits.”<sup>52</sup> Finally, the law of business associations enables businesses to enjoy economies of scale not possible for sole proprietors. Under those laws, businesses can separate business: making business decisions, carrying them out, and serving as the residual claimant of any profits. Fiduciary duties address potential conflicts among the participants in a business that can arise because of the separation between ownership and control. All of these doctrines are geared towards the market.<sup>53</sup> As Boule puts it, “the dominant corporate objectives remain the pursuit of profit and, in this sense, they maintain and reproduce the market system.”<sup>54</sup>

Boule notes that contract, property, and company law function primarily on the domestic level, but at the international level, treaties have tried to harmonize domestic law in economic matters, particularly in contract. The UN Convention on the International Sale of Goods<sup>55</sup> is a well-known example. International trade and investment agreements, such as the General Agreement on Tariffs and Trade and the United States–Mexico–Canada Agreement represent national concerns, but they also embody market norms. Finally, international dispute resolution mechanisms often allow businesses to pursue claims apart from the domestic law of host countries.<sup>56</sup>

## National Law

As discussed, the laws of the market are often dependent on formal law such as contract, property, and the law of business associations. Boule argues that globalization has in some respects weakened the power of states and yet depends on state institutions and laws.<sup>57</sup> This is not only because contract, property, and business associations law operate primarily on the domestic level. Of equal

---

<sup>51</sup> *Id.*, at 75.

<sup>52</sup> *Id.*, at 76–77.

<sup>53</sup> *Id.*, at 80.

<sup>54</sup> *Id.*, at 81.

<sup>55</sup> United Nations Convention on Contracts for the International Sale of Goods, U.N. Doc. A/CONF.97/18CONF.97/18, art. 5 (1980).

<sup>56</sup> *Id.*, at 92–97.

<sup>57</sup> Boule, *supra* note 1, at 100–101.

importance, the state effectuates and gives legitimacy to economic norms that are codified at the international level.<sup>58</sup> International treaties are often not self-executing, so their provisions must be implemented by domestic legislation before they become operative. Measures taken by international organizations that have domestic effects also follow this pattern. The International Monetary Fund, for example, offers financial assistance to countries with balance of payment problems and often conditions such assistance on structural changes within the economies of those countries. Those conditions are never implemented directly; rather, the recipient country must enact legislation that does so. This restructuring is often controversial and unpopular because of the negative impacts it can have on people, but at a minimum, it can be argued that the legislative bodies of the recipient countries have approved them.

### **International Law of Treaties**

In Boule's schema, international treaties are an important source of rules that govern economic globalization. They require states to comply with their terms, which are often influenced by the laws of the market.<sup>59</sup> Treaties also influence the form of domestic law and regulations.<sup>60</sup> They regularly provide for dispute resolution mechanisms that often transcend those of individual states.<sup>61</sup> At present there are no overarching, bespoke international agreements concerning the use of artificial intelligence. At the same time, parts of existing treaty regimes, such as the Agreement on Trade-Related Aspects of Intellectual Property (TRIPS), and the treaties and customary international law that make up humanitarian law, do apply to specific uses of artificial intelligence or the data on which many applications rely, or at a minimum create a legal infrastructure for its operations. Further, as will be discussed in Chapter 5, at the regional level, the EU's existing and proposed directives and regulations on data protection, artificial intelligence systems, and access to data platforms form an overarching transnational regime that reaches far beyond its borders.

### **International Organizations**

International organizations are formed by treaties between states, but they themselves are international actors. With economic globalization the most influential international economic institutions are the International Monetary

---

<sup>58</sup> *Id.*, at 117–18.

<sup>59</sup> *Id.*, at 149–56.

<sup>60</sup> *Id.*, at 156–58.

<sup>61</sup> *Id.*, at 169–73.

Fund, the World Bank, and the World Trade Organization.<sup>62</sup> Concerning artificial intelligence, there is no corresponding set of institutions devoted to the development and regulation of technology, let alone to AI applications. However, several international organizations, the United Nations and the Organisation for Economic Co-operation and Development (OECD) in particular, are playing important roles in the development of international norms for AI. Further, as discussed above, the EU as a regional organization is very much the source of binding law that has worldwide impacts on artificial intelligence. AI legislation enacted by the EU will be discussed in Chapter 5, and other international organizations will be explored in more detail in Chapter 8.

## THE REGULATION OF COMPLEX INDUSTRIES AND THE LANDSCAPE OF REGULATION

### A Complex Regulatory Landscape

Recall from Chapter 1 Collingridge's Dilemma: when a technology is just developing, it is hard if not impossible to predict what impacts that technology will have. As a result, any regulations imposed at that stage are likely to be inapt, but by the time those impacts have become known, it is often too late to regulate them. The previous section has discussed how major sources of hard and soft law and the actors that both form and are governed by them interact in the international economy. This section focuses on the ways in which those actors and tools for regulation might play out within large and complex industries or fields such as artificial intelligence.

As discussed at the introduction to this chapter, Iris Chiu is interested in the regulatory space of the financial industry. She views that landscape as: "on the one hand decentred, populated by resourceful, competent and powerful industry participants, and on the other hand consisting also of agencies of authority that have a public character, whether state-based, regional ... or international..."<sup>63</sup> For Chiu, governance in finance is largely public-private governance, which she defines as "a sharing and morphing together of competence between actors of a public and private character, in addressing regulatory concerns and problems."<sup>64</sup> This type of relational regulation "reflect[s] the changing role of regulators not as platforms of authority, but as participants in a relational paradigm with the industry in providing governance over the financial services

---

<sup>62</sup> *Id.*, at 175–76.

<sup>63</sup> Chiu, *supra* note 2, at 171.

<sup>64</sup> *Id.*, at 171–72 (citation omitted).

landscape.”<sup>65</sup> As will be seen, the same can be said of the relationship between the developers of artificial intelligence applications and regulators in jurisdictions such as the United States.

Chiu mentions that the various forms of relational regulation between regulated parties and regulators are often paired with risk-based regulation, although the two are not necessarily linked. In this aspect of regulation, “the private and social risks of [certain] activities are measured, adjusted and assessed so as to inform the regulator of the appropriate regulatory approach to take with respect to a firm or the industry in general, having regard to the regulator’s resources.”<sup>66</sup> This approach often appears in artificial intelligence regulation. An obvious example is the proposed EU Artificial Intelligence Act (AIA).<sup>67</sup> The draft legislation categorizes artificial intelligence systems by the degree of risk they pose to the public. Practices that pose an unacceptable risk to the public, such as those that manipulate persons through subliminal techniques, are expressly prohibited.<sup>68</sup> High-risk practices<sup>69</sup> are permitted but subject to certain mandatory requirements over the design and uses of those practices, as well as assessment after implementation.

Chiu also provides a taxonomy of regulations that emerge when regulation takes a relational form, each with certain strengths and weaknesses. The first is self-regulation by actors, in which essentially there is no outside regulation by authorities.<sup>70</sup> Under some jurisdictional and political approaches, this is the baseline for regulation and often couched in terms of individual freedom. Supporters of self-regulation argue that individuals are in the best position to

---

<sup>65</sup> *Id.*

<sup>66</sup> *Id.*, at 172.

<sup>67</sup> Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Legislative Acts, COM (2021) 206, Brussels 21.4.2021 [hereinafter AIA].

<sup>68</sup> *Id.*, art. 5.1(a).

<sup>69</sup> High-risk systems are listed in Annex III. They are certain systems used in the following areas: biometric identification and categorization of natural persons; management and operation of critical infrastructure; education and vocational training; employment; access to and enjoyment of essential private and public services and benefits; law enforcement; migration, asylum and border control management; and the administration of justice and democratic processes.

<sup>70</sup> Chiu, *supra* note 2, at 173. Other scholars divide regulations along a similar spectrum. For example, Ugo Pagallo and his co-authors refer to three levels of regulation. The first is legal regulation, which is a top-down approach. Second is self-regulation, a bottom-up approach to regulation. In the middle is co-regulation in which legal regulation and self-regulation interact. Ugo Pagallo et al., On Good AI Governance: 14 Priority Actions, a S.M.A.R.T. Model of Governance, and a Regulatory Toolbox, *AI4People*, at 10–11.

decide what is the appropriate action for themselves and will govern themselves because it is in their best interest to do so. Under this view, hard and soft norms enable individuals to solve coordination problems and to avoid getting trapped in prisoner's dilemmas.

A related form of regulation is responsabilization. In the context of the finance industry, responsabilization refers to "the enhancement of awareness of personal responsibility on the part of investors or consumers for their choices and in finding solutions for their investment or financial problems."<sup>71</sup> This approach is closely related to self-regulation in that individuals are given tools and more information to make their own transactional decisions. Responsibilization arises in proposals to educate the public about the benefits and risks of AI applications or requirements to disclose how personal data is used. The presumption is that individuals can then make more informed decisions about using an AI application. The proposed AIA's requirement that users be informed when they are interacting with an AI system also illustrates this approach. Chiu points out, however, that this form of governance works best when there are no information asymmetries and inequality in bargaining power.<sup>72</sup> Such asymmetries, however, almost always exist when technology is involved. Further, Chiu suggests that like self-regulation, responsabilization is individualistic in nature and might not necessarily contribute to the production of public collective goods.<sup>73</sup>

Yet another form of regulation is transactional governance. This is governance by contract. Contract leads to the creation of private law that regulates the relationship between parties. Like self-regulation, contract in the classic form enables individuals to decide for themselves how their economic relationships will operate. In Chiu's view, this kind of governance occurs from the ground up, and norms that emerge from private contract law often become the basis for harder forms of regulation. Yet, transactional governance has its shortcomings. Transactional governance "could be myopic as to wider communitarian effects" and, like self-regulation and responsabilization, it might not be able to deal with externalities.<sup>74</sup> In artificial intelligence, this tension between public and private goods is reflected in non-disclosure agreements between AI vendors and municipalities that prevent officials from disclosing information about algorithmic decision-making in law enforcement or the provision of government services.

---

<sup>71</sup> Chiu, *supra* note 2, at 174.

<sup>72</sup> *Id.*

<sup>73</sup> *Id.*

<sup>74</sup> *Id.*

Self-governance, responsabilization, and transactional governance are often closely tied to the market, which as described above is often believed to create its own norms. Markets can self-regulate as certain practices that emerge from the market become standardized.<sup>75</sup> Such standardization can arise from several sources: individual contracts, as described above, or from voluntary trade associations and organizations. Using the market framework, Chiu points out that there are supply-side and demand-side aspects to standardization. On the supply side, industries might have an incentive to create their own standards to stave off rules imposed by government agencies.<sup>76</sup> Part of this is motivated by the desire to adopt standards that are technically appropriate to the industry. Chiu continues that on the demand side for regulation, these standards, many of which come from trade associations, might enjoy greater legitimacy from the perspective of the regulated because the standards are considered more technologically relevant than regulations imposed by outsiders.<sup>77</sup> At the same time, as discussed above, standards that emerge solely from market forces can result from races to the top or races to the bottom.<sup>78</sup> Therefore there is no assurance that these norms will be optimal for the broader public.

Second, the market can also be seen as exerting disciplinary forces. Chiu has in mind market participants in finance such as bondholders, other creditors, and shareholders who have an incentive to ensure financial institutions avoid undue risk.<sup>79</sup> This is true also for the investors in major technology companies. One issue of course is whether such financial stakeholders have enough power relative to management to steer corporate policy in meaningful ways. Obviously, such parties do have some measure of influence—as an example, shareholders’ resolutions are a means to express concerns about adverse effects of technology.<sup>80</sup> At the same time, such activities have their limits. As Kate Neville et al. discuss, the distance between individual investors and the place where corporate policy is made, the way in which resolutions are proposed and management responds to them, and the way in which corporate capital and voting rights are structured, lessen shareholder influence.<sup>81</sup> Min-Dong Paul Lee and Michael Lounsbury point out that whether corporations are responsive to shareholder demands depends in part on whether they are subject to high disruption costs from shareholder activism, whether they are dependent

---

<sup>75</sup> *Id.*, at 175.

<sup>76</sup> *Id.*

<sup>77</sup> *Id.*

<sup>78</sup> *Id.*

<sup>79</sup> *Id.*, at 175–76.

<sup>80</sup> Such resolutions are discussed in Chapter 3.

<sup>81</sup> Kate J. Neville et al., *Can Shareholder Advocacy Shape Energy Governance? The Case of the US Antifracking Movement*, 26 REV. OF INT’L POL. ECON. 104 (2019).

on their reputation, and whether their business models place them closer to end-user consumers.<sup>82</sup> These dynamics will be explored further in Chapter 3.

Third, Chiu points that stock exchanges have their own sets of rules that have some disciplinary effects on companies.<sup>83</sup> This is relevant to this study because most of the major technology companies are public companies. Moreover, as discussed above, trade and voluntary organizations and the norms they develop could be said to be market-based. Such organizations can perform a regulatory function through membership rules, denials of access to organization benefits, reviews, and reputational effects. At the same time, Chiu refers to studies that indicate that companies can use membership in trade organizations to camouflage bad behavior.<sup>84</sup> Companies can join trade organizations to avoid the direct gaze of regulators. And, when there are competing organizations, they might be tempted to lower standards to attract new membership.<sup>85</sup>

Finally, if regulation can be understood as forming its own market, stakeholders such as ratings agencies, lawyers, and others can be seen as suppliers of governance tools.<sup>86</sup> These actors facilitate transactions but also monitor them. Proponents of these actors as informal regulators argue that they can provide expertise and counter information asymmetries by providing certifications of creditworthiness and the like. This approach is reflected in proposals to have third-party audits of algorithms or the results of algorithmic decisions. Moreover, broader participation lends greater legitimacy to regulatory activities. But Chiu has two concerns about using third parties in this way. One is that formal regulators might cede the field to surrogate regulators, thus lessening the advantage of having more regulatory parties involved.<sup>87</sup> The other is that competitive incentives can weaken the incentives of such gatekeepers to remain neutral when participating in a business transaction.<sup>88</sup>

In reviewing these four facets of market-based regulation, Chiu points out that although market-based governance has its place, the fundamental issue is whether an irresolvable tension exists between governance for the delivery of public goods, which must take into account a number of factors and imperatives on the one hand, and market-oriented governance, which is largely

---

<sup>82</sup> Min-Dong Paul Lee and Michael Lounsbury, *Domesticating Radical Rant and Rage: An Exploration of the Consequences of Environmental Shareholder Resolutions on Corporate Environmental Performance*, 50 *Bus. & Soc'y* 155 (2011).

<sup>83</sup> Chiu, *supra*, note 2, at 176.

<sup>84</sup> *Id.*

<sup>85</sup> *Id.*

<sup>86</sup> *Id.*

<sup>87</sup> *Id.*, at 174–75.

<sup>88</sup> *Id.*



motivated by profit and decreasing costs.<sup>89</sup> But the tension may be a necessary evil. If rules, standards and other governance devices are themselves subject to a market for rules, in a sense they compete with each other. Rules deemed inefficient by some criteria will fall into disuse. The Delaware Corporation Code is a good example of this kind of ‘competition.’ More than half of all corporations in the United States are incorporated under Delaware law, which is seen as being amenable to the needs of corporate management. This widespread use has enabled the Delaware corporate law to enjoy economies of scale, and the law is supported by judges who are experts in corporate law. This in turn has created case law that in turn allows for greater predictability for corporations.

### **Enforced Self-regulation**

So far, this section has discussed forms of self-regulation that are affected to some extent by market imperatives and actors within the market. Chiu goes on to describe other forms of governance, all of which involve to some degree the regulated. In all of them, the government provides certain guidelines, standards, or parameters, to which the regulated adhere as they regulate themselves. One form is what Chiu terms enforced self-regulation.

Enforced self-regulation allows the regulated to set out the rules by which it is bound. The regulator vets and approves those rules in accordance with minimum standards and regulatory objectives. The regulator is then entitled to inspect the regulated to see if compliance is taking place and is able to enforce against the regulated where the rules are breached.<sup>90</sup>

Enforced self-regulation has the advantage of allowing regulated parties to use their expertise in devising rules that are better targeted toward the regulated parties’ reality. It is also seen as better than pure self-regulation because the government does play a supervisory role.<sup>91</sup> At the same time, the issue of expertise arises again in this regulatory paradigm; regulators will be tempted to defer too much to the industry if the regulated domain is highly technical. This runs the danger of returning to self-regulation with the problems discussed in the previous section. Further, firms often adopt practices of other competitors, and there is no guarantee that such practices are optimal from the point of view of the public.<sup>92</sup>

---

<sup>89</sup> *Id.*, at 178.

<sup>90</sup> *Id.* (footnote omitted).

<sup>91</sup> *Id.*

<sup>92</sup> *Id.*

Recent work by Nolan McCarty supports these concerns. McCarty proposes a model for the regulation of complex domains, in which the outcome of such regulation tends to be those adopted by the regulated firms themselves. In complex industries it is difficult for regulators to find independent sources of expertise largely because experts stand to earn more by working in the private sector. This is the case in artificial intelligence, as will be discussed in Chapter 4. This lack of expertise on the part of regulators means that it is difficult to predict the impact of a regulation. As a result, regulators become heavily dependent on the regulated industry itself and on its willingness to self-regulate.<sup>93</sup> “Unless the agency is willing and able to commit significant resources to building its own expertise, it can learn only through monitoring the firm’s activities.”<sup>94</sup>

McCarty continues that “public regulation generally supplements and builds upon the efforts of firms to self-regulate.”<sup>95</sup> In this regard, he observes that firms tend to have strong incentives to engage in such self-regulation through internal measures, trade associations, and the like. This creates a dilemma between the regulating agency and the firm’s efforts at self-regulation. Because the agency lacks expertise, it becomes rational to either allow the industry to regulate itself, or to impose easy-to-administer bright-line rules such as bans.<sup>96</sup> The latter move, however, can be inefficient (again reaffirming the first prong of Collingridge’s dilemma) and has the effect of dampening efforts of the industry to regulate itself since the agency has imposed its own set of rules.<sup>97</sup> The result often is to defer to the industry’s rules or to create regulatory agencies that are pro-industry. According to McCarty, the latter happens when expertise and training are only available through the industry itself or through schools that cater to industry interests.<sup>98</sup>

As a way of addressing the dilemma created by asymmetries in knowledge and expertise, governance can take the form of meta-regulation. This is an attempt to affect and align the internal values and principles of private actors with public ones. According to Chiu, emphases on corporate governance, risk management, ethical business practices and corporate social responsi-

---

<sup>93</sup> Nolan McCarty, *The Regulation and Self-Regulation of a Complex Industry*, 79 J. POL. 1220 (2017).

<sup>94</sup> *Id.* at 1221. On the role of consumers in influencing corporate behavior, see DIETLAND STOLLE AND MICHELE MICHELETTI, *POLITICAL CONSUMERISM: GLOBAL RESPONSIBILITY IN ACTION* (2013).

<sup>95</sup> McCarty, *supra* note 93.

<sup>96</sup> *Id.*, at 1222.

<sup>97</sup> *Id.*, at 1225, 1232.

<sup>98</sup> *Id.*, at 1228, 1232.

bility can be seen as manifestations of this approach.<sup>99</sup> An emphasis on risk management, for example, can be seen as providing various principles for sound business practices such as “the establishment of a compliance function, risk-management systems, internal audit, and the responsibility of senior management for compliance generally.”<sup>100</sup> In data privacy, the seven principles set out in Article 5 of the General Data Protection Regulation—lawfulness, fairness and transparency; the purpose limitation; data minimization; accuracy; the storage limitation; integrity and confidentiality; and accountability on the part of a data controller—can be seen as a form of meta-regulation. Chiu has several criticisms of this approach. Like enforced self-regulation, meta-regulation can be captured by industry if the regulators are overly deferential to the expertise of the regulated.<sup>101</sup> Such approaches then create a ‘presumed legitimacy’ of the regulatory scheme.<sup>102</sup> Such an approach can lead to an overemphasis on process and learning. Further, because the locus of regulation rests largely in the firm, this can lead to “microcosms” of inward-looking firms, with little concern for larger systemic or public effects.<sup>103</sup>

Closely related to meta-regulation is ethical self-regulation. In for-profit companies, this takes the form of business ethics.<sup>104</sup> Ethical self-regulation is pertinent to AI governance because, as discussed above, AI ethics has become a major, if not the predominant, approach to guiding the development and deployment of AI applications, particularly within private firms. Jeffrey Moriarty describes the field as concerned with several issues,<sup>105</sup> several of which are relevant to the decisions of technology companies that are developing artificial intelligence. There is a general debate whether the purpose of corporate governance is to further the interests of shareholders, or whether other stakeholders need to be considered as well. There are also questions about a business’s obligations to consumers, so that in some cases it might be necessary to decline offering certain products or services. There are larger questions about the proper role of the firm in society. Corporate social responsibility,

---

<sup>99</sup> Chiu, *supra*, note 2, at 179.

<sup>100</sup> *Id.* (footnote omitted).

<sup>101</sup> *Id.*

<sup>102</sup> According to Chiu, in corporate governance, such legitimacy results from ‘legal endogeneity’, the “self-legitimizing effect of corporations’ implementation of their compliance systems, resulting in *de facto* self-regulation.” Iris H-Y Chiu, *An Institutional Theory of Corporate Regulation*, 71 CURRENT LEGAL PROBS. 279, 315 (2018) (footnote omitted).

<sup>103</sup> Chiu, *supra*, note 2, at 179.

<sup>104</sup> *Id.*, at 180.

<sup>105</sup> Jeffrey Moriarty, *Business Ethics*, in THE STANFORD ENCYCLOPEDIA OF PHILOSOPHY (Edward N. Zalta ed., Summer 2021 ed.), <https://plato.stanford.edu/entries/ethics-business/>.

which will be examined in Chapter 3, falls into this area of business ethics. The extent to which corporations should participate in the political process is another issue.<sup>106</sup> Chiu and others argue, however, that business ethics alone is an inadequate form of governance. In part, the critique is the same as the critiques of enforced self-regulation and meta-regulation. Business ethics has the advantage of acknowledging the ubiquity of firms and their importance in a modern economy. Yet, questions such as that of shareholder primacy and the role of the firm in the larger society remain hotly contested. Moreover, with for-profit companies, there will always be the drive to maximize profits and minimize costs. As discussed, market dynamics produce strong incentives that could run counter to larger societal goals. In this regard, Chiu writes, “Internal dispositions towards moral behavior are often challenged by countervailing forces, such as rational calculations in game scenarios, organizational indifference or the market perceptions of the value of such behavior.”<sup>107</sup>

Finally, there is regulator-led regulation, the regulation that we tend to think of when using that term. Here, regulation is primarily the responsibility of regulatory agencies, which promulgate rules, standards, and other norms to which the regulated must conform.<sup>108</sup> Such regulations are of a public nature because government agencies are accountable to a broader range of stakeholders. Yet, as Chiu points out, private firms often have great influence in the formation of these ‘public’ rules. As will be seen later in this book, large, regulated businesses or associations comprising private businesses expend significant resources to influence rule-making processes.

### Complex Interactions in Response to Complex Technology

In the regulatory space set out in the previous subsection, the participants described in this chapter are engaged in a dynamic relationship, “co-producing and co-enforcing norms of governance,”<sup>109</sup> so that the various forms of regulation are no longer exogenous to a system, but emerge from the participants themselves, each with particular interests and concerns, some overlapping and others potentially conflicting. Moreover, as Chiu argues, it is possible for any regulatory space to be dominated by one or more participants, and thus, not only determine which regulatory ‘heuristic’ will be used, but in theory come to dominate all heuristics, hence her recommendation for the financial services

---

<sup>106</sup> *Id.*

<sup>107</sup> Chiu, *supra*, note 2, at 180.

<sup>108</sup> *Id.*

<sup>109</sup> *Id.* at 170.

industry that other stakeholders be strengthened so that they can participate in the formation of these norms for financial governance.<sup>110</sup>

The emergence of artificial intelligence technologies and their regulation are similar in this regard. The nature of advanced technologies in general and artificial intelligence in particular, in addition to the power of some of the actors, lend themselves to regulatory arrangements in which the lines between regulator and regulated and between public and private functions become indistinct.<sup>111</sup> Roxana Radu points out that this mixing of roles poses challenges for the governance of artificial intelligence. Radu surveys several national artificial intelligence strategies and observes how these blurred lines come into play.<sup>112</sup> A number of the strategies contemplate hybrid forms of governance in which a number of actors participate in the regulation of AI.<sup>113</sup> For Radu, such actors are informed by different institutional logics, discussed above.<sup>114</sup> In Radu's view, one of the results of this hybridity is what she terms "functional indeterminism." This appears to be the inability to determine what role a particular organization or institution is to play because that organization and institution might have several roles relevant to governance, with each role informed by different institutional logics. This can have several effects.

The blurring boundaries between the public and the private may give rise to new institutions and to redefined functions. The ambiguity and uncertainty related to the future of AI may structure interactions as a way to preserve freedom of action. From this perspective, the process of function indetermination reflects a decision to allow for open-ended possibilities: 'tasks so far assigned to the polity can be transposed with increasing ease to a web of "authorities" created for the purpose of making decisions on technical and scientific issues.'<sup>115</sup>

Radu observes that although, as articulated in the national strategies she reviewed, governments indeed see roles for themselves in terms of oversight,

---

<sup>110</sup> *Id.* at 301.

<sup>111</sup> Brian Balogh points out, however, that in the United States there has been a long history of shifting governmental functions to the private sector. BRIAN BALOGH, *THE ASSOCIATIONAL STATE: AMERICAN GOVERNANCE IN THE TWENTIETH CENTURY* (2015). *See also*, CATHERINE E. RUDDER, A. LEE FRITSCHLER, AND YON JUNG CHOI, *PUBLIC POLICYMAKING BY PRIVATE ORGANIZATIONS: CHALLENGES TO DEMOCRATIC GOVERNANCE* (2016) (arguing that private organizations can form a 'fifth' branch of government that creates policy).

<sup>112</sup> Roxana Radu, *Steering the Governance of Artificial Intelligence: National Strategies in Perspective*, 40 *POL'Y & Soc'y* 178 (2021).

<sup>113</sup> *Id.*, at 182–83.

<sup>114</sup> *Id.*, at 183–84.

<sup>115</sup> *Id.*, at 183–84 (citation omitted).

control, and the promotion of AI technologies, “[y]et all strategies remain vague on the concrete measures enabling them to act in these roles.”<sup>116</sup>

## MULTISTAKEHOLDER PARTICIPATION AS A GOVERNANCE TOOL

Thus far, the chapter has discussed different types of regulatory approaches outlined by Chiu and other scholars. Chapter 1 discussed how recent approaches to the governance of technology try to resolve the second horn of Collingridge’s dilemma by allowing early interventions that allow all relevant stakeholders to be consulted before action is taken to develop a particular technology. This approach is not limited to the governance of technology; it is part of a much larger trend, one in which companies,<sup>117</sup> states, and international organizations have tried to incorporate aspects of this participatory process for different purposes. Stakeholder governance has been used to build value for business firms and to support democratic governance on the global level.<sup>118</sup>

Stakeholder engagement has been framed in strategic, moral, and pragmatic terms.<sup>119</sup> Barbara Gray and Jill Purdy describe multistakeholder partnerships as a means to collaborate. It is “a process that engages a group of autonomous stakeholders interested in a problem or issue in an interactive deliberation using shared rules, norms, and structures, to share information and/or to take coordinated actions.”<sup>120</sup> Stakeholders in turn are “all individuals, groups, or organizations that can directly influence the ... problem or issue by their actions or are directly influenced by actions others take to address it.”<sup>121</sup> Such stakeholders can be grouped generically into businesses, governments, non-governmental organizations, and civic society.<sup>122</sup> Several features of this approach are attractive. Some problems are complex and benefit from different perspectives to appropriately address them.<sup>123</sup> Similar to Abbott and Snidal, Gray and Purdy point out that state partnerships with NGOs and the private

---

<sup>116</sup> *Id.*, at 188.

<sup>117</sup> See Johanna Kujala et al., *Stakeholder Engagement: Past, Present, and Future*, 61 BUS. & SOC’Y 1136 (2022) (discussing stakeholder engagement and the governance of businesses).

<sup>118</sup> TERRY MACDONALD, GLOBAL STAKEHOLDER DEMOCRACY: POWER AND REPRESENTATION BEYOND LIBERAL STATES (2008).

<sup>119</sup> Kujala et al., *supra* note 117, at 1152–55.

<sup>120</sup> BARBARA GRAY AND JILL PURDY, COLLABORATING FOR OUR FUTURE: MULTISTAKEHOLDER PARTNERSHIPS FOR SOLVING COMPLEX PROBLEMS 1 (2018) (citation omitted).

<sup>121</sup> *Id.*, at 2.

<sup>122</sup> *Id.*, at 3.

<sup>123</sup> The authors refer to these as “wicked problems.” *Id.*, at 15.

sector might increase capacity for the development of policies that in turn can increase their legitimacy. NGOs can benefit from help with organizational development, resources, and “increased recognition and status whereby the organization’s political influence can be strengthened.”<sup>124</sup> By engaging in community involvement, businesses might gain benefits such as “improved reputation and brand value, license to operate, innovation, development of local markets, improved risk management, and greater ability to attract, and retain their employees.”<sup>125</sup>

The authors point out, however, that multistakeholder processes have challenges. Stakeholders might have a history of conflict that has generated distrust among them. They might have different ways of viewing a particular issue. Other times, stakeholders are required by law to engage in this process, so there is no internal impetus for collaboration. Stakeholders often have different senses of identities and values. They might differ in terms of risk assessment and tolerance. There might be a reluctance to trade limited resources needed to reach agreement. Differences in power might leave some stakeholders feeling marginalized, thus causing them to withdraw from the process.<sup>126</sup> Since it can be argued that stakeholder governance has become an international norm for addressing AI governance issues at the global level some appreciation for these strengths and weaknesses is warranted.

## SPECIFIC AI REGULATORY TOOLS

Chiu’s taxonomy of regulations provides a framework for understanding the kinds of norms that are emerging in AI governance. These range from forms of self-regulation to conventional law, and there is a wide array of approaches. In a recent study, Carlos Gutierrez and Gary Marchant identified 634 soft law ‘programs’ that are used for the governance of artificial intelligence. In order of frequency, Gutierrez and Marchant find that soft law takes the form of recommendations and strategies, principles, standards, professional guidelines and codes of conduct, partnerships, certification or voluntary programs, voluntary moratoriums, and finally, bans.<sup>127</sup> Recommendations are “suggestions, proposals, or evidence-based actions meant to improve an organization’s status quo.”<sup>128</sup> Strategies are “roadmaps that highlight the direction an entity

---

<sup>124</sup> *Id.*, at 5 (citations omitted).

<sup>125</sup> *Id.* (citations omitted).

<sup>126</sup> *Id.*, at 97–104. *See also* Kujala et al., *supra* note 117, at 1165–68 (describing similar problems with the stakeholder approach).

<sup>127</sup> Carlos Ignacio Gutierrez and Gary Marchant, *A Global Perspective of Soft Law Programs for the Governance of Artificial Intelligence* (2021), at 11.

<sup>128</sup> *Id.*

wishes to or should pursue.”<sup>129</sup> Principles are “broad statements that serve as high level norms.”<sup>130</sup> Standards are developed by standard-setting organizations and address a technical need in the field. Twelve such organizations were responsible for 60 standards.<sup>131</sup> Professional guidelines and codes of conduct set out expectations for individuals who work with AI applications or methods.<sup>132</sup> Partnerships are alliances between two or more entities to advance some agenda regarding AI.<sup>133</sup> A certification program “indicates compliance to a set of pre-defined characteristics,”<sup>134</sup> while a voluntary program is “a government initiative that invites non-governmental entities ... to comply with a non-binding set of actions or guidelines.”<sup>135</sup> Finally, a voluntary moratorium or ban is “characterized by a call of action to avoid or cease the usage of an AI application or method.”<sup>136</sup>

Most of these governance programs are recommendations and strategies (54.26% of total programs identified) and principles (25%). Standards make up about 9.5 percent of programs, and professional guidelines or codes of conduct, partnerships, certification or voluntary programs, and voluntary moratoriums and bans were significantly less frequent (3.6%, 3.3%, 2.5%, and 1.9% respectively).<sup>137</sup> Most of them are recent, with the bulk of such programs emerging between 2016 and 2019.<sup>138</sup> The large majority of the programs (82%) have been adopted for both internal and external influence on stakeholders.<sup>139</sup> However, in contrast to other forms of soft law, recommendations and strategies tend to be internally focused, but there are some that are both inside- and outside-facing.<sup>140</sup>

As one studies these types of programs, one begins to see that the governance of artificial intelligence takes on a fractal nature. At any given level of organization, from the firm to the international level, the same basic tools are being used or proposed, either individually or in combination with one another. The book will discuss several of these tools in more detail in the following chapters, but it is useful here to identify some of them and to outline their features.

---

<sup>129</sup> *Id.*

<sup>130</sup> *Id.*, at 12.

<sup>131</sup> *Id.*

<sup>132</sup> *Id.*

<sup>133</sup> *Id.*

<sup>134</sup> *Id.*, at 13.

<sup>135</sup> *Id.*

<sup>136</sup> *Id.*

<sup>137</sup> *Id.*, at 11.

<sup>138</sup> *Id.*, at 35.

<sup>139</sup> *Id.*, at 11.

<sup>140</sup> *Id.*



## Statements of Ethics or Principles

As discussed above, statements of ethics or principles are high-level norms, and according to Gutierrez and Marchant's survey, are the second-most common type of soft law. There are a plethora of such statements for AI (Gutierrez and Marchant count 158 of them) and they have been written by all the actors named in this chapter. Specific statements of principles will be discussed in later chapters of this book. Here what should be noted is that "this intense interest from such a broad range of stakeholders reflects not only the need for ethical guidance, but also the desire of those different parties to shape the 'ethical AI' conversation around their own priorities."<sup>141</sup>

Despite their large number, Jessica Morley and her co-authors argue that broad themes can be derived from these statements, even though there will be argument whether these themes are exhaustive.

[E]thically aligned [machine learning is] that which is (a) beneficial to, and respectful of, people and the environment (beneficence); (b) robust and secure (non-maleficence); (c) respectful of human values (autonomy); (d) fair (justice); and (e) explainable, accountable and understandable (explicability).<sup>142</sup>

In a similar study, Kai Jia and Nan Zhang reviewed 123 ethics guidelines proposed by different stakeholders. They identified seven areas of concern addressed in the guidelines: transparency and trust, bias and equity, privacy, imputation (i.e., accountability), human autonomy, robustness, and social security.<sup>143</sup> If researchers such as Morley and her co-authors and Jia and Zhang

---

<sup>141</sup> Jessica Morley et al., *From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods, and Research to Translate Principles into Practices*, 26 SCI. & ENG. ETHICS, 2141, 2144–45 (2020), citing Anna Jobin, Marcello Ienca, and Effy Vayena, *Artificial Intelligence: The Global Landscape of Ethics Guidelines*, 1 NATURE MACHINE INTELLIGENCE 389 (2019); Luciano Floridi, *Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical*, 32 PHIL. & TECH. 185 (2019). See also, Luciano Floridi and Josh Cowls, *A United Framework of Five Principles for AI in Society*, in ETHICS, GOVERNANCE, AND POLICIES IN ARTIFICIAL INTELLIGENCE (Luciano Floridi ed., 2021) (articulating the five principles); Jessica Fjeld et al., *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI*, Berkman Klein Center for Internet & Society at Harvard University Res. Pub. Ser. No. 2020-1 (Jan. 15, 2020) (finding eight key themes from a survey of sets of principles: privacy, accountability, safety and security, transparency and explainability, fairness and non-discrimination, human control of technology, professional responsibility, and promotion of human values).

<sup>142</sup> Morley et al., *supra* note 141, at 2145 (emphasis omitted).

<sup>143</sup> Kai Jia and Nan Zhang, *Categorization and Eccentricity of AI Risks: A Comparative Study of the Global AI Guidelines*, 32 ELECTRONIC MARKETS 59, 64–66 (2022).

are correct, it can be argued that despite the various motives of actors who have produced these statements, one can identify broad sets of values that can inform the development and deployment of artificial intelligence. Gutierrez and Marchant point out, however, that principles and guidelines are criticized because they are voluntary.<sup>144</sup> Moreover, by definition, principles are abstract, thus it can be hard to determine compliance with them. Further, as Morley and her co-authors note, principles can be difficult to implement as AI systems are designed.<sup>145</sup> Finally, Jia and Zhang suggest that although statements of principles are often expressed in broad terms, they might not necessarily be comprehensive in scope. In their study, they evaluated how well the statements address risks that are posed by artificial intelligence. They found that “extant AI guidelines do not cover all risks evenly but have eccentric distribution.”<sup>146</sup> In particular, “[c]ollective risk and generational risk are substantially underestimated and even ignored by different stakeholders.”<sup>147</sup>

Despite these limitations, statements of artificial intelligence ethics or principles have value. They can represent a public commitment by the actors who adopt them, thereby making adopters accountable for fulfilling them. Moreover, although they are voluntary, they can nevertheless serve as a framework for organizational behavior.<sup>148</sup> Daniel Schiff and his co-authors add that these statements can impact other ethics documents, which indeed cite one another; they can influence other parties to conform to their principles; shape public discourse about issues addressed by the statements; inform harder forms of regulation as they are being developed; and also shape the “stakeholder ecosystem.”<sup>149</sup> Concerning this ecosystem, the authors have in mind policies that will educate the public and other participants in AI policy.<sup>150</sup> In addition, these statements often call for broad participation in the development of artificial intelligence, thus further identifying stakeholders as legitimate actors in the formation of policy and further embedding the norm of stakeholder participation.

---

<sup>144</sup> Gutierrez and Marchant, *supra* note 127, at 16.

<sup>145</sup> Morley et al., *supra* note 141, at 2147.

<sup>146</sup> Jia and Zhang, *supra* note 143, at 67.

<sup>147</sup> *Id.* For the authors, collective risk refers to risks posed to groups as opposed to risks to individuals. *Id.*, at 63, 66. Generational risk refers to risks posed by current levels of technology, as opposed to transgenerational risk, which has to do with future developments of artificial intelligence, such as the achievement of general intelligence. *Id.*

<sup>148</sup> Accord, Daniel Schiff et al., *What’s Next for AI Ethics, Policy and Governance? A Global Overview*, AIES ’20, at 157 (Feb. 7–8, 2020).

<sup>149</sup> *Id.*, at 156–57.

<sup>150</sup> *Id.*

## Bans

On the opposite end of the spectrum, bans are arguably the starkest form of regulation, and it is not surprising that they seem to be the last resort for regulatory approaches. At the same time, bans have indeed been imposed in connection with artificial intelligence applications. Several municipalities in the United States have prohibited the use of facial recognition for law enforcement purposes.<sup>151</sup> As discussed above, the proposed EU Artificial Intelligence Act would forbid the use of artificial intelligence to engage in social or individual manipulation. In Chapter 1, it was reported that efforts are being made to outlaw the use of lethal autonomous weapons.

Like all regulatory tools, bans have their strengths and weaknesses. They are most effective when there is a clear consensus that a practice should be prohibited. In those circumstances, a ban expresses a community's opposition to that activity. Prohibiting the use of machine learning techniques to subliminally manipulate human decisions or to assist in the cloning of human beings are examples. Further, a ban often has the effect of curbing undesired behaviors. To take an example outside of artificial intelligence, bans on smoking in restaurants reduce the effects of second-hand smoke. At the same time, a ban poses the risk of being under- or overinclusive. H.L.A. Hart's famous example of the ordinance, "no vehicles in the park," illustrates the problem because it is not obvious what is meant by the term 'vehicle.' Likewise, a ban on the use of facial recognition techniques for law enforcement purposes or a ban on the manipulation of human decisions will need to be interpreted as it is applied.<sup>152</sup> Finally, a ban can be ineffective if it is not broadly supported, and even if it is, might have unintended consequences that in some cases undermine the broader purpose of the ban.<sup>153</sup>

---

<sup>151</sup> According to the website, Ban Facial Recognition, in the United States, as of this writing 20 municipalities, one county, and one state had imposed such a ban. Ban Facial Recognition, Interactive Map, <https://www.banfacialrecognition.com/map/>.

<sup>152</sup> For the constitutional issues raised by bans, see Joseph Blocher, *Bans*, 129 YALE L. J. 308 (2019).

<sup>153</sup> See, e.g., Diana S. Weber et al., *Unexpected and Undesired Conservation Outcomes of Wildlife Trade Bans—An Emerging Problem for Stakeholders?* 3 GLOB. ECOL. & CONSERVATION 389, 391 (2015) (citing literature that argues that sometimes the international ban on trafficking in wildlife parts or products increases exploitation). Much will depend on the ban in question. For example, the popular assessment of the Great Prohibition in the United States is that it was largely unsuccessful in reducing alcohol consumption. However, recent studies have argued that the alcohol ban did in fact lead to less consumption while still retaining public support. For discussions of the conventional view of the Prohibition and its reappraisal, see Jack S. Bloker, Jr., *Did Prohibition Really Work? Alcohol Prohibition as a Public Health Innovation*, 96

## Data and the Training of Algorithms

As discussed in Chapter 1, several artificial intelligence techniques involve machine learning, which in turn depends on models that have been trained or improved via exposure to data sets. Some models of artificial intelligence governance recommend oversight over the data used to train these algorithms as well as over the design and training of the models themselves. For example, Singapore lists a set of practices that govern this process. With data, this includes practices to understand the lineage or provenance of data; to ensure data quality; to minimize inherent bias in data sets, including selection bias and measurement bias; to use different data sets for training, testing and validation of the models; and to review and update data sets.<sup>154</sup> With regard to the models and the algorithms used to develop them, several measures can be taken. These include techniques to enhance explainability, repeatability, robustness, regular tuning, traceability, and auditability.<sup>155</sup> Explainability and auditability will be discussed in more detail below. Repeatability refers to the ability of an AI system to reproduce the same result when presented with the same data or scenario.<sup>156</sup> Robustness is the degree to which an AI system functions as desired when presented with invalid data.<sup>157</sup> Regular tuning refers to practices to refresh models with new data as they emerge.<sup>158</sup> Traceability concerns documenting an AI model's decisions and "the datasets and processes that yield the AI model's decisions (including those of data gathering, data labelling and the algorithms used)...."<sup>159</sup> As a form of Chiu's enforced self-regulation, an AI regulatory regime could require developers to adopt some or all of these measures in the design, development, and deployment of AI systems. Further, as will be explained in Chapter 3, techniques are being developed to perform some of these functions through software itself. Questions, however, arise concerning the feasibility and the tradeoffs required of these measures, as will be discussed below.

---

AM. J. PUB. HEALTH 233 (2006); Wayne Hall, *What Are the Policy Lessons of National Alcohol Prohibition in the United States, 1920–1933?* 105 ADDICTION 1164 (2010).

<sup>154</sup> SG:D, Infocomm Media Development Authority and Personal Data Protection Commission Singapore, Model Artificial Intelligence Governance Framework (2d ed. 2020), at 36–40 [hereinafter Model AI Governance Framework], <https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.ashx>.

<sup>155</sup> *Id.*, at 43–51.

<sup>156</sup> *Id.*, at 46.

<sup>157</sup> *Id.*, at 47.

<sup>158</sup> *Id.*, at 48.

<sup>159</sup> *Id.*, at 48–49.

## Human Control

A major concern about artificial intelligence systems is that they will be capable of taking and carrying out decisions without human intervention. Although decision-making systems continue to improve in their ability to assess states of the world and act on them, for the present, this ability remains limited. In addition to limitations in current AI techniques, there is a worry that AI systems will make decisions that are inconsistent with human values, even though those decisions are ‘rational’ from the frame of reference of the model in question.

To address these concerns, several commentators and regulatory schemes urge or require that a human remain in the decision-making process.<sup>160</sup> Human control is intended to avoid mishaps caused by algorithmic decision-making and to increase public trust in artificial intelligence.<sup>161</sup> This tool is ubiquitous in AI regulation. It informs current requirements that human drivers remain at the wheel in autonomous vehicles and the principle that autonomous weapons systems must be designed so that a human remains in control of targeting and other tactical decisions. Further, many AI systems are being designed to work with human beings on certain tasks, so that as a matter of course humans will participate in decisions that are aided by artificial intelligence. Finally, human collaboration or intervention can also take place during the machine learning process.<sup>162</sup>

The main drawback with keeping a human in the decision loop, however, is that some systems might be hampered by human intervention. For example, driver error is by far the primary cause of automobile accidents, and an expected benefit of autonomous vehicles is that accidents will decrease precisely because human drivers will no longer control most driving operations. For some systems, perhaps in health care, the increased trust and enhanced performance that comes from human involvement could outweigh the costs of human failure to adequately supervise. Thus, for system designers and regula-

---

<sup>160</sup> See, e.g., Ross P. Buckley et al., *Regulating Artificial Intelligence in Finance: Putting the Human in the Loop*, 43 SYDNEY L. REV. 43 (2021).

<sup>161</sup> Naomi Aoki, *The Importance of the Assurance that “Humans Are Still in the Decision Loop” For Public Trust in Artificial Intelligence: Evidence from an Online Experiment*, 114 COMPUTERS IN HUM. BEHAV. art. no. 106572 (2020).

<sup>162</sup> Tor Grønsund and Margunn Aanestad, *Augmenting the Algorithm: Emerging Human-in-the-Loop Work Configurations*, 29 J. STRAT. INFO. SYS. 101613 (2020); Julia Ostheimer, Soumitra Chowdhury, and Sarfraz Iqbal, *An Alliance of Humans and Machines for Machine Learning: Hybrid Intelligent Systems and their Design Principles*, 66 TECH. IN SOC<sup>3</sup>Y art. no. 101647 (2021).

tors, the task is to identify when a system requires human participation<sup>163</sup> and then to develop strategies to reduce the possibility of human error when they are kept in the loop.<sup>164</sup>

### Algorithmic Impact Assessments and Audits

As discussed in Chapter 1, one approach to Collingridge's dilemma is to assess the impact of a technology as soon as possible during its lifecycle. Chiu discussed earlier that relational regulation often is accompanied by risk-based methods of regulation. Various regulatory schemes now urge that algorithmic impact statements be required in the design process, akin to impact assessments on finance, the environment, data privacy protection, and human rights that are mandated in other regulatory regimes. These assessments try to predict and to prevent or mitigate possible adverse impacts of a contemplated use of artificial intelligence. Further, assessments provide information that can be used for future planning.<sup>165</sup> As Andrew Selbst points out, they can vary in their complexity from processes akin to an environmental impact statement to questionnaires.<sup>166</sup> In Canada, such an impact assessment is mandatory for electronic services provided by the federal government.<sup>167</sup> A similar assessment is voluntary for US federal agencies.<sup>168</sup> As of this writing, the US Congress is considering the Algorithmic Accountability Act, which, if enacted, would require large firms and other entities to assess and to mitigate the impacts of certain algorithmic decision-making systems that are being developed or are already in service.<sup>169</sup>

Proponents of assessments urge that if designed well, they allow for broad-based participation before an algorithmic system is adopted—again, an affirmation of stakeholder processes.<sup>170</sup> However, they raise several issues.

---

<sup>163</sup> Singapore suggests using a risk analysis to determine the level of human control needed over an AI system. Model AI Governance Framework, *supra* note 154, at 31–32.

<sup>164</sup> Lorrie Faith Cranor, *A Framework for Reasoning About the Human in the Loop*, Proc. 2008 UPSEC'08 PROC. CONF. ON USABILITY, PSYCH., AND SECURITY.

<sup>165</sup> Andrew D. Selbst, *An Institutional View of Algorithmic Impact Assessments*, 35 HARV. J. L. & TECH. 117, 140 (2021).

<sup>166</sup> *Id.*, at 139–40.

<sup>167</sup> Directive on Automated Decision-Making, sec. 6.1., <https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592>.

<sup>168</sup> CIO.gov, Algorithmic Impact Assessment, <https://www.cio.gov/aia-eia-js/#/>.

<sup>169</sup> Algorithmic Accountability Act of 2022, S.B. 3572, 117<sup>th</sup> Cong., 2d. Sess. (2022).

<sup>170</sup> Emanuel Moss et al., *Assembling Accountability: Algorithmic Assessment for the Public Interest*, Data and Society, (2021), at 5.

According to Jacob Metcalf and Emanuel Moss, there is no consensus on how to “do [an] assessment, how to report the impacts of that assessment in ways that lead to democratic governance of algorithmic systems, or how to effectively and fairly integrate the interests of impacted communities in such practices.”<sup>171</sup> With regard to ‘doing’ assessments, Josephine Yam and Joshua Skorburg argue that assessments can be malleable: an assessment of whether a proposed application is fair will of course depend on what is meant by that term.<sup>172</sup> Moreover, assessments can be too narrow or too broad in their reach. For example, Yam and Skorburg are concerned that using algorithms in hiring decisions might violate various human rights concerning labor. However, they are of the view that several existing assessment methods do not adequately address the impacts through the framework of those specific rights, although some do.<sup>173</sup> Selbst points out in his critique of the Canadian questionnaire discussed above, that, if not designed well, questionnaires can close off important avenues of exploration.<sup>174</sup> Further, there continues to be debate about who should have access to these assessments. Assessments that follow the environmental impact statement model would be subject to public comment. Those based on the data privacy assessment model involve a company and its regulators, but not necessarily the public.<sup>175</sup>

### **Explainability, Interpretability, and Transparency**

Another concern with artificial intelligence is that certain types of artificial intelligence techniques, particularly machine learning via neural networks, can lead to inexplicable decisions. Thus, some regulatory schemes, including Article 22 of the General Data Protection Regulation, require that AI applications be designed so that decisions made or recommended by them can be explained.<sup>176</sup> Explainable artificial intelligence, however, poses several

---

<sup>171</sup> Jacob Metcalf and Emanuel Moss, *Raising Algorithmic Impact Assessment*, DATA & SOC’Y: POINTS, Mar. 17, 2022, <https://points.datasociety.net/raising-algorithmic-impact-assessment-1ebc6ba1bbeb>.

<sup>172</sup> Josephine Yam and Joshua August Skorburg, *From Human Resources to Human Rights: Impact Assessments for Hiring Algorithms*, 23 ETHICS & INFO. TECH. 611, 619 (2021).

<sup>173</sup> *Id.*, at 618–20.

<sup>174</sup> Selbst, *supra* note 165, at 150.

<sup>175</sup> *Id.*, at 151.

<sup>176</sup> Other legal frameworks require some degree of explainability. See Bibal Adrien et al., *Legal Requirements on Explainability in Machine Learning*, 29 ARTIFICIAL INTELLIGENCE & LAW 149 (2021).

challenges.<sup>177</sup> In a recent literature review, Giulia Vilone and Luca Longo identify four major areas of research in the field, thus indicating that there are several issues that must be addressed before explainability can be helpful in governance.<sup>178</sup> One is the need to develop “universal, objective criteria on how to build and validate explanations.”<sup>179</sup> The authors list some 36 parameters that have been identified as possible criteria for explainability. Among these 36 parameters are causality,<sup>180</sup> efficiency,<sup>181</sup> interpretability,<sup>182</sup> and mental fit.<sup>183</sup> Other research has focused on the structure of explanations. Structure has more than one aspect. It can refer to the need for different classes of explanations because the type of explanation required will differ depending on the person to whom the explanation is given. A programmer might require a complete explanation of how a model works, but a layperson would not likely be able to understand or find value in it.<sup>184</sup> Structure also involves identifying the components of a conventional explanation and trying to design explanations in the most effective way.<sup>185</sup> Another task is devising methods to assess the quality of an explanation. Recent approaches involve using objective criteria taken from the many parameters discussed above or by assessing human reactions to various explanations.<sup>186</sup> Moreover, as is true with keeping a human in the loop, explainability could require a tradeoff between the accuracy of an AI system and its ability to explain its decisions.<sup>187</sup>

---

<sup>177</sup> Alun Preece, *Asking ‘Why’ in AI: Explainability of Intelligent Systems—Perspectives and Challenges*, 25 INTELLIGENT SYS. IN ACCT. FIN & MGMT. 63 (2018).

<sup>178</sup> Giulia Vilone and Luca Longo, *Notions of Explainability and Evaluation Approaches for Explainable Artificial Intelligence*, 76 INFO. FUSION 89 (2021). For a more technical survey of this area, see Nadia Burkart and Marco F. Huber, *A Survey on the Explainability of Supervised Machine Learning*, 70 J. ARTIFICIAL INTELLIGENCE RES. 245 (2021).

<sup>179</sup> *Id.*, at 91. See also Mladan Jovanović and Mia Schmitz, *Explainability as a User Requirement for Artificial Intelligence Systems*, COMPUTER, Feb. 22, at 90, 91 (“It is challenging to develop a clear yet shared definition of explainability because the term does not originate from computer science (namely, AI)”).

<sup>180</sup> The term is defined as “[t]he capacity of a method ... to clarify the relationship between input and output”. Vilone and Longo, *supra* note 178, at 92, Table 1.

<sup>181</sup> “The capacity of a method ... to support faster user decision-making” *Id.*

<sup>182</sup> “The capacity to provide or bring out the meaning of an abstract concept” *Id.*

<sup>183</sup> “The ability for a human to grasp and evaluate the [algorithmic] model” *Id.*

<sup>184</sup> *Id.*, at 93–94.

<sup>185</sup> *Id.*, at 94–95.

<sup>186</sup> *Id.*, at 96–100.

<sup>187</sup> Giorgios Myrianthous, *Understanding the Accuracy-Interpretability Trade-Off*, TOWARDS DATA SCIENCE, Oct. 6, 2021, <https://towardsdatascience.com/accuracy-interpretability-trade-off-8d055ed2e445>.



## Regulatory Sandboxes

Sandboxes are limited areas where AI applications can be tested so that possible ill effects of artificial intelligence can be contained. This allows developers and regulators to gain more experience with deploying AI without the concern that unexpected events might have larger systemic effects. The hope is that such experience will instill greater confidence once an application is used beyond the sandbox. The best example is the designation of London as an area where fintech can be developed and deployed.<sup>188</sup> When a sector is highly regulated, Hilary Allen points out that regulatory sandboxes can also lower regulatory barriers to entry for startups.<sup>189</sup> Sandboxes also tend to be highly relational in nature, so that the regulators and the regulated often share knowledge and their respective social networks.<sup>190</sup> This leads to benefits that are thought to accrue from a relational approach to regulation discussed earlier in this chapter.

Some cautions, however, are in order. Jacob Sherkow argues that sandboxes pose several risks. Consumers can be harmed even in a regulatory sandbox. Because these spaces often involve relaxing existing regulations, the public might have less confidence in products or services that are ‘approved’ via a sandbox process. The approach might also erode trust in the agency itself. The collaborative approach used in sandboxes is also resource-intensive for regulators. For Sherkow, finally, there is a concern that allowing sandboxes will strengthen a tendency to move away from more restrictive regulatory approaches.<sup>191</sup> Allen adds that the ease of creating a sandbox depends on the general regulatory stance of the jurisdiction in question. She argues that implementing sandboxes in countries whose regulatory approach is more principle-based, such as the UK’s stance towards finance, is easier because those principles need not necessarily be suspended in the sandbox. However,

---

<sup>188</sup> Hilary J. Allen, *Regulatory Sandboxes*, 87 GEO. WASH. L. REV. 579, 596–600 (2019) (discussing regulatory sandboxes in the United Kingdom and Australia). According to Allen, Bahrain, Brunei, Canada, Hong Kong, Malaysia, Mauritius, the Netherlands, Singapore, Switzerland, Thailand, and the United Arab Emirates have also created sandboxes for financial services. *Id.*, at 592. Jacob Sherkow argues that the U.S. Food and Drug Administration’s Emergency Use Authorization Program, which was used for the development and approval of COVID-19 treatments and vaccines is a type of regulatory sandbox. Jacob S. Sherkow, *Regulatory Sandboxes and the Public Health*, 2022 U. ILL. L. REV. 357.

<sup>189</sup> Allen, *supra* note 188, at 587–91.

<sup>190</sup> Ahmad Alaasar, Anne-Laure Mention, and Tor Helge Aas, *Exploring how Social Interactions Influence Regulators and Innovators: The Case of Regulatory Sandboxes*, 160 TECH. FORECASTING & SOC. CHANGE 12257 (2020), at 6–11.

<sup>191</sup> Sherkow, *supra* note 188, at 369–71.

in countries such as the United States, where in her view the approach is more rules-based, legislation is required to suspend otherwise applicable law.<sup>192</sup> As Allen suggests, the question becomes whether the gains to be had by promoting a particular sector in a sandbox outweighs the administrative costs of creating one,<sup>193</sup> as well as the additional risks Sherkow identifies.

## ENFORCEMENT MECHANISMS

Enforcement mechanisms are of course an important part of any system of governance and pose a challenge for any group that promulgates norms. This is true even for states. Governments have the power to impose legal sanctions, sometimes by force. But as discussed, governments often lack financial resources, personnel, and expertise to monitor compliance with laws and regulations and to enforce them. Recall the long list of types of soft programs for AI identified by Gutierrez and Marchant. As mentioned, all such programs are criticized as being ineffective precisely because they are voluntary.<sup>194</sup> Compliance with soft norms thus depends on structures for enforcement short of the coercion available to states. The authors observe that companies and other organizations can adopt ‘organizational mechanisms’ that can encourage compliance with these programs.

Gutierrez and Marchant divide enforcement mechanisms into four quadrants. These mechanisms can be internal or external to the organization that has adopted a set of norms.<sup>195</sup> They can take the form of ‘levers’ or ‘roles.’ Levers are the “toolkit of actions or mechanisms that an organization can employ to implement or enforce a program.”<sup>196</sup> Roles describe the arrangement of persons or persons who use these levers.<sup>197</sup>

There are several internal levers and roles for enforcement of AI norms. Some levers are tied to the development lifecycle. These include educating the workforce about an organization’s approach to AI or training them on how to consider the broader implications of the technology they are developing or marketing. Some organizations also implement procedures for allowing developers to assess possible impacts and to report them, as well as procedures for

---

<sup>192</sup> Allen, *supra* note 188, at 592–93.

<sup>193</sup> *Id.*, at 605.

<sup>194</sup> Gutierrez and Marchant, *supra* note 127, at 16.

<sup>195</sup> *Id.*, at 16.

<sup>196</sup> *Id.*, at 17.

<sup>197</sup> *Id.* Peter Cihon, Jonas Schuett, and Seth Baum identify ways in which internal and external actors participate in the governance of artificial intelligence within corporations. Peter Cihon, Jonas Schuett, and Seth D. Baum, *Corporate Governance of Artificial Intelligence in the Public Interest*, 12 INFORMATION art. no. 275 (2021).

reviews throughout the development process. Allocating budgetary resources towards soft law implementation is another lever. Organizations often make public statements about their stance towards artificial intelligence that also have a leveraging effect. Finally, organizations such as governments that have decided to adopt AI applications into their work have set out indicators or benchmarks to assess progress in integrating AI applications.<sup>198</sup> Concerning internal roles, some organizations authorize specific individuals “to promote, educate, or assess issues related to an AI program.”<sup>199</sup> Others establish units within the organization to perform the same functions. Internal committees or taskforces represented by all parts of an organization is another way to do so.<sup>200</sup>

There are several external levers and roles. External levers include third-party verification mechanisms, standards, labels or certifications, professional associations who require their members to comply with standards concerning artificial intelligence, and the ability of some AI companies to require customers and suppliers to comply with norms set by those companies.<sup>201</sup> External roles are closely related to these levers. Outside members of internal committees are common in corporate boards or board committees; advisory boards with external members often are used by governments and even research organizations, nonprofits, and professional associations (often members of the associations but not employed by them).<sup>202</sup>

## CONCLUSIONS: COMPLEX INTERACTIONS BETWEEN ACTORS AND NORMS

This chapter has argued that an emerging international ‘law’ of artificial intelligence results from the interactions between actors: private firms, nation states, and international organizations identified by Boule’s work, and other actors that bear study, such as individuals, industry associations, universities, professional associations and non-governmental organizations, all of which have interests in artificial intelligence applications and their impacts, and all of which press for norms to govern them. These actors, perhaps informed by inner logics, advocate a variety of tools along a range of soft and hard norms and laws to govern artificial intelligence: general principles, statement of ethics, private ordering through contract, national strategies, and formal regulations, laws, and treaties. The participants are engaged in a dynamic relationship, “co-producing and co-enforcing norms of governance,” so that

---

<sup>198</sup> Gutierrez and Marchant, *supra* note 127, at 17–19.

<sup>199</sup> *Id.*, at 19.

<sup>200</sup> *Id.*, at 19–20.

<sup>201</sup> *Id.*, at 20–21.

<sup>202</sup> *Id.*, at 22–23.

the various forms of regulation discussed above are no longer exogenous, but emerge from the participants themselves, each with their own interests and concerns, some overlapping and others in potential conflict.

Other scholars have tried to anticipate how actors might operate within complex systems and how that environment might change as a result. Karen Alter and Sophie Meunier argue that the increasing density of interlocking international legal regimes in areas such as trade has led to a complex system of rules.<sup>203</sup> Such complexity also exists around the emerging sets of rules that apply to artificial intelligence, which run through several international regimes and between domestic and international layers of norms. According to Alter and Meunier, one can expect several consequences from this kind of complexity. For example, one can expect international actors to engage in “chessboard politics,” where international actors pursue various goals in several institutions; to forum-shop within a particular legal region; shift of regimes themselves; and to engage in strategic inconsistency: creating contradictory rules in a parallel regime to undermine rules in another.<sup>204</sup>

As another consequence, actors are likely to use the decision-making tools of bounded rationality, among them a reliance on experts and the use of heuristics, as such experts make the first rough cuts at assessing and responding to various policy issues. In this regard, as discussed earlier this approaches a kind of fractal geometry, the book will show how the different actors use the same regulatory tools and deliberative approaches towards AI governance at organizational, state, and international levels. Alter and Meunier find that international regime complexity and its reliance on experts results in small group environments, thus bringing into play the positive and negative aspects of small group dynamics.<sup>205</sup> Here, I would add that because many of the actors in AI governance are large organizations, one also sees tensions between small and large group dynamics. Finally, for Alter and Meunier, the feedback effects inherent in complex systems can lead to competition among various institutions; unintentional reverberations in other regimes; a lack of accountability because feedback and secondary effects make it hard to identify who might be responsible for a particular issue; a greater emphasis on trustworthiness

---

<sup>203</sup> Karen J. Alter and Sophie Meunier, *The Politics of International Regime Complexity*, 7 PERS. POL. 13 (2009). The article is an introduction to a symposium on regime complexity. Some of the symposium contributions include Alexander Betts, *Institutional Proliferation and the Global Refugee Regime*, 7 PERS. POL. 53 (2009); Emilie M. Hafner-Burton, *The Power Politics of Regime Complexity: Human Rights Trade Conditionality in Europe*, 7 PERS. POL. 33 (2009); Laurence L. Helfer, *Regime Shifting in the International Intellectual Property System*, 7 PERS. POL. 39 (2009).

<sup>204</sup> Alter and Meunier, *supra* note 203, at 16–17.

<sup>205</sup> *Id.*

because of the actions of an international actor in one area will affect how it is perceived in others; and lastly, greater ease of exit from international regimes through non-compliance, regime-shifting, or as a last resort, withdrawal from international organizations. As the book progresses, we will see many of these consequences at play in the governance of AI applications.