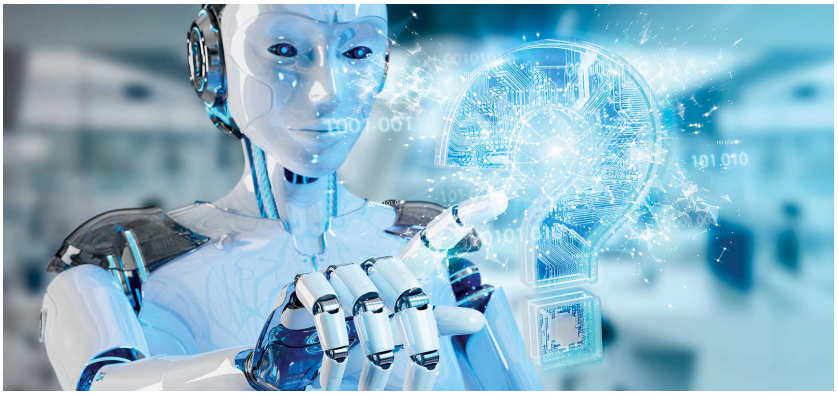
Toward the Agile and Comprehensive International Governance of AI and Robotics

By WENDELL WALLACH
Yale Interdisciplinary Center for Bioethics, Yale University, New Haven, CT 06520 USA

GARY MARCHANT
Sandra Day O'Connor College of Law, Arizona State University, Phoenix, AZ 85004 USA



I. NEED FOR AGILE GOVERNANCE

Rapidly emerging technologies, such as AI and robotics, present a serious challenge to traditional models of government regulation. These technologies are advancing so quickly that in many sectors, traditional regulation cannot keep up, given the cumbersome procedural and bureaucratic procedures and safeguards that modern legislative and rulemaking processes require. Consequently, regulatory systems will predictively fail to put in place appropriately tailored regulatory measures by the time new applications of fast-moving technologies begin to affect society. Perhaps even worse, if a regulatory system does somehow manage to rush into place new regulations for an emerging technology, they will likely be obsolete by the time the ink dries on the enactment. Given this so-called "pacing problem," traditional regulatory approaches will either produce no regulation or bad regulation [1].

Emerging technologies, such as AI and robotics, present additional regulatory challenges beyond the pacing problem [2]. While scientific uncertainty affects all types of product and process regulations, the novelty and rapid pace of emerging technologies, such as AI and robotics, present exceptionally broad and intractable uncertainties about their benefits, risks, and future trajectories. Emerging technologies often span many industry sectors and cross the jurisdictions of multiple regulatory agencies, creating large and diverse sets of stakeholders in government, industry, and civil society. Emerging technologies also generally present a broad range of concerns that go beyond the safety risks and efficacy issues that government agencies are often tasked with addressing. For example, AI presents not only the safety risks in a variety of contexts ranging from autonomous vehicles to financial algorithms but also presents the

Digital Object Identifier 10.1109/JPROC.2019.2899422

concerns relating to privacy, autonomy, enhancement, bias, fairness, justice, relationships to others, unemployment, national security, and existential risk. Finally, AI, robotics, and other emerging technologies develop in an international context, often making national regulation disadvantageous, inept, or incomplete.

For all these reasons, there is a growing consensus that traditional government regulation is not sufficient for the oversight of emerging technologies, such as AI and robotics. While government regulators and policymakers still play a critical role, oversight must be expanded to also include new institutions and methods that are more agile, holistic, reflexive, and inclusive.

II. SOFT LAW GOVERNANCE

"Soft law" has been advanced as a strategy to try to overcome limitations and challenges of traditional government regulation for emerging technologies, such as AI and robotics. Whereas "hard law" consists of legally enforceable requirements imposed by the governments, "soft law" consists of substantive expectations that are not directly enforceable. Soft law measures can be promulgated by a variety of stakeholders, including governments, industry actors, nongovernmental organizations, professional societies, standard-setting organizations, think tanks, public–private partnerships, or any combination of the above. Examples of soft law include voluntary programs, standards, codes of conduct, best practices, certification programs, guidelines, and statements of principles.

Soft law can address many of the limitations of traditional regulation for emerging technologies [3]. They can usually be adopted and revised relatively quickly. Not limited by an agency's jurisdiction and delegation to certain concerns or applications, they can address a technology holistically. They can involve a broad range of stakeholders and create a cooperative approach, rather than an adversarial approach. In addition, soft law can be inherently international in scope, such as with standards set by the IEEE and the ISO.

Notwithstanding these advantages, soft law approaches have their own limitations, perhaps most significantly the lack of enforceability and the absence of any coordination that the top–down government regulation provides. While the soft law requirements are not directly enforceable, there does exist various mechanisms for indirect enforcement [7]. One such mechanism is for the governments to eventually adopt soft law requirements into traditional regulatory enactments after they have been first field-tested as soft law. For example, the Future of Life Institute promulgated its Asilomar principles as a soft law tool for AI governance, but now the State of California has adopted those principles into its statutory law. In other words, soft law governance can be a first stage toward enacting the hard law where necessary, with the proviso that once laws and regulatory oversight have been codified, agility is sacrificed as often inflexible regulatory requirements and intransigent bureaucracy sets in. Another indirect enforcement tool is for the Federal Trade Commission in the U.S. (or its equivalent in other countries) to take enforcement against companies that break their promises to comply with soft law programs as "unfair or deceptive" business practices. Other indirect enforcement approaches include insurers requiring compliance with soft law risk management programs as a prerequisite for liability coverage, journals requiring compliance with selected soft law provisions as a condition of publication, and grant funding agencies requiring soft law compliance of its grantees.

The second major problem with soft law programs is that because any entity can develop or propose soft law, there tends to be a proliferation of such programs. Companies and other entities may have a difficult time in navigating the tangle of soft law programs and traditional regulatory programs, struggling to make sense of how all the different guidelines fit together, or whether they even do. This lack of coordination problem is what led us to propose the governance coordinating committees (GCCs) in articles and books [4], [5].

III. GCC MODEL

Our 2015 GCC proposal has recognized that for emerging technologies, such as AI, there would likely be a multitude of governance actors, issues, and programs, both in hard law and soft law. The GCC would be situated outside government but would include participation by government representatives, industry, nongovernmental organizations, think tanks, and other stakeholders. It would not itself have any role in promulgating new governance instruments but rather would act as an "orchestra conductor" for the instruments that have already been promulgated or proposed, analyzing how they fit together (or not), where they agree (or not), and where gaps were left which perhaps needed to be addressed. The GCC would also provide a forum for dialog and debate among stakeholders and would be a contact point for the public and media seeking information about the relevant technology and its governance. In addition, the GCC can help to coordinate some of the indirect enforcement options for the above-mentioned soft law. Presumably, many of the governance tasks the GCC will loosely coordinate will have been taken on by nongovernmental institutions, industry, and even by governments, and therefore, the GCC itself will be quite lean and agile.

When we first proposed the GCC in 2015, we suggested that AI and synthetic biology would be the good first candidates for a GCC because they were the relatively new technologies in terms of governance activities and polarized opinions had not yet taken hold. In the approximately four years since we first proposed the GCC, the concept has received much attention and some traction, and we extend and elaborate on our proposal and its application to AI here.

IV. PROCESS-BASED SOFT LAW GOVERNANCE

Within AI and robotics, soft law governance can extend to methods by which the engineering teams integrate ethical considerations into the creation of computational systems and the processes through which a corporation or research laboratory oversees the development of AI. We refer to engineering strategies and the oversight of development as process-based soft law.

A. Ethics and Engineering

The commitment to engineering ethics, best practices, and compliance standards needs to be renewed constantly by industry and research teams. In addition, value-added design and machine ethics have been proposed as methods that the engineering teams can adopt to ensure that the systems they design will reflect the prevailing norms and have a positive societal impact. Value-added design offers an approach to treat values, in addition to safety, as design specifications. For example, systems can be designed to maximize users privacy. Even a determination of who will be held responsible for a system failure can be treated as a system specification by the team engineering the system. In effect, this often already happens, as companies turn away from the features or platforms that increase their responsibility for failures to the one's that hold users liable. Nevertheless, determining the liability and responsibility as a design specification can sensitize the engineers to the risks and societal impacts of the technologies they develop. This value-added design process can be facilitated by integrating ethicists and social theorists into design teams, not as naysayers, but as fellow designers sensitive to ethical and societal concerns.

The prospect of developing AI systems and robots sensitive to values and norms, capable of reognizing the morally significant situations, and factoring these into their choices and actions, offers a particularly intriguing means for designers and engineers to ensure the safety and beneficence of the systems they deploy. Progress in machine ethics or what is referred to by the AI researchers as solving the value alignment problem will, however, be relatively slow. The overall project of aligning the behavior of AI systems with that of the acceptable human behavior is a daunting challenge. Moral decision-making machines will initially be designed for bounded applications, in which the options they confront are limited.

Progress in machine ethics will become a key factor for expanding the environments in which cognitive systems can be safely deployed. Success in designing computational systems that factor ethical considerations into their choices and actions will be achieved when the machines govern their own behavior. Thus, progress in building moral machines eases demands for governmental regulations and for the other forms of soft law governance while expanding the markets in which the intelligent systems can be deployed. In other words, the explicit ethical decision making by intelligent machines can become a core component in the comprehensive and agile governance of AI and robotics.

Nevertheless, moral machines also pose their own governance concerns. Similar to any complex adaptive or learning systems, it will be difficult to test and empirically demonstrate that the moral machines will reliably act in an acceptable manner, for example, in the field where the system might encounter a situation that it had not been trained upon and which had features it was not designed to recognize. Furthermore, if the moral machine is a learning system, it might well alter its responses to challenges in new ways that it had never been tested on. Paradoxically, machine ethics will increase the likelihood that a system will act ethically and in a predictably appropriate manner while, at times, displaying unpredictable and potentially risky behavior. Occasionally, unpredictability will occur because at this stage of development, we cannot presume that moral machines will fully understand all the ethical considerations arising in a complex situation.

B. Oversight of Research

If industry leaders and the heads of other research facilities are serious about the responsible development of AI and robotics, we recommend that they establish technology review boards (TRBs). A TRB would perform an ethical risk assessment (BS 8611:2016) to evaluate the impact of the tools and techniques that the institution's engineers are developing and share its findings with both the design team and with management. Among the activities a TRB would engage in is consideration of worse case scenarios, disaster planning, determination of who might be held responsible when a system fails, the fairness and privacy implications of the data that the system will use, and analysis of societal impacts should the system be widely deployed. From the perspective of a management or a corporate board, the TRB will help assess the liability of a company for a system it markets and protect the corporation from class-action lawsuits. From the perspective of the design team, the TRB will ensure that the technology is developed responsibly and will make suggestions for improving the product's success. Reports from review boards often get lost or ignored. An effective TRB should include a corporate AI Ethics Officer with the power to bring concerns to the attention of management.

V. CALL FOR AN INTERNATIONAL CONGRESS FOR GOVERNANCE OF AI AND ROBOTICS

Many of the concerns that AI and robotics pose can and will be addressed by regional (e.g., the EU),

national, or local ethical/legal oversight, while others will require compliance with the standards set by international bodies. In addition to the IEEE, many other international institutions and partnerships have begun projects directed at the oversight of AI and robotics. Nevertheless, for coordination purposes and for comprehensive monitoring of the field and its impacts, we propose the establishment of an international GCC (IGCC). An IGCC could function as a good-faith broker mediating among the various stakeholder groups. One role for an IGCC might be to underscore "best practices" and outline considerations for various national and regional bodies, as they consider the most appropriate soft and hard laws for their culture [8]. This would be particularly helpful for smaller countries that lack the resources to develop their own policies. Indeed, these "best practices" might even be considered *de facto* international standards, subject to variations introduced by the national and regional bodies.

Establishing a new national or international governance mechanism is certainly not easy and entails overcoming an array of implementation challenges from establishing effectiveness, trust, and credibility to funding and adequate insulation from political or economic influence. An IGCC, working in cooperation with all the other international and national bodies in the AI space, is unlikely to initially have any capacity to enforce standards or best practices, but it can have a great deal of influence if perceived as a good-faith broker. More importantly, its influence and authority will grow over time.

As a forerunner to the establishment of a global mechanism for the governance of AI, we further propose the convening of the International Congress for the Governance of AI (ICGAI) in the fall of 2019 or early 2020. An ICGAI provides a first step in multistakeholder engagement over the challenges arising from these new technological fields. The ICGAI will need representation from not only leading states but also from major AI industry leaders, research laboratories, and nongovernmental organizations, which may also represent the broader publics' concerns. Furthermore, it would be helpful if the ICGAI is not perceived as dominated by those in North America and Europe. Therefore, we propose it be convened in Asia (e.g., Singapore, Hong Kong, or Tokyo) or the Islamic World (e.g., Dubai or Bahrain). A venue in Europe which is perceived as a neutral site might also be acceptable (e.g., Geneva, Prague, Oslo, or Venice). We are hopeful that an ICGAI will endorse steps toward building agile and responsible institutions for the continuing oversight of AI and robotics. At an event in NYC on September 26, 2018, during the UN General Assembly yearly meeting, 70 representatives from leading bodies in the AI space and in the international governance space endorsed the convening of ICGAI in November 2019.

The ICGAI may elect to govern AI and robotics differently than the mechanism that we have proposed. Whether something like an IGCC can be established is less important than the fact that this general model contains suggestions that will be helpful for forging agile and comprehensive governance of AI and robotics.

AI is a technology which will help shape so many aspects of life over the upcoming century. It is closer to being like electricity than a sector-specific technology, such as the automobile. Governance mechanisms for AI are rapidly being formulated and proposed, such as those being considered by the EU High-Level Expert Group on Artificial Intelligence. The choice is whether this technology is governed by a kludge of overlapping and sometimes conflicting laws, regulations, standards, and guidelines or whether we put in place a governance mechanism that embraces the comprehensive monitoring and the gentle coordination of the stakeholders and policies. ■

REFERENCES

[1] G. E. Marchant, "The growing gap between emerging technologies and the law," in *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem*. Dordrecht, The Netherlands: Springer, 2011, pp. 19–33.

[2] G. E. Marchant and W. Wallach, "Introduction," in *Emerging Technologies: Ethics, Law and Governance*. London, U.K.: Routledge, Nov. 2016, pp. 1–12.

[3] G. E. Marchant and B. Allenby, "Soft law: New tools for governing emerging technologies," *Bull. At. Scientists*, vol. 73, no. 2, pp. 108–114, 2017.

[4] G. E. Marchant and W. Wallach, "Coordinating technology governance," *Issues Sci. Technol.*, vol. 31, no. 4, pp. 43–50, 2015.

[5] W. Wallach, *A Dangerous Master: How to Keep Technology From Slipping Beyond Our Control*. New York, NY, USA: Basic Books, 2015.

[6] BSI Group. (2016). *BSI-BS 8611 Guide to Ethical Design Application Robots Robotic Devices*. Accessed: Dec. 20, 2018. [Online]. Available: https://standards.globalspec.com/std/10005027/bsi-bs-8611

[7] G. E. Marchant, "'Soft law' mechanisms for nanotechnology: liability and insurance drivers," *J. Risk Res.*, vol. 17, pp. 709–719, Feb. 2014.

[8] W. Wallach and G. E. Marchant, "An agile ethical/legal model for the international and national governance of AI and robotics," in *Control and Responsible Innovation in the Development of AI and Robot*, W. Wallach, Ed. The Hastings Center, 2018.