An Agile Ethical/Legal Model for the International and National Governance of AI and Robotics

Wendell Wallach[1] and Gary E. Marchant[2]

1. Yale Interdisciplinary Center For Bioethics; email: wendell.wallach@yale.edu
2. Center for Law, Science & Innovation, Arizona State University; email: gary.marchant@asu.edu

Abstract

The accelerating pace of emerging technologies such as AI has revealed a total mismatch between existing governmental approaches and what is needed for effective ethical/legal oversight. To address this "pacing gap" the authors proposed governance coordinating committees (GCCs) in 2015 as a new more agile approach for the coordinated oversight of emerging technologies. In this paper, we quickly reintroduce the reasons why AI and robotics require more agile governance, and the potential role of the GCC model for meeting that need. Secondly, we flesh out the roles for government, engineering, and ethics in forcing a comprehensive approach to the oversight of AI/robotics mediated by a GCC. We argue for an international GCC with complementary regional bodies in light of the transnational nature of AI concerns and risks. We also propose a series of new mechanisms for enforcing (directly or indirectly) "soft law" approaches for AI through coordinated institutional controls by insurers, journal publishers, grant funding agencies, courts and governments. The GCC is particularly well-adapted and situated for coordinating this type of enforcement of soft law requirements. Finally, we show how a GCC can support and reinforce the governance initiatives of organizations such as the IEEE, WEF, the Partnership in AI, and various AI research centers.

Introduction: The Need for Agile Governance

The accelerating pace of emerging technologies such as AI, and the onset of a Fourth Industrial Revolution, has revealed a total mismatch with existing governmental approaches and what is needed for effective ethical/legal oversight (Marchant and Wallach 2016). These emerging technologies exceed the regulatory scope, capabilities and jurisdiction of any one agency or nation. For example, AI raises ethical, legal and social concerns in need of governance relating to military use, safety, privacy, transparency, bias, unfair business practices, antitrust, human enhancement, criminal justice, impacts on personal, family and societal relationships, economic equality, technological unemployment, existential risk, and no doubt many others. These diverse issues span many different industries, regulatory authorities, non-governmental organization, experts and other stakeholders. While these concerns raise distinct issues that often must be addressed in their own way, they are also connected in that they relate to the same underlying technologies and therefore necessitate a more holistic approach.

In addition to the complexity of emerging technologies such as AI, the pace at which they are being developed also presents a major obstacle to traditional government regulation. AI is developing at an accelerating trajectory, surprising even many AI experts about its recent speed and impact (Executive Office of the President, 2016). At the same time, our traditional governmental institutions of legislation, regulation and judicial review are slowing down rather than speeding up, creating the "pacing problem" (Marchant 2011).

To address these governance challenges, the authors proposed governance coordinating committees (GCCs) in 2015 as a new more agile approach for the coordinated oversight of emerging technologies such as AI (Marchant and Wallach 2015). In addition, we further proposed that pilot projects be started for AI/robotics and also for synthetic biology. The selection of these fields for pilot projects was occasioned by the fact that AI/robotics and synthetic biology are relatively new fields of research, largely unencumbered by rules and regulations. Much has happened since our initial proposal. Machine learning approaches have led to breakthroughs in AI, and CRISPR/Cas9 has speeded up gene editing and in turn the development of genomic products and synthetic organisms. Governments have taken notice and are studying ways to regulate AI and genomics. In addition, many existing and new non-governmental organizations (NGOs or Civil Society), alliances, and research centers have sprung up to address potential benefits, societal impacts, risks and dangers posed by the deployment of AI and gene editing. In this paper, we expand on our initial proposal to describe how a GCC, perhaps at an international level, could help to break the current logjam with respect to agile and effective governance of AI.

The GCC Approach

The basic idea behind the GCC is that an orchestra needs an orchestra conductor – not to play the instruments for the various players, but to coordinate all these important parts of the performance. For emerging technologies like AI, there is an explosion of governance strategies, actions, proposals and institutions. All have an important role to play – whether they are from government, industry, NGOs, academia or some combination of the above, but no one entity or program can hope to govern the fields of AI and robotics *in toto*. What is missing is some mechanism for communication, coordination, synchronization and synergy.

It is for this reason we proposed the idea of an "issue manager" for the governance of individual emerging technologies like AI, which we named a Governance Coordinating Committee or GCC. The GCC would serve several coordinating functions. One function would be as an information clearinghouse, by collecting and reporting in one place all significant programs, proposals, ideas or initiatives for governing AI. The GCC could also perform a monitoring and analysis function, such as identifying gaps, overlaps, and inconsistencies with respect to existing and proposed governance programs. It could serve as an early warning system, by noting emerging issues or problems that are not addressed or covered by existing governance programs. It could provide an evaluation program that scores various governance programs and efforts for their metrics and compliance with stated goals. The GCC could provide a forum for stakeholders to meet and discuss governance ideas and issues and to produce recommendations, reports, and roadmaps. It could serve as a trusted "go-to" source for the media, the public, scholars and stakeholders to obtain information about AI and its governance. Finally, the GCC could serve as a convener for interested stakeholders on specific issues to meet and try to forge a negotiated partnership program for addressing unaddressed problems or governance needs.

There are many practical implementation issues that need to be addressed for the creation of a GCC. Who would fund it? What type of governance system would be needed to operate the GCC? How would the GCC be evaluated, and by whom? How much staff would the GCC have, and who would hire them? Would the GCC have a direct or indirect government role, or would it be solely outside of government? How would stakeholders have a say and role in the operation of the GCC? What would be the specific goals and functions of the GCC? These are critical questions, but they do not lend themselves to one obvious set of answers. Rather, they are issues that need to be negotiated and discussed in the context of a specific proposal and effort to create a GCC by as broad of range of stakeholders as possible.

To facilitate this creative process, we extend our GCC proposal below to provide some additional insights, timely possibilities and benefits that a GCC could play in the governance of AI.

International and National Oversight of AI and Robotics

In the spring of 2016, a project to Build Global Infrastructure to Ensure that AI and Robotics are Beneficial was initiated (Wallach 2016). This project is referred to as the BGI project (https://bgi4ai.org/). The BGI project is a pilot for applying the GCC model to the ethical/legal oversight of those two fields of research (AI and robotics), but while a GCC was initially proposed for the U.S., this new project begins as an international program with complementary national or regional bodies. A complementary GCC could cover the needs of one country or a region, such as a Pan-Arab GCC. For theoretical purposes the international body might be referred to as a global governance coordinating committee (GGCC). But once initiated, within the United Nations or as a new NGO, we expect the monitoring, multi-stakeholder engagement, coordinating, and other functions to be established under a new name. A central role of a GGCC and its complementary national or regional GCCs will be to underscore gaps in existing mechanisms for the oversight of AI/robotics, to propose new mechanisms to address those gaps after considering an array of available tools, and to build the governance infrastructure necessary to sustain those proposals.

Some of the concerns AI and robotics pose must be addressed globally while others are better left to regional, national or local ethical/legal oversight. Lethal autonomous weapons, and whether their deployment should be restricted by an arms control treaty, can only be resolved internationally. The Conventional on Certain Conventional Weapons at the UN in Geneva has already taken up this question. Regulatory policies for the deployment of fully autonomous vehicles are being developed by many countries independently, and even states or regions within those countries.

While each country could in theory establish its own technical standards, testing procedures, compliance requirements, and quality management standards that must be met before products can be marketed, commonly they adopt those developed by international bodies such as the IEEE and ISO. For other regulatory and soft law concerns, many countries are unable to establish their own requirements, or adopt those set by other countries. One role for a GGCC might be to underscore "best practices" and outline considerations for various national and regional bodies as they consider the most appropriate soft and hard law for their culture. Indeed these "best practices" might even be considered de facto international standards, subject to variations introduced by national and regional GCCs. This would be particularly helpful for poorer regions.

The BGI project will begin with the establishment of a GGCC and several independent yet complementary GCCs.

Enforcement

In our initial GCC proposal we emphasized the importance of soft governance mechanisms, which include industry standards, professional society codes of conduct, laboratory practices and procedures, insurance policies, statements of principles, voluntary government programs, certification programs and similar measures. Soft law measures impose substantive expectations or obligations that are not directly enforceable by government. We suggested that soft governance mechanisms should be favored in the GCC over hard governance (laws, regulations, and regulatory bodies) because they often involve multi-stakeholder participation, and can be adopted and modified more quickly and nimbly than traditional legal instruments. Another benefit of soft law mechanisms is that because they are usually not associated with a specific regulatory agency or jurisdiction, they can be applied at the international level (Marchant and Abbott 2013).

However, the obvious weakness of soft governance mechanism lies in the difficulty, if not inability, to enforce them directly. There are nonetheless a number of indirect ways to enforce soft law measures, and the GCC can provide an appropriate forum for bringing the relevant players together to implement such soft law enforcement mechanisms. For example, we propose an additional and new role for governments, which is to create means to punish those who violate soft governance standards in a manner that leads to harm to people, non-human animals, the environment, institutions, or establishes practices that have undesirable societal impacts. Such an indirect government enforcement opportunity can be created using soft law instruments, perhaps negotiated or ratified through the processes of the GCC.

While the specific legal authority for such a government role will vary country by country, an example is provided by the Federal Trade Commission (FTC) in the United States. The FTC has a long-standing statutory authority to take enforcement action against "deceptive and unfair" business practices. The FTC has in recent decades re-interpreted this authority to apply it to companies that fail to comply with their publicly stated commitments, including adherence to soft law instruments such as private standards or codes of conduct (Hetcher 2000). The FTC's legal position is that a company's failure to live up to its public commitment misleads and deceives consumers, in violation of the statutory prohibition of deceptive and unfair practices. A GCC could help create or promote a private code of best practices for AI and robotics that participating companies would agree to, with the understanding that the FTC is empowered to take enforcement action against companies that fail to comply with their commitments. Similar indirect governmental enforcement mechanisms may be possible in other jurisdictions.

Courts may also have some enforcement role for soft law instruments created or publicized by GCCs. Private standards can set the standard of care for industry actors, particularly in the absence of any regulatory standards. Thus, private standards can provide a partial liability shield for those entities that comply with the standards, and can be used as a sword to establish the lack of due care by those entities that fail to comply with the private standards (Marchant 2014). The more recognized and accepted the private standard, the more force it has as a shield or sword for liability in personal injury or other tort lawsuits. GCC endorsement of a soft law instrument could therefore give it more salience in private lawsuits, and could provide another indirect enforcement mechanism.

In addition, many governments and major corporations strengthen standards by requiring that they be met for products and services purchased by the government and industry. ISO 9000, for example, is an international quality management and quality assurance standard designed to increase business efficiency and the quality of products. Organizations that demonstrate the ability to provide products and services consistently can apply for ISO 9001 certification (a subset of ISO 9000). While organizations that fail to meet ISO 9000 standards are not directly punished, they are indirectly punished in their inability to sell their products and services into large markets.

Insurance companies also provide an indirect enforcement mechanism. After the asbestos debacle, liability insurers realized they cannot afford to insure companies or products that present unknown and potentially unlimited liability if harms occur. As a result, liability insurers are increasingly taking a more active risk management role for emerging technologies that present highly unknown but potentially widespread risks. For example, liability insurers for companies that manufacture or handle nanotechnology materials are increasingly requiring their clients to adopt an active risk management program as a condition for coverage (Marchant 2014). These risk management programs often involve a commitment to comply with a voluntary standard or code of conduct. A GCC could work with companies and insurers to identify an appropriate set of risk management standards for companies working with AI applications that present significant risks.

There are other mechanisms for indirectly enforcing soft law instruments that a GCC could help facilitate. Journal publishers could agree to only publish articles that comply with applicable codes of conduct or professional standards. Funding agencies could condition funding on compliance with appropriate soft law standards. Research institutions could mandate compliance with soft law standards by their employees, perhaps enforced by an institutional review committee based on Institutional Biosafety Committees (Fatehi et al. 2012). All of these mechanisms hold significant potential for indirectly enforcing soft law norms, and a GCC could provide the impetus and focus to enable such efforts.

Ethics and Engineering

Societal and ethical considerations can indicate the need for new soft law approaches to responsible innovation. Testing and compliance standards, inspection and certification of AI systems by a third party, corporate AI ethics officers, data ethics committee, and AI and robotic ethics review boards are among the soft governance mechanisms that would be helpful.

The feasibility of addressing gaps through ethics and social engineering or as problems that can be tackled through engineering solutions should be explored before turning to either soft law or hard regulations. An array of challenges can be met by the reinforcement of existing societal values, establishment of new norms, or by feasible technological solutions.

For example:

a.) There is no “right answer” to the popular trolley car dilemmas as to whether or how a self-driving vehicle should be designed and programmed to act in a situation where all options will lead to the death of humans. Nevertheless, societies can, after reflecting on the challenges, elect to establish new norms that provide manufacturers with guidelines as to what is acceptable.

b.) In the imagination of some engineers, all problems can be solved technologically. But often their proposals are based upon fanciful gadgetry that is not feasible with the available tools and techniques. Determination of the feasibility, timeliness, and cost of developing technological solutions can be weighed against other options such as social engineering or government regulation.

c.) Information systems have already undermined some country’s established standards for privacy. Autonomous systems threaten to undermine the foundational principle that an agent, either human or corporate is responsible, culpable and liable for any harm caused by a device they deploy. If societies demand that the principle of a responsible agent be maintained and enforced by the legal system, manufacturers will turn away from marketing AI systems and robots whose safety they cannot guarantee.

Furthermore, technology and ethics can be combined together in forms such as imbuing products with values and even means to integrate values and ethical considerations into their choices and actions. Ethicists and social theorists can be introduced as members of design teams to help engineers become more fully aware of the potential societal impact of the systems they are developing, and this in turn may suggest the use of different platforms or design components.

The prospect for developing AI systems and robots capable of factoring norms, ethical concerns, and laws into their choices and actions has received significant attention by science fiction writers, then philosophers and computer scientists (Wallach and Allen 2008), and has now become an engineering challenge referred to as machine morality, machine ethics (Anderson & Anderson eds. 2011), or value alignment (Russell 2015).

Support and Coordination, Not Competition

How can a GGCC or GCCs support and complement (not compete with) the many international NGOs (e.g., the IEEE), governments (e.g., the EU), industry promoted consortiums (e.g., the Partnership in AI), and research centers (e.g., AI Now) that have emerged to address challenges arising in the development and deployment of AI and robotics? AI is beginning to affect every facet of modern life, and as it does so, an array of existing institutions and a proliferation of new centers and consortium have arisen to tackle emerging challenges. Ethical guidelines, standards, protocols, policy recommendations, research findings, tools for data analytics, and technical means to ensure safety and fairness are appearing, and will continue to be developed by these various initiatives.

Most of the initiatives are siloed attempts to deal with specific concerns, or institutions, such as research centers, whose influence will be limited unless its recommendations come to the attention of policy makers or industry leaders. None of these entities or programs can hope to govern the fields of AI and robotics *in toto*.

Nevertheless, each body is sensitive to competition from similar institutions, and will be unwilling to participate in joint deliberations if it feels the deliberating body will merely usurp its ideas and authority. In order to be effective a GCC (or GGCC) must attract the involvement of these other institutions, respect and support their contributions, and provide services that they cannot provide by themselves. It should not usurp the authority of other institutions. Rather, it should support their activities and facilitate their working together to ensure that best practices come to the fore and that the resources of individual institutions are not wasted through unnecessary duplication of effort. It will be helpful if the various institutions are aware of similar projects being performed by other researchers and institutions. Furthermore, it will be helpful for those proposing new policies, standards, and guidelines to be able to bring their work to the attention of others who have a good prospect of affecting their adoption.

In addition to governments, a few of these bodies have, or are expected to have, significant worldwide impact on the development of AI and robotics. Among the most influential internationally are the IEEE and the World Economic Forum. The Partnership in AI (PAI) – formed by Amazon, Apple, Facebook, Google, IBM, and Microsoft – is young, and yet it is casting a wide net and has already embraced many NGOs and research laboratories. PAI will certainly be influential, but it remains unclear whether this initiative will eventually include representation from all regions and all industry leaders. A few programs have also begun within the UN to address concerns posed by AI. But, as of this writing, no one institution can claim to speak for or include all the key stakeholders and on the broad array of issues arising from the development of AI and robotics.

Furthermore, much of the focus on the various emerging concerns is dominated by industry leaders and researchers from Europe and North American, as well as Japan and South Korea. Meanwhile, the BRICS economies (Brazil, Russia, India, China, and South Africa) have to date been less active in international forum on AI and robotics. Furthermore, China and it primary IT corporations (Alibaba and Baidu) rival counterparts in the U.S. in the development of AI. The Arab world, Africa, South and Central America have still to make their voices fully heard. In other words, there is a need for a GGCC.

A GGCC must draw upon and serve all of the stakeholders from industry and civil society to governments and international standard setting bodies. Hopefully, it will also find means to represent the interests of under-served nations and people, even those that lack stable governments. There is no shortage of opportunities for the fruits of AI and robotics to benefit all of humanity, but this can only occur when risks, dangers, and undesirable societal impacts are also being mitigated. A comparable level of responsibility will also fall upon national and regional GCCs.

Outcomes Not Merely Process

The GCC model offers a process and a framework for responsive and agile governance. The details of putting a GCC or GGCC into place are extensive and will be complicated, given the fact that AI and robotic applications are context specific, and each context, such as healthcare, will require its own supporting mechanisms and institutions. An agile process, however, is only worth pursuing if it effectively leads to significant outcomes. The development of the supporting institutions must proceed hand-in-hand with the pursuit of specific goals. Furthermore, the goals will help dictate the structure of the institutions and mechanisms put in place. The challenge lies in forging mechanisms that will serve both immediate goals and longer-term needs for responsive and agile oversight. With this in mind we will propose a project for the creation of a GGCC and a first national and regional GCCs.

Three near-term issues have emerged regarding the fairness, transparency, and integrity of machine learning systems (Wallach 2018):

a) There is a lack of transparency as to how neural networks achieve their outputs, i.e., reach conclusions. This is particularly problematic given the explosion in use of deep-learning algorithms for a vast array of applications. Should an accident or harms occur, there may be no way to even forensically determine what went wrong.

b) The output of a deep learning algorithm will be unfair and biased if there is bias inherent in the dataset the system is trained upon. A deep learning algorithm might also yield a false or dangerous output if the data it is fed is filled with inaccuracies or simply too limited (insufficient in depth) to reach an accurate conclusion.

c) Concentration of power, claims of data ownership, and the inability of individuals to opt out of IT services whose handling of their data can jeopardize their security, privacy, and finances is receiving considerable attention in the media and by some governments. These issues are exacerbated when the use of an individual's data by social media companies makes users susceptible to political and ideological campaigns, behavior manipulation, and undesired marketing campaigns.

AI engineers, data analysts, standard-setting institutions, multi-stakeholder forums, research centers, policy planners, and some governments are working on means to address various facets of these challenges and mitigate potential harms. A General Data Protection Regulation, which comes into force in 2018, has been enacted by the EU. Among other provisions, these regulations provide for a right to obtain an explanation for any decision made by an algorithm, as well as the right to opt-out of various forms of data collection. Arguably the EU's regulations on these matters may be too broad and may even unnecessarily stultify innovation and economic progress. Nevertheless, they underscore the importance attributed to these issues.

Furthermore, the landscape is changing. Policy makers will hopefully clarify when a lack of algorithmic transparency is problematic and when it is not. Data analysts are likely to produce tools that help illuminate biases and other limitations inherent in training data, as well as biases in system outputs. The difficulty lies in the fact that while research progress and standards are being formulated, leading industry players, healthcare providers, legal decision-makers, and other parties are rapidly deploying systems and marketing products whose safety and societal impacts have not been determined.

We propose a Global Congress as a forerunner to the establishment of a GGCC. Focusing on the issue of algorithmic transparency is particularly appropriate as an agenda for such an international gathering. This Congress would establish preliminary guidelines for the deployment of algorithms that are not fully transparent. It would clarify when learning systems can be exempt from transparency requirements, what testing and compliance must be performed before potentially risky systems are deployed, and in which situations or contexts systems that lack transparency (opacity) should never be deployed. The standards, practices and procedures for setting these preliminary guidelines may have already been clarified by other bodies. However, these other players are likely to be dominated by industries and institutions concentrated in North America or Europe. It therefore becomes important for other companies, institutions, countries and regions to evaluate whether such guidelines are appropriate given their needs. In other words, the Congress provides an opportunity for stakeholders to endorse (or modify if necessary) the best practices that have emerged to date. We propose that this Congress be held in a nation and at a venue that is considered relatively neutral. Given the dynamic state of the research and lessons learned from monitoring best practices, preliminary guidelines will need to be revisited in a few years, modified and hopefully made more precise. The very act of convening a Congress also provides an opportunity to lay foundations for the ongoing multi-stakeholder responsibilities of a GGCC. The Congress itself might also endorse steps towards building agile and responsible institutions for the continuing oversight of AI and robotics.

References

Anderson, M. and Anderson, S.L. (eds.) 2011. *Machine Ethics.* Cambridge, UK: Cambridge University Press.

Executive Office of the President. 2016. Artificial Intelligence, Automation, and the Economy, December 2016. https://obamawhitehouse.archives.gov/sites/whitehouse.gov/files/documents/Artificial-Intelligence-Automation-Economy.PDF

Fatehi, L. 2012. Recommendations for Nanomedicine Human Subjects Research Oversight: An Evolutionary Approach for an Emerging Field. *Journal of Law, Medicine & Ethics* 40: 716-750.

Hetcher, S. 2000. FTC as Internet Privacy Norm Entrepreneur. *Vanderbilt Law Review* 53:2041-2062.

Marchant G.E. 2011. The Growing Gap between Emerging Technologies and the Law. In *The Growing Gap between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem* 19-33. Dordrecht: Springer.

Marchant, G.E. 2014. 'Soft Law" Mechanisms for Nanotechnology: Liability and Insurance Drivers. *Journal of Risk Research* 17: 709-719 (2014).

Marchant, G.E. and Abbott, K.W. 2013. International Harmonization of Nanotechnology Governance through "Soft Law" Approaches. *Nanotechnology Law & Business* 9: 393-410

Marchant. G.E. and Wallach, W. 2015. Coordinating Technology Governance. *Issues in Science & Technology*, Summer 2015: 430-450.

Marchant, G.E., and Wallach, W. 2016. Introduction. In *Emerging Technologies: Ethics, Law and Governance,* 1-12. London, U.K.: Routledge.

Russell, S. 2015. Value Alignment: Stuart Russell. Talk given at World Economic Forum meeting. Accessed October 15, 2017. https://www.youtube.com/watch?v=WvmeTaFc_Qw

Wallach, W. and Allen, C. 2008. *Moral Machines: Teaching Robots Right From Wrong.* New York, NY: Oxford University Press.

Wallach, W. 2016. Building Global Infrastructure to Ensure AI and Robotics Are Beneficial. Unpublished but widely circulated article.

Wallach, W. 2018. How to Keep AI from Slipping Beyond Our Control. Geneva, Switzerland: The World Economic Forum. A preliminary draft of this article has been available to WEF members during fall 2017. The final draft will be made available on the World Economic Forum's public website in January 2018.