SAVANNAH
LAW REVIEW
VOLUME 5 | NUMBER 1

RISE OF THE AUTOMATONS

Wendell Wallach[*]

If you listen to the techno-*optimists* you would think that we are on a highway to heaven on Earth and the self-driving buses are speeding up at an exponential pace. On the other hand, if you just listen to the techno-*pessimists* you would think we are going straight to hell in a handbasket.

But most of us do perceive technology as a source of promise and productivity, and yet there is considerable disquiet—disquiet about specific areas of research and disquiet about the overall trajectory of scientific discovery and technological innovation. You can see that disquiet in so many different areas: the worldwide prohibition on human cloning, the use of human growth hormones in sports, the E.U. has had its debate on genetically modified foods, and the U.S. is debating embryonic stem cell research. And there are the never-ending issues surrounding bio-security and info-security. Over the past few years we've had this dramatic uptake in concerns around what can happen with AI and the biotechnologies due to breakthroughs in machine learning algorithms and CRISPR Cas9, which dramatically eases the editing of the genome.

I see the self-driving car as giving us a metaphor for what we're dealing with. Technology is moving into the driver seat as a primary determinant of humanity's

[*] Chair, Technology and Ethics Studies, Yale University Interdisciplinary Center for Bioethics. Author, A DANGEROUS MASTER: HOW TO KEEP TECHNOLOGY FROM SLIPPING BEYOND OUR CONTROL (Basic Books, 2015). I was truly honored to give the plenary talk at *Rise of the Automatons*. The input of ethicists is not always valued by lawyers or law schools. The quality of the student and faculty presentations was excellent, and the southern hospitality of *Savannah Law Review* and its students could not have been surpassed. I was particularly touched when the students presented me with an oil portrait of my future cyborg self, painted by artist Fauvista Artista. Finally, let me express special appreciation to Marlan Eller who invited me, served as my primary host, and made sure that my trip to the Savannah Law School went very smoothly.

future. It's up to us to shape the trajectory of technological development. It is not inconceivable that we are on a path to inventing the human species, as we known it, out of existence. But hopefully we can come up with a more life-centric, biological-centric, environmental-centric, human-centric path.

I'm going to start out with a little bit about where we are to date with AI, just to make sure that we're all on the same page. First, I will very quickly review the three big issues that come up as concerns around artificial intelligence. Then I am going to turn more to nearer-term concerns and a few recommendations as to how we might manage those concerns through ethics, engineering, policy, and oversight.

The term AI was coined at a conference at Dartmouth College in 1956—a very small conference whose attendees are now seen as the fathers of the field, with Marvin Minsky and John McCarthy among those who gathered. They were a little naïve and amazingly optimistic. They thought that we were ten years away from having a computer program that could beat a grandmaster at chess. They also thought in ten years they would have computers communicating with natural language (sort of like Star Trek) where they would talk to the computer and it would talk back to them clearly answering whatever they asked it. Furthermore, they believed that if they assigned computer vision to one graduate student for the summer, he would solve that challenge. Well, it was forty years before Deep Blue beat Garry Kasparov at chess. To this day, we don't have computers that have full natural language intelligence, and we still don't have computers that have full visual capabilities—even though we've made significant strides in both of those capabilities. We now even have translation programs that are pretty satisfactory.

The field of AI has gone through a series of what are sometimes characterized as summers and winters. AI summers occur when there is a new trajectory which generates a great deal of excitement; everybody thinks we are going to make breakthrough after breakthrough. AI winter sets in when that does not turn out to be the case.

One of the most famous summers was the onset of neural networks. A neural network is a computer platform that uses multiple processors, each representing one neuron, in an attempt to emulate the thought processes of human beings. It was expected that neural networks would create great breakthroughs, but it didn't happen. For a couple decades, there were only a few scientists who were seriously engaged with neural networks. One of them was Goeffrey Hinton who was at the University of Toronto.

Goeff stayed with this approach until sufficient computing power made it possible to realize some of the earlier visions of what would happen with neural networks. He, for all practical purposes, is the father of what we call deep learning. All the excitement surrounding AI recently is due to this approach known as deep learning.

But let's be clear about what deep learning is and what it is not. Deep learning is not the unstructured learning method of children where they just wander around their environment learning from everything. It's a specific kind of structured learning, and it is only one sub-set among several different approaches to machine learning algorithms. What deep-learning algorithms do is they can look at a massive amount of data (in fact it's required that they look at a massive amount of data) about a particular subject, and they will find significant relationships within that data. Often those are significant relationships that humans would not discover or recognize without the help of great computing power.

The importance of deep learning is that it can be applied to any massive database, and for a wide variety of applications. In some applications it can be combined with other approaches to computing and new fruitful paths can be realized. For example, deep learning was combined with other learning approaches to create AlphaGo (a computer program). Many of you may have heard of Go. It is a popular game played in Asia, and it's more difficult than Chess. AlphaGo first got a great amount of publicity when it beat Lee Sedol, one of the greatest players in the world, in a five game match.

I sat on a panel with Lee Sedol in 2016. Even though he lost that tournament to this computer program, I was in awe of what he had achieved. Here was a man who perhaps in his life had played 15,000–20,000 games of Go and maybe studied another 5,000–10,000. He actually won one game against AlphaGo and almost won another. By that time, AlphaGo had played a million and a half games of Go. But it didn't really *play* Go in the way a human does. It engaged in a kind of logical processing that it had deduced from playing all those games. It was not actually playing the game of Go, that is, it didn't understand that it was playing Go the way you or I would. We only know that by a standard we call "winning," it won.

All of this is a way of underscoring that even if we lose to a machine, that doesn't necessarily mean that our remarkable human faculties have been fully recreated. We, humans, run on roughly twenty watts of energy. There is unbelievable efficiency going on in our bodies, and, in many respects we still can solve all kinds of problems that computers cannot. In fact, the only cognitive capability that AlphaGo has emulated or solved is the problem of perception. But other higher order problems such as common sense, planning, analogies, reasoning, language—those have not been solved. Unstructured learning has not been solved, which is essential to emulate human mental prowess. An earlier speaker talked a little bit about the Chinese room today. The Chinese room was about the distinction between syntactical programming and semantic understanding—the capacity to *understand* what you are doing. There is no sense in which AlphaGo *understood* when it was playing Go, and there was no sense in which Deep Blue knew it was playing Chess—or at least not in the way a human would understand the game.

There exists a plethora of thresholds that have not been crossed by AI systems, and the question remains as to how quickly computers are going to cross them. Of course, no one knows. The deep learning breakthrough has been truly exciting and we will have a lot of machines doing astounding things over the next decade or two. But whether or not all human cognitive capabilities are within the capacity of AI systems, we really don't know. It is all speculation that AI will exceed human capabilities in all respects. For some theorists, reverse reasoning where they argue that each human is a machine; therefore, the machines that we create should be able to eventually solve all problems, seems satisfactory.

None of this, however, alters the fact that this breakthrough in deep learning has got many people wondering again:

What happens if the AI engineers actually succeed?

What will happen when machines can do everything that we can do?

What will happen when we have superintelligence, machines far superior to humans in their cognitive capabilities?

The original fathers of AI were not interested in creating little discrete machines that could each do different tasks. They wanted what is now referred to as artificial *general* intelligence. They thought that they were really only a few decades away from a machine that could do everything that we can do and that were comparable to us in all capabilities. This is now often referred to as superintelligence largely because computers will very quickly go from human-level intelligence to being far superior to humans. Just look at the fact that they're going to be able to work 24/7. AlphaGo was able to play a million and a half games with, by the way, a tremendous warehouse of energy and infrstructure behind it—not as simple as our twenty watts.

You've all been hearing these stories—and you've all watched movies like *The Terminator*—so the possibility that superintelligence might be dangerous for humanity in on your mind. Some suggest that we will produce superintelligence very quickly, and we should start planning now. One of those people is Elon Musk. Elon Musk has been particularly vocal in the last year or so. Elon Musk also happens to be one of the people who is investing the most money in developing AI. He has said that we need the government to oversee the development of AI.

So what is Elon saying? Is he saying, "Stop me before I destroy humanity"?

Not all the AI researchers got into the field because they dreamed of building AI with superintelligence. There is widespread disagreement in the field regarding how far we will progress simulating human intelligence by computational means. I characterize my position on that as your friendly skeptic—friendly to the engineering spirit that will produce remarkable things, but skeptical that we know enough about human intelligence to know whether it can be fully simulated. Yet, that does not mean that I think we should ignore potential dangers.

My friend Stuart Russell is one of the leading AI researchers who, with Peter Norvig (V.P. of Google) wrote the textbook on Artificial Intelligence that everybody who studies AI learns from [*Artificial Intelligence: A Modern Approach* is in its Third Edition]. He asked: "What if we have a 10% chance that we can reach superintelligence in the next 15 years? Shouldn't we start working now to be sure that we can control it?" He has started a whole new trajectory within the field of AI, which until now has been about getting machines to do what the engineers want them to do—get them to function properly. Russell started what is called the AI Safety research program—research directed more at safety, controllability, and transparency. Given my focus, I think this is the right course. At the present stage in AI development, the safety of superintelligent systems is not a problem for the government or lawyers, but it is a challenge for engineers.

Think about whether to let government get involved in this stage of the game. How many of you would have wanted us to restrict the past *sixty years* of research in genetics and in computer science based on 1950s fears of robot takeovers and giant mutant locusts?

The second issue we hear a great deal about is weaponized AI, particularly lethal autonomous weapons and cyberwarfare.

First, let us discuss lethal autonomous weapons systems [LAWS]. LAWS do not refer to weapons systems that necessarily have a great deal of intelligence. It simply refers to weapon systems that can pick their own target and kill people without immediate human involvement in the decision-making process. The debate is largely about what is meaningful human control? How close do the decision makers have to be to give the kill order or the selection of the target order, and to what extent it is acceptable for them to delegate that to machines? The debate on this has been going on in the UN for the last three years at The Convention on Certain Conventional Weapons, which oversees treaties to regulate various weapon systems, including biological weapons and nuclear weapons. An international campaign has been initiated to try to ban what are known to the public as "killer robots."

There's a great deal of resistance by strategic planners to ban LAWS. However, there are also many military leaders on the front lines who say that this is nuts because they wouldn't have robust command or control. Furthermore, soldiers are asking where virtue lies in going onto a battlefield against a weaponized machine gunner. A LAW may need as little as a millisecond to fire or at best a few milliseconds. We, humans, need roughly 250–400 milliseconds to react to anything. The difference between 250 versus 400 milliseconds is when we are already prepared to react versus something we have to think about first.

I am among those who champion a ban on LAWS, or at least dramatic restrictions on their deployment. When people think about LAWS, they often think about drones hanging out over a village. Using facial recognition software, the drone picks out and kills the terrorists when they appear. But *any* weapons system can be a lethal autonomous weapons system—including a nuclear weapon or an unmanned, nuclear-powered submarine that could launch high-powered munitions. At the very least, there is need to outlaw some LAWS if not all LAWS. The argument should be over how broadly we extend the ban. For all of you interested in getting into international law, the arguments are cast in the language of international humanitarian law, also referred to as the Laws of Armed Conflict.

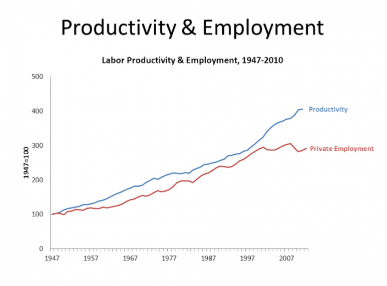
A series of different accords and treaties have been developed (some concerns that really date back to the Roman era). The arguments in international circles are largely centered on whether LAWS could commit atrocities by breaking international humanitarian law, and whether they may lower the barriers for starting new wars or even escalate an existing conflict in ways that no one ever intended.

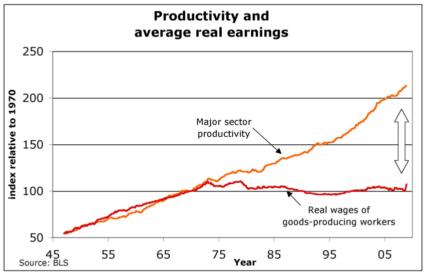
When I testified at the expert meetings of the Convention on Certain Conventional Weapons, my focus was the unpredictability of LAWS. LAWS are complex adaptive systems. They can sometimes react in ways that we cannot anticipate. And if they are learning systems, they may alter their behavior based on what they've learned. So, you have no way of testing them and no way of totally validating whether these systems, in fact, are working as intended.

Cyber warfare has been talked about a bit today. Weaponized AI is a term for this unholy alliance among social networks, cyber warfare, hacking systems, and fake news and what can result from that. In the U.S. we are just beginning to understand that this is going on all over the world.

Let me move on to technological unemployment. Technological unemployment was John Maynard Keynes' term for the longstanding Luddite fear that each new technology was going to rob more jobs than it created. He used the term in a 1930s paper entitled, *Technological Possibilities for Our Grandchildren*. Roughly 200 years have passed since the original Luddite Rebellion, and it hasn't happened yet. Each new technology has created more jobs than it has robbed. But I'm among those who say it's different this time. It was only a few years ago that if I were to say that, there would have been vociferous critics, and there may still be some in this audience today. I am fascinated to see today, how it was more or less taken for granted that some jobs are going to be decimated by AI-enabled automation.

What you see in the two graphs below is often referred to as the "great decoupling." That's the decoupling between productivity and job creation or wage growth. For a long time, wage gains and job gains moved in parallel with productivity gains—they don't anymore. Furthermore, the productivity gains we do witness are either anemic or returning to baseline, which isn't really all that fast, somewhere around 2.5% per year.





Source: BLS

Research by Osborne and Frey published by the Martin School at Oxford University concluded that 47% of U.S. jobs were automatable based on present day technology. Later on they applied the same methodology to England. Surprisingly, only 37% in the UK were automatable. In 2015, their methodology indicated that in India 69% of jobs were automatable, and in China 77% of jobs. The debate underway focuses on: "How true is this? How quickly will this happen?"

Other methodologies suggest how these figures are computed may exaggerated the impact of automation. The Organisation for Economic Co-operation and Development (OECD) did some research using a different methodology—rather than looking at occupations, it looked at tasks that might contribute to an occupation. The OECD decided that most occupations have certain tasks that are not automatable. Looking across 21 countries, the OECD came up with the average of 9% of jobs that were fully automatable by technology presently available.

I think the truth probably lies between these two ranges, as it often does. Consider a law firm, a large law firm that has lots of paralegals. The computers can't do all the tasks paralegals do. They can, however, do about 70% of those tasks. So, let's just imagine the firm has ten paralegals engaged in property searches and other tasks. While it can't replace all the paralegals, this will not stop the firm from getting rid of maybe six of those individuals, maybe more. So that's what we're more likely to see. We are more likely to see jobs re-organized so that those tasks that cannot be automated are performed by one individual and those that can be automated contribute to overall staff reductions. I think that's what you're going to see over the next few decades. Imagine if 25% of jobs in the United States were decimated in the next twenty or thirty years. That would be truly dramatic. Could we create 25% new jobs? It would actually have to be more than that, based on how many new people enter the workforce yearly.

That brings us to another challenge—self-driving vehicles. Long before we will have large numbers of self-driving cars on city streets, we're going to have self-driving trucks on highways. Truckers happen to be one the largest occupations in the United States (1.7 million of those truckers are long-distance haulers). Let's imagine that over 10 years all long-distance trucks are either replaced by self-driving trucks or upgraded to self-driving trucks. Maybe it will take 20 years, maybe it will be longer, but let's say for purposes of this example that it takes 10 years. That means that we are going to need to create 14,000 new jobs every month suitable for truckers—and that's in addition to all the new jobs we will need for new people entering the work place and other jobs that have been decimated. We've never seen the creation of new kinds of work at such a sustained pace. This would be a totally different adventure.

Among the more immediate concerns is the transparency of learning algorithms. Learning algorithms largely require massive data input from which specific outputs are determined. But no one fully understands what happens between that input and output. The algorithms can't tell you, and the computer scientists can't go back and tell you what the learning system has done. There is no transparency. Those in the E.U. have become concerned about this, and they have enacted a law, the General Data Protection Requirements [GDPR], which specifies all systems must be explainable. Many computer scientists are trying to figure out how to get at least a degree of transparency in learning algorithms. They've had indirect success, not great success.

Here's one difficulty: if a learning algorithm performs better than a human, should it be deployed or shouldn't it? We probably wouldn't mind deployment in contexts where no harm is going to come to humans. But there is no guarantee of that. If a self-driving vehicle is constantly analyzing its environment, and it's getting more and more accurate data so it's more efficient in what it's doing—that's okay. But if it has an accident, at the *very least* we would like to have forensic data wouldn't we? We'd want an explanation as to why that accident occurred in order to find means to ensure that that accident would not reoccur.

Learning algorithms are going to be deployed to analyze a vast array of existing databases, yet there is no actual oversight in place as to which systems do and which systems do not get employed. Furthermore, which systems should or should not get deployed may change depending upon what they've learned. *There* is a new occupation for all you young lawyers in the room. Get involved in thinking through the criteria for deploying or not deploying systems that don't have transparency.

Here is another dilemma: if the overall impact of a learning algorithm is good but it carries some dangers, should we be willing to go ahead? For example, some analysts contend that self-driving vehicles will have 93% fewer accidents based on information from National Highway Traffic Safety Administration (NHTSA (2008)), which indicated that human error is a factor in 93% of accidents. If we could really have 93% less accidents, think of how many lives would be saved. We're approaching 40,000 lives lost yearly in the U.S. alone. But self-driving vehicles aren't ready for prime time. If you put them on city streets today, they would kill people that attentive human drivers would not. But if they're really going to save even 50% of the lives presently lost, maybe we should be willing to accept that—even if they aren't fully ready at this time.

In other words, if the utilitarian calculation is that the benefits will outweigh the harms, should we accept systems that have very large benefits even if they have some harms? That's not something that lawyers are going to solve alone. That is going to require a public conversation, a creation of new norms. I'm not arguing that utilitarian benefits should trump all dangers, but merely that we may need to consider whether this might be acceptable for self-driving cars or other applications.

The next area: data. Data, as they say, has become "the new oil." He who owns a great deal of data is rich. The owners of the most data are the IT companies. The IT oligopoly will be the AI oligopoly. The IT oligopoly already has more power than the oil companies ever did, and they are going to continue accumulating additional power. They have the resources to do so. This raises issues of ownership, power and the responsibility of companies to the people whose data they hold.

What happens if Google goes into a small African country and says we will give you all of these resources such as wireless connectivity; we just want all your data. Will that country even know [the value of] what they are giving up? And will the short-term benefits they derive justify accepting the deal? This has happened. Google has put high speed internet in every major train station in India. Even if they didn't make a deal with the Indian government, this 'gift' has given them [Google] tremendous access to the data of the Indian people.

Another issue that has been given considerable attention recently is algorithmic bias. The biases are largely there not because of the algorithms, but because they are implicit in the databases the algorithms are looking at from the very beginning. There is some question as to whether any means to compensate for those biases would just be biases themselves. What to do about bias is a serious problem. It is possible to perform some data analytics to determine what biases are in the initial dataset. But some stakeholders might argue that artificial intelligences should have the same prejudices that we have for certain applications. Biases within datasets poses technical, legal, and a governance challenges.

Then there are questions as to who has a right to your data? Do Facebook and Google own the data on their sites? Do you have a right to opt out? Should some of the existing HIPAA and other restrictions on who can look at your data, as specified by the FDA, be overridden because the conglomeration of data might be helpful for curing diseases or to help those who are least advantaged? These are among the debates that our society needs to have, and they have already started.

Not all solutions or suggestions relating to these challenges are legal. Sometimes ethics is the way forward. Sometimes we need the society to set new standards or new norms. Sometimes we need to reinforce existing ethical principles that embody what we hold dear.

Consider technological unemployment, for example. It's not only a question of what to do about all of the people for whom wages are no longer the way they get goods and services, but also whether they will have meaningful lives. Many people get some sense of fulfillment and meaning through their work. If jobs are going to be decimated, from where will people derive meaning and purpose? What kind of world do we want to create? This is largely an ethical question, a discussion about what kind of world we want to create and what values we need to reinforce to nudge the world in that direction.

Consider who is responsible for deploying an autonomous system. First the legal question: who is responsible and potentially culpable and liable? It looks like the manufacturers are largely responsible, but we are teetering into a zone where autonomous systems threaten to undermine the foundational principle that there is an agent (either individual or corporate) that is responsible, and potentially liable, for the deployment of any artefact.

There are ways of getting around this dilemma. For example, when engineering systems, rather than simply looking at the technical aspects of designing the system, the engineers can also treat ethical considerations as design features. Let's say that the engineers were told that part of the design specification is that you design a system that will clarify who will be responsible and potentially liable if that system fails. Designing for responsibility could be treated as a feature, the same as building in a servo that doesn't overheat. We can also design for privacy. We can design for all kinds of values, but it requires educating engineers in a new way, or bringing social scientists who are more sensitive to the societal impacts of these technologies onto the design teams.

Another option is to make the systems, themselves, sensitive to human ethical and legal considerations and factor those into their choices and actions. This is the subject of my 2008 book, titled *Moral Machines: Teaching Robots Right from Wrong*, which I co-authored with Colin Allen. *Moral Machines* is about mapping this new field of inquiry called machine morality, machine ethics, and more recently value alignment.

To be honest, the book was just as much about human decision making and ethics. It may have been the first book to look comprehensively at ethical human decision making. Ethical decision making for humans takes certain cognitive capabilities for granted—emotions, theory of mind, consciousness, empathy.

Moral psychology was just beginning to be studied during the era when we wrote *Moral Machines*. We recognized that computers and robots making moral decisions was not merely about what ethical theories were utilized by the machine, but it was also about what function these other capacities served. Could we instantiate, for example, the functional equivalent of empathy or consciousness in artificial intelligence?

A number of people turned to us and said: "this is a simple problem, wasn't it already solved by Isaac Asimov with his three laws for robots?" But nearly all of Asimov's stories were about how systems programmed with these three laws failed. The second law, for example, is to obey humans. One story was about a robot breaking down because it had two conflicting orders from two different humans. Asimov invented a robot psychologist named Susan Calvin to analyze why this robot went into the equivalent of a catatonic fit. In story after story, Asimov effectively demonstrated that a simple ruled based ethical system would not be sufficient.

Building moral machines has suddenly caught on with AI researchers and it is referred to by them as the "value alignment problem." The first time I heard Stuart Russell use this term, I went up to him afterwards and I said, "that sounds like the bottom-up approach to machine ethics." He looked at me like he had no idea what I was talking about. This anecdote underscores another problem—we lack the transdisciplinary knowledge from other people that have done research in areas that complement our own. My interaction with Stuart Russell evolved into a research project and a series of workshops where we bring leaders in these various fields together so that they can learn from each other's insights. Hopefully we can move toward a more collaborative community, but it's very hard to get scientists to understand that maybe the social sciences have something to contribute.

Joi Ito, head of the MIT Media Lab, and a few other people wanted to create a transdisciplinary community where scientists and social scientists work together in designing a research project. They were willing to give substantial funds to transdisciplinary teams, but hardly anyone understood what they were asking for.

The computer scientists said, "the social scientists can tell us what their problem is, and then we'll go and solve it." That isn't what Joi Ito had in mind. There is need for some of you in this audience to develop adequate technical skills so that you can work together with engineers and help them appreciate what you understand about the societal impact that their inventions may have.

Let me move on to governance. There's a total misfit between existing governmental systems and the rapid pace of scientific discovery and technological innovations. This is sometimes referred to as the pacing problem. It's the growing gap between emerging technology and legal-ethical oversight.

There are a lot of reasons why we have this pacing problem. One is that legislators don't understand the sciences. Furthermore, legislators in the U.S. are disinclined to regulate in any way that might interfere with development. European leaders don't have the same disinclination. This gives Americans a lead if we don't put regulatory reins on, and we wait until after the first crisis happens and then we try to regulate. The other problem is that there is such a rapid and vast array of technology appearing that it's almost impossible to address them all.

Gary Marchant is the director of the Center for Law, Science, and Innovation at Sandra Day O'Connor School of Law at Arizona State University. He and I were sitting together one day and rather than lamenting the state of affairs in our governance of emerging technologies, we turned the question on itself and asked "if we had our choice, what do we need—what kind of governance would we try to implement?" In response to that question we developed a model that we call "Governance Coordinating Committees."

The idea was to create some kind of new entity that would coordinate the activities of the many stakeholders, function as a good-faith broker, and monitor the field to look for gaps that perhaps were not being addressed by anybody else. A central concern is to not usurp the authority of any other body that would take on the responsibilities for addressing issues, but just clarify who was taking on the responsibility for what, and what was not being dealt with. As new challenges came up we would try to look at an array of mechanisms from feasible technological solutions to industry oversight to hard governance and determine the best means to address the gap.

We considered hard law and regulatory oversight a *last* resort. In fact, what we really wanted to build on was (and this is Gary's specialty) soft law. Soft law and soft governance refers to industry standards, laboratory practice and procedures, insurance policies, and a whole plethora of mechanisms which can be much more adaptive than hard laws.

One difficulty has been that soft governance usually has no enforcement mechanism or seldom has an enforcement mechanism. But implicit in soft law is nudging industry to be the responsible agent. This, however, underscores the second issue with soft law, that of the fox guarding the henhouse. One way forward might be to forge within hard law a means to enforce violations of well-established standards, such as those of the IEEE [Institute of Electrical and Electronic Engineers]. The IEEE sets all kinds of standards such as the size of [light]bulbs and how much electricity can go over various gages of wire. Over the past couple years, IEEE has jumped into AI, hosting conferences and work groups. Over 250 people have participated in these activities and weighed in proposing new standards and ethical guidelines for AI.

It's heartening to see the issues in AI receive attention.

Any new governance mechanism will pose implementation challenges. From where would this body get its authority? Its legitimacy? Its influence? How would the officers and administrators be selected? Would it be governmental, private, or perhaps some conglomeration of the two? Who would fund it and to whom would it be accountable? And all of these implementation challenges bring up the basic reaction: this is just too complicated or this is hopelessly naïve. Perhaps it is both too complicated and hopelessly naïve. Nevertheless, something like this is needed, and we've done complicated things in the past. After all, the U.S. did land a man on the moon.

Gary and I recommend that we focus on two areas as pilot projects—one area being robotics and AI and the other area being synthetic biology, which is the creation of genetic products and genetically modified organisms. These are both relatively new areas of research and neither of them is heavily encumbered by laws and regulations. In each of these fields, we have had major breakthroughs recently—CRISPR Cas9, which dramatically eased and sped up editing of the genome—and advances in machine learning, particularly deep learning in AI. The proposal for these pilot projects goes back about 4–5 years now including a significant paper in the journal, *Issues in Science and Technology*. We heard often what a good idea this is and that somebody should work to implement it, but nobody did anything.

I woke up one morning, and I thought "guess what, you're it." You can no longer wait for somebody else to act. Furthermore, it wouldn't be satisfactory if a GCC is started as a national project, hopeful that other countries will follow suit. Thus, I initiated what's called the project to Build Global Infrastructure to Ensure that Artificial Intelligence and Robotics are Beneficial. Some of you may consider that "beneficial" word a little clumsy. It actually comes from the AI researchers who didn't want to admit that AI might be problematic, so they called for "beneficial AI."

My partners in this endeavor come from The Hastings Center, which was the first bioethics think tank, and also the Carnegie Council for Ethics in International Affairs. We already have advisors including the heads of the IEEE AI initiative, the Director General of the United Nations in Geneva and a few other significant figures. In addition, we are discussing with various countries who might host an International Congress for the Governance of AI.

In all of this it is important to keep in mind that technological development can both stagnate or overheat. A central role for public policy and law is to modulate the rate of development. Governments can stimulate development through investment and can slow it down through regulation. However, if technological development is truly accelerating, as it appears to be, then the need for foresight and planning becomes pressing.