AMITAI ETZIONI
OREN ETZIONI

Should Artificial Intelligence Be Regulated?

New technologies often spur public anxiety, but the intensity of concern about the implications of advances in artificial intelligence (AI) is particularly noteworthy. Several respected scholars and technology leaders warn that AI is on the path to turning robots into a master class that will subjugate humanity, if not destroy it. Others fear that AI is enabling governments to mass produce autonomous weapons—"killing machines"—that will choose their own targets, including innocent civilians. Renowned economists point out that AI, unlike previous technologies, is destroying many more jobs than it creates, leading to major economic disruptions.

There seems to be widespread agreement that AI growth is accelerating. After waves of hype followed by disappointment, computers have now defeated chess, Jeopardy, Go, and poker champions. Policymakers and the public are impressed by driverless cars that have already traveled several million miles. Calls from scholars and public intellectuals for imposing government regulations on AI research and development (R&D) are gaining traction. Although AI developments undoubtedly deserve attention, we must be careful to avoid applying too broad a brush. We agree with the findings of a study panel organized as part of Stanford University's One Hundred Year Study of Artificial Intelligence: "The Study Panel's consensus is that attempts to regulate 'AI' in general would be misguided, since there is no clear definition of AI (it is not any one thing), and the risks and considerations are very different in different domains."

One well-known definition is: "Artificial intelligence is that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment." A popular understanding of AI is that it will enable a computer to think like a person. The famous Turing test holds that AI is achieved when a person is unable to determine whether a response to a question he or she asked was made by a person or a computer. Others use the term to refer to the computers that use algorithms to process large amounts of information and draw conclusions and learn from their experiences.

AI is believed by some to be on its way to producing intelligent machines that will be far more capable than human beings. After reaching this point of "technological singularity," computers will continue to advance and give birth to rapid technological progress that will result in dramatic and unpredictable changes for humanity. Some observers predict that the singularity could occur as soon as 2030.

One might dismiss these ideas as the provenance of science fiction, were it not for the fact that these concerns are shared by several highly respected scholars and tech leaders. An Oxford University team warned: "Such extreme intelligences could not easily be controlled (either by the groups creating them, or by some international regulatory regime)… the intelligence will be driven to construct a world without humans or without meaningful features of human existence. This makes extremely intelligent AIs a unique risk, in that extinction is more likely than lesser impacts." Elon Musk, the founder of Tesla, tweeted that: "We need to be super careful with AI.

Potentially more dangerous than nukes." He added: "I'm increasingly inclined to think there should be some regulatory oversight [of AI], maybe at the national and international level." Oxford philosopher Nick Bostrom believes that just as humans out-competed and almost completely eliminated gorillas, AI will outpace human development and ultimately dominate.

Attorney and legal scholar Matthew Scherer calls for an Artificial Intelligence Development Act and the creation of a government agency to certify AI programs' safety. The White House organized four workshops on AI in 2016. One of the main topics: does AI need to be regulated?

The AI community has not been indifferent to these concerns. In 2009, the president of the Association for the Advancement of Artificial Intelligence (AAAI) appointed a panel of leading members to examine "the value of formulating guidelines for guiding research and of creating policies that might constrain or bias the behaviors of autonomous and semi-autonomous systems so as to address concerns." Some called for a pause, but in the end the AI researchers decided that there was not yet any reason for concern or to halt research.

As we see it, the fact that AI makes machines much smarter and more capable does not make them fully autonomous. We are accustomed to thinking that if a person is granted more autonomy—inmates released from jails, teenagers left unsupervised—that they may do wrong because they will follow their previously restrained desires. In contrast, machines equipped with AI, however smart they may become, have no goals or motivations of their own. It is hard to see, for instance, why driverless cars would unite to march on Washington. And even if an AI program came up with the most persuasive political slogan ever created, why would this program nominate an AI-equipped computer as the nominee for the next president? Science fiction writers might come up with ways intelligence can be turned into motivation, but for now, such notions probably should stay where they belong: in the movies.

One must further note that regulating AI on an international level is a highly challenging task, as the AI R&D genie has already left the bottle. AI work is carried out in many countries, by large numbers of government employees, business people, and academics. It is used in a great variety and number of machines, from passenger planes to search engines, from industrial robots to virtual nursing aids.

Most important, one must take into account that restrictions on the development of AI as a field are likely to impose very high human and economic costs. AI programs already help detect cancer, reduce the risk of airplane collisions, and are implemented into old-fashioned (that is, nonautonomous) cars' software that makes them much safer.

In a study in which a robot and human surgeons were given the same task (to sew up part of an intestine that had been cut), the robot outperformed the humans. Although the surgeons did step in to assist the Smart Tissue Autonomous Robot in 40% of the trials, the robot completed the task without any human intervention 60% of the time, and the quality of its stiches was superior.

AI is used in search and rescue missions. Here algorithms are used to survey aerial footage of disaster zones to identify quickly where people are likely to be stranded, and the increased speed means that there is a better chance that the victims will be found alive.

AI-equipped robots are used in child, elder, and patient care. For example, there are robotic "pets" used to reduce stress for elderly patients with dementia. The pets are programmed to learn how to behave differently with each patient through positive and negative feedback from the patients. AI is also used in the development of virtual psychotherapists. People appear more willing to share information in a computer interview because they do not feel judged the same way they might in the presence of a person.

Computerized personal assistants such as Apple's Siri, Microsoft's Cortana, and Amazon's Alexa use AI to learn from their users' behavior how to better serve them. AI is used by all major credit card companies in fraud detection programs. Security systems use AI programs to surveil multiple screens from security cameras and detect items that a human guard often misses.

One must weigh losses in all these areas and in many others if AI research were to be hindered as part of hedging against singularity. It follows that although there may be some reasons to vigilantly watch for signs that AI is running amok, for now, the threat of singularity is best left to deliberations during conferences and workshops. Singularity is still too speculative to be a reason at this time to impose governmental or even self-imposed controls to limit or slow down development of AI across the board.

Autonomous killing machines?

In contrast, suggestions to limit some very specific applications of AI seem to merit much closer examination and action. A major case in point is the development of autonomous weapons that employ

AI to decide when to fire, with how much force to apply, and on what targets.

A group of robotics and AI researchers, joined by public intellectuals and activists, signed an open letter that was presented at the 2015 International Conference on Artificial Intelligence, calling for the United Nations to ban the further development of weaponized AI that could operate "beyond meaningful human control." The letter has over 20,000 signatories, including Stephen Hawking, Elon Musk, and Noam Chomsky, as well as many of the leading researchers in the fields of AI and robotics. The petition followed a statement in 2013 by Christof Heyns, the UN special rapporteur on extrajudicial, summary, or arbitrary executions, calling for a moratorium on testing and deploying armed robots. Heyns argued that "A decision to allow machines to be deployed to kill human beings worldwide, whatever weapons they use, deserves a collective pause."

A pause in developing killing machines until the nations of the world come to agree on limitations on the deployment of autonomous weapons seems sensible. Most nations of the world have signed the Treaty on the Non-Proliferation of Nuclear Weapons, which was one major reasons that several nations, including South Africa, Brazil and Argentina, dropped their programs to develop nuclear weapons and that those who already had them reduced their nuclear arsenals. Other relevant treaties include the ban on biological and chemical weapons and the ban on landmines.

We note, though, that these treaties deal with items where the line between what is prohibited and what is not covered is relatively clear. When one turns to autonomous weapons, such a line is exceedingly difficult to draw. Some measure of autonomy is built into all software that uses algorithms, and such software is included in numerous weapon systems. At this point, it would be beneficial to discuss three levels of autonomy for weapons systems. Weapons with the first level of autonomy, or "human-in-the-loop systems," are in use today and require human command over the robot's choice of target and deployment of force. Israel's Iron Dome system is an example of this level of autonomy. The next level of weapons, "human-on-the-loop," may select targets and deploy force without human assistance. However, a human can override the robot's decisions. South Korea has placed a sentry robot along the demilitarized zone abutting North Korea whose capabilities align with this level of autonomy. Finally, there is the level of fully autonomous weapons that operate entirely independent of human input. It seems worthwhile to explore whether the nations of the world, including Russia, China, and North Korea, can agree to a ban on at least fully autonomous weapons.

We suggest that what is needed, in addition, is a whole new AI development that is applicable to many if not all so-called smart technologies. What is required is the introduction into the world of AI the same basic structure that exists in practically all non-digital systems: a tiered decision-making system. On one level are the operational systems, the worker bees that carry out the various missions. Above that are a great variety of oversight systems that ensure that the work is carried out within specified parameters. Thus, factory workers and office staff have supervisors, businesses have auditors, and teachers have principals. Oversight AI systems—we call them AI Guardians—can ensure that the decisions made by autonomous weapons will stay within a predetermined set of parameters. For instance, they would not be permitted to target the scores of targets banned by the US military, including mosques, schools, and dams. Also, these weapons should not be permitted to rely on intelligence from only one source.

A pause in developing killing machines until the nations of the world come to agree on limitations on the deployment of autonomous weapons seems sensible.

To illustrate that AI Guardians are needed for all smart technologies, we cite one example: driverless cars. These are designed as learning machines that change their behavior on the basis of their experience and new information. They may note, for instance, that old-fashioned cars do not observe the speed limits. Hence, the driverless cars may decide to speed as well. The Tesla that killed its passenger in a crash in Florida in 2016—the first known death attributed to a driverless car—was traveling nine miles per hour over the speed limit, according investigators from the National Transportation Safety Board. An oversight system will ensure that the speed limit parameter will not be violated.

One may argue that rather than another layer of AI, human supervisors could do the job. The problem is that AI systems are an increasingly opaque black box. As Viktor Mayer-Schönberger and Kenneth Cukier note in their book *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, "Today's

computer code can be opened and inspected ... With big-data analysis, however, this traceability will become much harder. The basis of an algorithm's predictions may often be far too intricate for most people to understand." They add that "the algorithms and datasets behind them will become black boxes that offer us no accountability, traceability, or confidence." Jenna Burrell from the School of Information at the University of California, Berkeley, distinguishes three ways that algorithms become opaque: intentional opacity, where, for example, a government or corporation wants to keep secret certain proprietary algorithms; technical illiteracy, where the complexity and function of algorithms is beyond the public's comprehension (and, we add, even by experts unaided by AI); and scale of application, where "machine learning" or the number of different programmers involved, or both, renders an algorithm opaque even to the programmers. Hence, humans will need new, yet-to-be-developed AI oversight programs to understand and keep operational AI systems in line. A fine place to start is keeping autonomous weapons under control. Also, only an AI oversight system can move fast enough to make a split-second decision to stop a mission in real time—for example, if a child runs into the target area.

One may wonder if the oversight AI systems are not subject to the same challenges faced by the first-line systems. First of all, it helps to consider the purpose and design of the different categories of AI. First-line AI programs are created to increase the efficiency of the machines they guide, and users employ them with this goal in mind. In contrast, AI oversight systems are designed and employed, well, to oversee. Moreover, just like human auditors, various programs build a reputation as being either more trustworthy or less so, and those that are less reliable are less used by those who do seek oversight. And just as in the auditing business, there is room in the field of AI for a third layer of overseers, who could oversee the lower-level oversight system. However, at the end of the day, AI cannot solve the issue raised by philosophers in Ancient Greece—namely "who will guard the guardians?" Ultimately, we are unaware of any way to construct a perfect system.

Finally, this is not meant to leave humans out of the loop. They not only are the ones to design and improve both operational and oversight AI systems, but they are to remain the ultimate authority, the guardian of the AI Guardians. Humans should be able to shut down both operational and oversight AI systems—for example, shutting down all killing machines when the enemy surrenders, or enabling a driverless car to speed if the passenger is seriously ill.

Finally, we hold that the study of killing machines should be expanded to include the opposite question: whether it is ethical to use a person in high-risk situations when a robot can carry out the same mission as well, if not better. This question applies to clearing mines and IEDs, dragging wounded soldiers out of the line of fire and civilians from burning buildings, and ultimately, fighting wars. If philosophers can indulge in end-of-the-world scenarios engineered by AI, then we can speculate about a day when nations will send only nonhuman arms to combat zones, and the nation whose machines win will be considered to have won the war.

Job collapse?

Oddly, the area in which AI is already having a significant impact and is expected to have major, worldwide, transformative effects is more often discussed by economists rather than by AI mavens. There is strong evidence that the cyber revolution, beginning with the large-scale use of computers and now accelerated by the introduction of stronger AI, is destroying many jobs: first blue-collar jobs (robots on the assembly line), then white-collar ones (banks reducing their back office staff), and now professional ones (legal research). The Bureau of Labor Statistics found that jobs in the service sector, which currently employs two-thirds of all workers, were being "obliterated by technology." From 2000 to 2010, 1.1 million secretarial jobs disappeared, as did 500,000 jobs for accounting and auditing clerks. Other job types, such as travel agents and data entry workers, have also seen steep declines due to technological advances.

The legal field has been the latest victim, as e-discovery technologies have reduced the need for large teams of lawyers and paralegals to examine millions of documents. Michael Lynch, the founder of an e-discovery company called Autonomy, estimates that the shift from human document discovery to e-discovery will eventually enable one lawyer to do the work that was previously done by 500.

These developments by themselves are not the main concern; job destruction has occurred throughout human history, from the weaving loom replacing hand-weaving, to steam boats displacing sail boats, to Model T cars destroying the horse-and-buggy industries. The concern, however, is that this time the new technological developments will create few new jobs. A piece of software, written by a few

programmers, does the work that was previously carried out by several hundred thousand people. Hence, we hear cries that the United States and indeed the world are facing a job collapse and even an economic Armageddon.

Moreover, joblessness and growing income disparities can result in serious societal distributions. One can see already that persistently high levels of unemployment in Europe are a major factor in fomenting unrest, including an increase in violence, political fragmentation and polarization, a rise in anti-immigrant feelings, xenophobia, and anti-Semitism.

Some economists are less troubled. They hold that new jobs will arise. People will develop new tastes for products and especially services that even smart computers will be unable to provide or produce. Examples include greater demand for trained chefs, organic farmers, and personal trainers. And these economists point out that the unemployment rate is quite low in the United States increased significantly, to which the alarmed group responds by pointing out that the new jobs pay much less, carry fewer benefits, and are much less secure.

Given the significance and scope of the economic and social challenges posed by AI in the very immediate future, several measures seem justified. The research community should be called on to provide a meta-review of all the information available on whether or not the nation faces a high and growing job deficit. This is a task for a respected nonpartisan source, such as the Congressional Research Service or the National Academy of Sciences. If the conclusion of the meta review is that major actions must be undertaken to cope with the side effects of the accelerating cyber revolution, the US president should appoint a high-level commission to examine what could be done other than try to slow down the revolution. The Cyber Age Commission that we envision would be akin to the highly influential 9/11 Commission and include respected former officials from both political parties, select business chief executive officers and labor leaders, and AI experts. They would examine alternative responses to the looming job crisis and its corollaries.

Some possible responses have been tried in the past, including helping workers find new jobs rather than trying to preserve the jobs of declining industries. In the United States, for example, Trade Adjustment Assistance for workers provides training and unemployment insurance for displaced workers. Another option would be government efforts to create jobs through major investments in shoring up the national infrastructure, or by stimulating economic growth by printing more money, as Japan is currently attempting.

More untested options include guaranteeing everyone a basic income (in effect, a major extension of the existing Earned Income Tax Credit); shorter work weeks (as France did but is now regretting); a six-hour workday (which many workplaces in Sweden have introduced to much acclaim); and taxes on overtime—to spread around whatever work is left. In suggesting to Congress and the White House what might be done, the commission will have to take into account that each of these responses faces major challenges from deeply held beliefs and powerful vested interests.

The response to the cyber revolution may need to be much more transformative than the various policies mentioned so far, or even than all of them combined. In the near future, societies may well need to adapt to a world in which robots will become the main working class and people will spend more of their time with their children and families, friends and neighbors, in community activities, and in spiritual and cultural pursuits. This transformation would require some combination of two major changes. The first would be that people will derive a large part of their satisfaction from activities that cost less and hence require only a relatively modest income. Such a change, by the way, is much more environmentally friendly than the current drive to attain ever higher levels of consumption of material goods. The second change would be that the income generated by AI-driven technologies will be more evenly distributed through the introduction of progressive value-added tax or carbon tax, or both, and a very small levy on all short-term financial transactions.

The most important service that the Cyber Age Commission could provide, through public hearings, would be to help launch and nurture a nationwide public dialogue about what course the nation's people favor, or can come to favor. If those who hold that the greatest challenges from AI are in the economic and social realm are correct, many hearts and minds will have to be changed before the nation can adopt the policy measures and cultural changes that will be needed to negotiate the coming transformation into an AI-rich world.

Amitai Etzioni is University Professor and professor of international affairs at George Washington University in Washington, DC. Oren Etzioni is chief executive officer of the Allen Institute for Artificial Institute and professor of computer science at the University of Washington.