STRATEGIC STUDIES QUARTERLY

Highlighting Artificial Intelligence: An Interview with Paul Scharre

Director, Technology and National Security Program Center for a New American Security

Conducted 26 September 2017

SSQ: What is the best way to prepare for an artificial intelligence future?

Mr. Scharre: People have been talking about AI for decades and there have been cycles of excitement, hype, and disappointment. We are in a period right now of intense excitement and progress. In just the last five years we've seen several things emerge. The first is big data that can be used to train learning machines. That combines with more powerful computer processing capabilities that can be used for parallel computations for deep neural networks. And finally there have been advances in the algorithms. All of this has come together to enable machine learning, often using deep neural networks, that can make machines very effective at solving a variety of problems. We are seeing this technology being applied to a whole range of industries including finance and transportation, and there are many national security applications as well. The best way to think about AI is not as a discrete kind of technology, like you might think of hypersonics, but something that is more like a basic enabling technology like electricity. Kevin Kelly, editor of *Wired* magazine, has suggested that just as electricity empowered and enlivened all sorts of objects, AI will similarly cognitize objects making them more intelligent and useful. Now, there are limitations to AI today. It is very narrow—domain specific—and does not have the kind of general-purpose reasoning capability that humans do or the kinds of scary AI one sees in science fiction. But even still, AI today is a very powerful technology. Many people compare it to a new industrial revolution in its capacity to change things. It is poised to change not only the way we think about productivity but also elements of national power. Just as past industrial revolutions transferred power to the more industrialized nations, AI will do something similar. But those elements of national power and advantage may look different. What is it that gives an actor a competitive advantage, whether a corporation or government? Is it better data, better algorithms, human capital, technology, or the right ideas for implementing them? The first-order questions we should be thinking about are: what is this technology, what is the essence of what is occurring, and how do we think about strategic advantage? What will position the United States for strategic advantage, and how do we maintain it? With all the disruption AI brings there is great opportunity and also a lot of risk, particularly for a nation like the United States that is heavily invested in the current way of doing things. We spend quite a bit of money each year on defense and national security programs and so far we don't rely on AI to any great extent. So how do we need to shift what we are doing as a result of AI?

SSQ: A recent study predicted that by 2025 over 70 billion objects would be network enabled. Should we be rethinking the internet of things?

Mr. Scharre: The trends in the internet of things are new and are happening out of anyone's control. The proliferation of the internet of things is going to force us to rethink elements of the internet and connectivity from a standpoint of cybersecurity and personal security. We need to better prepare for the world that is coming. William Gibson, the science fiction writer who coined the term cyberspace, has said: "Cyberspace, not so long ago, was a specific elsewhere, one we visited periodically, peering into it from the familiar physical world. Now cyberspace has everted. Turned itself inside out. Colonized the physical." In many ways, cyberspace is not a place but rather a layer on top of our existing reality. So through our various connected devices, whether in our pockets or our cars, we are now able to connect with others around the world. The trend is toward more internet connections and more devices whether in our homes or as wearable devices. There are a number of challenges that come with this trend. The baseline challenge for these devices is that the cybersecurity for these devices is incredibly poor. We tend to rush these devices to market even if they contain many vulnerabilities. Then we try to close the vulnerabilities later after deploying them and think about security as the last step. Many of these devices are very insecure, and the effect is not only that people can hack the devices in your home to spy on you but also that these devices can be leveraged as part of bot nets for things such as DDOS [distributed denial of service] attacks. The Mirai bot net in 2016 is one example. So this is a major problem. Societies need to reevaluate their views on cybersecurity as a whole and in particular the risks to their personal security that come with these devices. One of my favorite hacks came from an episode of the TV show "South Park." The scenario used a character talking to "Alexa" [the virtual assistant AI from Amazon] during the show commanding Alexa to do things. Now if you had one of these devices in your home, it would respond to the television program rather than you. Again, the risk comes from someone being able to reach into your home via the network or some other method. So the internet of things is an interesting challenge. But most people who are buying these devices do not know how secure or insecure these objects are since there is no way for a consumer to know this. So these limitations create a big challenge.

SSQ: Are the risks of AI overblown, or do we have reason to be concerned?

Mr. Scharre: It depends on the kind of risks we are talking about. With any type of new technology there is going to be risk associated with implementation. Because we don't always understand the capabilities of the technology, we miss some of the risks involved and many of the unresolved safety concerns. For instance, consider electricity. It's not going to rise up and kill us all, but if one is careless, it can be dangerous and life threatening. We have learned the safety protocols of electricity, such as grounding and other precautions. Now we need to do the same with AI. We also need to think about people intentionally using AI for malicious purposes—something that is inevitable. State and non-state actors are going to use AI for nefarious ends and we must be prepared for this. Given the safety risks and vulnerabilities, we also need to be worried about AI systems that might be exploited or manipulated in some way. Current generation AI systems have safety problems that are not yet solved. These are also very serious concerns some experts have raised not about today's systems but more about the long-term implications. If AI systems become more intelligent, particularly if they develop in the direction of a general-purpose learning ability—which doesn't exist today—then this would raise significant long-term safety questions.

SSQ: As artificial intelligence becomes more ubiquitous and more powerful, should the United States attempt to control AI by enhancing human intellect through gene manipulation?

Mr. Scharre: This is a great question. Let me reframe the issue just a bit. During the first industrial revolution, we were able to create machines that were much stronger than human beings to perform various kinds of tasks. We are now creating machines that are smarter than humans—if the task is narrow enough and we have enough data to support it. So it seems as if for many applications we will be able to leverage machines in very specific ways. In many cases, even if machines are not qualitatively as smart as humans in making the best quality decisions, machines are faster than humans and can be employed cheaply and at scale, which is a great advantage. We have seen this kind of application in stock trading where the speed advantage emerges. We have seen this in Twitter bots where the advantage of scale would not be possible if you were trying to use a million people to replicate content. At the same time, the best general-purpose learning system on the planet is the human brain in terms of quality, robustness, perception, flexibility, and responding to novelty. This is unlikely to change any time soon. While it's possible there may be something in AI that changes this, it does not appear likely in the near future. Given these limitations, we should be thinking about the best way to blend intelligence—human cognition and machine learning working together. One challenge is going to be how we avoid making it more difficult for humans to stay engaged as the speed of action increases due to automation. It doesn't matter that humans make better qualitative decisions for stock trading and are more cognizant of manipulation; you simply cannot compete at the speed of automated stock-trading algorithms. There is potential for using AI and automation in warfare or national security applications, particularly in domains that are native to machines, such as the electromagnetic spectrum or cyberspace. In this type of world, how do humans cope with an environment where we may be approaching a battle for singularity, where the pace of battle becomes so fast that humans are not able to comprehend what is happening and react to events fast enough? We have some narrow settings in the military today where this already is the case, for instance with missile defense systems operating in automatic mode. The domain in which humans can no longer react fast enough is expanding over time. There are certainly risks when automating. Machines today are very brittle and do not have the common sense we expect from humans or the ability to understand context. This limits the machine's ability to recognize errors and stop if it malfunctions. So we might want to also think about how to increase human performance directly. Today there are many ways to enhance human performance through medicine—for instance, drugs such as Modafinil and Adderall to increase stamina, alertness, or concentration. The military is conducting some interesting studies in this area by using some of these drugs—mostly in aviation—but adoption is extremely slow in the military overall. This is the case even though the new drugs are better than the ones the military is currently using. For example, we give dextroamphetamine to pilots and caffeine of course to all sorts of troops in an unregulated fashion. But studies have shown that Modafinil is more effective at enhancing cognitive performance with fewer side effects than dextroamphetamine or caffeine. We should be looking at ways to enhance human performance, including genetics that we may see happening in the coming decades. Now anything that alters humans directly raises a host of serious legal, ethical, and social issues, and I don't want to dismiss them. We need to be careful to ensure that we're not exposing our troops to potentially harmful treatments. But we also don't want to miss out on an opportunity to enhance their performance and potentially save lives. The way to deal with this challenge is to confront these issues directly and work through them. There are things we could be doing with technologies that are well understood, effective, and reasonably safe that we are not doing because so far we have not been willing to grapple with these question in the military.

SSQ: Some people claim the US will "never" use autonomous lethal military systems. Is this realistic?

Mr. Scharre: The Pentagon is taking a cautious, hedging approach to autonomous weapons. The official policy, which I was involved in while working for DOD, approves certain things that we are already doing, such as autonomous missile defense systems. The policy then also creates a new process for approving new technology if people want to use autonomy in a novel way that's never been done before in weapon systems. So now there is a process for stakeholders to come together and evaluate ideas before adopting new uses of autonomy in weapons. When former Deputy Secretary of Defense Bob Work was at the Pentagon, he spoke about this and in essence stated we are not planning to use lethal autonomous weapons but if others do, we might have to. Air Force Gen Paul Selva has spoken on this a number of times and has said he feels it is essential to keep humans responsible for using lethal force. This raises a slightly different question: how do we think about accountability and responsibility? One of the challenges here is making a clear, bright line. Look at the example of self-driving cars. In theory, there is a clear difference between a car driven by a human and a car driven by a machine autonomously. But what we see in practice is creeping autonomy in a wide range of functions, such as intelligent cruise control, automatic collision avoidance, automatic lane keeping, and automatic parking. We are seeing a slow shift in various functions to the machine. The human is still sort of responsible for driving, but what we mean by "driving" begins to change over time and it starts to look a lot more like what we see in commercial airlines. The plane can basically fly itself and the pilot is there in case of an emergency and in some cases to be a scapegoat if something goes wrong. As automation continues to creep forward, how does this change the role of the human, and how do we ensure the human is ultimately responsible for what happens on the battlefield?

SSQ: What is the most futuristic AI technology we will see in the next 20 years? And the next 100 years?

Mr. Scharre: What we are likely to see in the next 20 years, given current advances in AI, is implementation of various narrow AI capabilities. I suspect we are likely to be surprised by how capable some of these applications might be but also how brittle they are. Consider DeepMind's program AlphaGo that learned to play the game of Go. Many people thought this application would take much longer to perfect than it actually did. Additionally, the system defeated its human Go adversary quite handily. So one of the effects we see is, often AI capabilities seem very distant but then, seemingly overnight, AI moves from not very good to much better than the best human player. Another aspect I think we are likely to see in the next two decades is the surprise factor—how machines can learn in novel ways. Sometimes these surprises are good, sometimes not so good. My favorite example is a bot that was learning to play the game of Tetris learned to pause the game right before the last brick fell so it would never lose. That was allowed according to its programming, but probably not what the designers meant it to do. Brittleness is another important attribute of AI and something we will have to grapple with as these technologies are implemented in various applications. AlphaGo learned to play on a standard 19-by-19-inch Go board and is better than any human at playing that game, but its intelligence is very narrow. AlphaGo cannot transfer its experience in playing to give it a leg up on learning how to play chess or checkers. It can't even play Go very well on a differently sized board. This is very different from a human player who can take concepts from one game and apply them somewhere else. So the systems will remain very brittle—being very powerful, but in an instant becoming very dumb.

In the longer term over the next century, I think it is very likely we will have systems that can overcome some of the weaknesses of AI systems today. One of these areas is the ability to transfer learning from one task to another. AI will be able to learn over multiple domains. The future will move from today's narrow learning systems to wider, general-purpose learning systems. Many will ask the question: when will AI reach human-level intelligence? But this is the wrong question. Why would we assume humans are the benchmark for intelligence? Why would we assume machines will evolve intelligence in the same way as humans? Humans today can still do things machines cannot do. But in the future we are more likely to see machines that have general-purpose abilities and manifest them in ways very different from today. One hundred years from now, I suspect people will continue to say, machines are very smart but they are not smart like people. This is only because we increasingly narrow down the things that make us uniquely human. We are likely to see very powerful general-purpose systems and that will create a range of tricky problems as we develop AI.

SSQ: Mr. Scharre, on behalf of Team *SSQ*, thank you for sharing your views on artificial intelligence with the *SSQ* audience and for peering into a future we hope will produce great promise for mankind. SSQ

For More Information

Scharre recommends the following links for those interested in learning more about artificial intelligence:

- https://www.wired.com/2014/10/future-of-artificial-intelligence/
- https://www.cnas.org/publications/reports/patriot-wars
- https://www.wired.com/2016/03/two-moves-alphago-lee-sedol-redefined-future/

- https://arxiv.org/pdf/1606.06565.pdf
- http://nautil.us/issue/40/learning/is-artificial-intelligence-permanently-inscrutable
- http://nautil.us/issue/27/dark-matter/artificial-intelligence-is-already-weirdly-inhuman
- http://www.evolvingai.org/fooling

Disclaimer

The views and opinions expressed or implied in SSQ are those of the authors and are not officially sanctioned by any agency or department of the US government. We encourage you to send comments to: strategicstudiesquarterly@us.af.mil