3 Takeaways Podcast Transcript

Lynn Thoman
(https://www.3takeaways.com/)

Ep. 151: War In The Age of AI. A Chilling, Mind-Blowing Talk With A Former Pentagon Defense Expert

[music]

INTRO male voice: Welcome to the 3 Takeaways podcast, which features short memorable conversations with the world's best thinkers, business leaders, writers, politicians, scientists, and other newsmakers. Each episode ends with the three-key takeaways that person has learned over their lives and their careers. And now your host and board member of schools at Harvard, Princeton and Columbia, Lynn Thoman.

Lynn Thoman: Hi everyone, it's Lynn Thoman. Welcome to another 3 Takeaways episode. Nuclear technology and nuclear weapons have become much more powerful. The 1986 Chernobyl disaster, which was the result of a flawed nuclear reactor design and human error released into the air 400 times the radiation emitted by the US nuclear bomb dropped on Hiroshima in 1945. If artificial intelligence controls today's powerful nuclear and other weapons, the impact could be orders of magnitude greater. Today, I'm excited to be joined by Paul Scharre, to find out what war looks like in the age of AI. Paul previously worked in the Office of the Secretary of Defense, where he helped establish policies on unmanned and autonomous weapons systems. Before that, he completed multiple tours in Iraq and Afghanistan. He's a graduate of the Army's Airborne Ranger and Sniper schools, and he is currently Director of Studies at the Center for a New American Security. He is also the author of several books, including Four Battlegrounds: Power in the Age of Artificial Intelligence. Welcome Paul, and thanks so much for joining 3 Takeaways today.

Paul Scharre: Thank you. Thanks for having me here today.

LT: It is my pleasure. Paul, when did you first realize that AI and robots would transform war?

PS: There was a discreet moment when I remember the light bulb coming on for me. I was in Iraq in 2008, so during the surge period when the US surge troops to the country to try to turn things around, and we were driving down the road and came across a roadside bomb or an improvised explosive device. And we saw it first, which is the preferred way of finding them rather than just running into it. And so, we called up the bomb disposal techs, and I was expecting to see this person come out in the big suit that they had, the big bomb suit, and I'd just been in-country maybe like a month, I thought it was, it'd be interesting to see. And instead out comes this little robot. And the light bulb went on my head and I was like, "Oh, that makes tons of sense, to have the robot diffuse the bomb - you don't want to be up there, sniping the wires, having a person do that.

PS: And then the more I started thinking about it, I was like, there's a lot of things that we're doing that are dangerous. There's a lot of things in warfare, and certainly we see this playing out in the war in Ukraine right now where people are in harm's way, people are killed and injured, and maybe robots could help create more standoff from these threats and protect US service members ' lives. And so, when I left the military and went to the Pentagon, that's one of the issues that I worked at - working to ensure that the US military could adopt robotics and later AI, as we've seen the

technology evolve, to help protect US service members.

LT: When an artificial intelligence fighter pilot beat an experienced human pilot 15 to 0 in the Defense Department's, DARPA [Defense Advanced Research Project Agency's] AlphaDogfight competition, it didn't just fight better than the human, it fought differently. Tell us about how AI fights differently.

PS: Yeah, this was mind blowing. So, in this competition, DARPA, the Defense Department's Department of Mad Scientists, they created an AI system and AI algorithm to go head-to-head against the human in a simulator in dog fighting. And as you said, AI totally crushed the person, but what was wild was that it actually used different tactics than how people fight. And one of the things that it did was it would make these split-second head-to-head gunshots. So, the aircraft are racing at each other, they're doing hundreds of miles an hour, and there's a brief second, like a split second where you could get a shot off and take out the other aircraft. This is basically impossible for humans to do this, but the AI could do this because machines can operate at superhuman levels of precision and speed and accuracy. And so, we can see that in this case, it opens up this new potential tactic. Even more interesting that AI learned to do this entirely on its own, that the AI systems are not just better - they actually fight differently than people, and that opens up a whole space of new possible tactics and strategies in warfare.

LT: Can you give some examples about how AI fights differently?

PS: Yeah. So, for example, in team fights, basically we have multiple units fighting against other ones. AI systems can operate with better teamwork than humans can do, so that AI agents can do things like time their attacks - so, they're going to land on the enemy unit at the exact same time with just the right amount of resources to take out that enemy unit without overkill and wasting energy or undershooting. We see in some settings that the AI systems demonstrate better situational awareness. They can just take in more information across the whole battle space, or the game space, better than humans can. This is true in computer games, and it's also true in games like chess. And then we'll also see that the AI systems can engage in some cases in more finely calibrated risk taking. This comes out quite a bit in poker, as you might imagine, we've got betting that the AI agents have to do, but also on other games where AI systems can be sometimes very aggressive, but in other times can pull back if they need to, and can exhibit these like huge swings in their level of risk taking and aggressiveness based on what's called for in the moment in ways that are hard for some of the best expert players to do in these games.

LT: How does AI learn its tactics?

PS: So, we've seen this huge paradigm shift in the last decade as part of the deep learning revolution towards machine learning systems. A lot of the breakthroughs recently come from machine learning where an AI system is trained on data. And so, for example, an early version of AlphaGo, the AI agent that achieved superhuman performance at the Chinese strategy game Go, it was trained on millions of human moves. So, they programmed in a database showing how humans moved in different situations, and they trained this large neural network, this big set of connections between these artificial neurons and this this big network. And it inputs this data, makes adjustments in the network to learn from the data and then outputs responses, which were possible moves to make. And that achieved just close to human level performance. And then to get beyond that, they had the AI playing against itself and that's a tactic that we see in a lot of areas. That's what

DARPA did to reach superhuman performance at
AlphaDogfight. They actually had a competitive league of over a hundred different AI agents, dogfighting against each other to come up with new tactics and new ways to beat humans at dogfighting. We're starting out training these AI systems on what humans can do, and then in many cases, pretty quickly they're going beyond. And then we have to turn around and we're learning from what the AI can do.

LT: Before the Industrial Revolution, the population and men under arms was the measure of military power. And the industrial revolution changed military power from men under arms to number of war machines such as tanks and planes. What happened, for example, to the UK and to Russia and their relative power?

PS: In the start of the 1800's, Russia was ahead of Europe, in economic and by extension military power because of their larger size. But Great Britain, and then somewhat later, Germany industrialized faster. And so, they [Great Britain and Germany] raced ahead in economic power and also military power. As we saw later on in World Wars I and World War II factories were transformed to churning out tanks and airplanes in World War II and countries turned their economic might to military power. And so, it's a cautionary tale about the importance of adopting technology quickly and finding ways to use it in your society for economic growth and for the military, for military advantage.

LT: How do you think AI will change the relative power of nations and war?

PS: Well, I think we'll see that it'll be largely based on who's able to adopt AI faster and make effective use of it in their society and in their military.

LT: Vladimir Putin has said, whoever becomes the leader in this sphere will become the ruler of the world. Why is being the leader in AI so important?

PS: I would compare it to getting an early lead on industrialization in the 19th century. And so, we can see how for, Great Britain and Germany that allowed them to race ahead in economic and military power. And I think that's likely to be the case for AI. We can already see that AI has transformative potential in society, and AI technology hasn't capped out. In fact, it's continuing to accelerate in terms of progress. And there's real advantages for countries that are going to be able to find ways to adopt this technology, increase their productivity, their societal welfare, their health, their economic and military power, but also to shape how these tools are used globally.

LT: Can you talk briefly about where the US, China, and Europe now stand on AI?

PS: When you look across a whole wide range of metrics looking at AI research and patents and adoption, one of the common sorts of overarching kind of big takeaways is that the US is in a leadership position in artificial intelligence today, but China is catching up and on track to overtake the US in some key areas in the next few years. China has said that their goal is to be the global leader in AI by 2030. And I take them pretty seriously at that goal. And so, I do think that at the end of the day, both China and the US are major powerhouses in AI and they both have a lot of opportunity here. And it's going to be really a question of how they're able to play the cards that they have, whether they're able to double down on the advantages that each of them has, or they miss some of those opportunities.

LT: If there are two countries or two, doesn't need to be countries, two actors, and one of them has AI controlled systems, weapons systems, self-defense systems, and the other does not, isn't there a big risk for the one that does not have the AI controlled systems that they could be wiped out?

PS: There's a huge risk, and that's part of this dilemma that we find ourselves in geopolitically because these AI systems that have vulnerabilities, they can do strange and surprising things. They can break in unexpected ways. They can be very opaque, but they're also going to have advantages. And so, there's this dilemma where using it comes with risks, but not using it also comes with risks. And I think one of the solutions here is to find ways to get countries to cooperate, to figure out, okay, even as countries are competing in AI, are there some things that we could take off the table and manage some of the most extreme risks in military AI?

LT: I hope you're right.

PS: One of the saving graces that we've had over the last century with nuclear weapons is it's really hard to build a nuclear weapon. Even if say a terrorist could get their hands on a nuclear weapon, thankfully it's just a physical thing it's not a piece of software where somebody could then copy it and post on the internet. AI systems are, and that's going to make controlling the proliferation pretty difficult.

LT: What are the dangers of AI that you see? Are they the ones we hear about in science fiction.

PS: Well, no, I mean science fiction has told us the story of AI systems getting smarter and then turning on us. And I'm more concerned about what people might be doing with AI systems. Now there are problems controlling AI systems today. They're not reliable, they're not robust. I think accident risk is a real problem, but it doesn't necessarily mean some AI is going to wake up and decide to exterminate humanity, like you see in science fiction. For one, AI systems are already being born with a gun in their hand - there are already weaponized AI systems and robotics autonomous systems. And so, this idea that somehow the AI systems seize control of the military, well they're already in the military. But, it doesn't necessarily take some AI system becoming self-aware to 'cause harm. You started by talking about Chernobyl, that's a scary kind, like an AI version of Chernobyl, an accident with powerful AI systems. If people aren't paying enough attention to safety, that's I think a big concern one should pay attention to in addition to putting in protections against deliberate misuse by people.

LT: Paul, let me ask you a question that you have posed yourself. Are we careening toward a world of AI systems that are powerful but insecure, unreliable and dangerous?

PS: I mean, the answer is yes, we are and that should really worry us. That's a problem. And so, I do think that when you look at the most cutting-edge systems, we need to be putting some protections in place because that is the trajectory that we're on. These systems are getting much more powerful very, very quickly. And then they rapidly proliferate. Just a few months ago, Meta formerly Facebook, had a very powerful language model released online. It leaked - they were sharing it with academic researchers and then someone put it up online. And once these things are released, there's no good way to get them back because they're trained AI models. It's basically a piece of software. There's not a good way to control that. And so, we need to have better, tighter protections on training these systems. So, people have started talking about a licensing regime for

training them and I actually think that makes a lot of sense, and also looking at proliferation to control the most powerful systems so that they don't spread out to the hands of people who might want to 'cause harm.

LT: Before I ask for the 3 Takeaways you'd like to leave the audience with today, is there anything else you'd like to mention? What should I have asked you that I did not?

PS: I guess one thing that that maybe I haven't mentioned yet, but I think is really important is the pace of progress right now is remarkable. So, if you're someone who hasn't been paying attention to AI, then all of a sudden AI is in the news, like what's going on? Is this real or is it hype? I think it's real. I've been working on these issues for a very long time and I'm pretty bullish on AI progress. I do think we're going to see more capable systems and I will say that I have been very surprised by the pace of progress in the last year. Things that I thought we might see 10 years from now are now happening. And that is what's driving a lot of serious AI scientists to raise the alarm to say, whoa, hold on. We need to slow down. Let's take a deep breath here because the systems are insecure, they're not safe and we need to be thoughtful going forward. We shouldn't be rushing blindly into a dangerous situation. And actually, I think some kind of government regulation is probably called for here.

LT: Although there will always be people that will avoid regulation.

PS: That's right.

LT: What are the 3 takeaways you'd like to leave the audience with today?

PS: One, I think the progress we're seeing in AI is real, the systems are powerful. The second is they have a lot of vulnerabilities. They can fail in strange, surprising ways, and it's hard to tease those vulnerabilities out ahead of time and that's a real risk to give us pause. And three, almost I think the most important thing to me, is even though these systems have some elements of intelligence, they don't think like people. So instead of thinking of intelligence, like a staircase where you've got bacteria and ants and mice and dogs and chimpanzees and humans and AI's moving up the staircase, in fact, what we see is that intelligence looks more like this vast space of different kinds of intelligence. And the AI systems that we're building, they are pretty capable, but they don't think like humans. And it makes it more challenging for us to interact with them because we tend to project that image of human intelligence onto them and then they do something that's weird and surprising. And so, we need to be cautious when we think about these systems to realize that it's less like artificial intelligence and more like an alien form of intelligence. And that should caution how we use these systems and how we employ them.

LT: And we cannot always understand how they come to their decisions or recommendations or actions because in many cases they could be processing trillions of parameters of data. It's a black box, is that right?

PS: That's right. It's a huge challenge. There are tools people are working on that make them more explainable, but right now they're in some ways very opaque and that's going to be a real hurdle when we think about how do we use them in a way that's safe.

LT: Thank you, Paul, this has been great.

PS: Thank you. Thanks for having me. Really enjoyed the discussion.

OUTRO male voice: If you enjoyed today's episode and would like to receive the show notes or get new fresh weekly episodes, be sure to sign up for our newsletter at https://www.3takeaways.com/ or follow us on Instagram, Twitter, LinkedIn and Facebook. Note that 3Takeaways.com is with the number 3, 3 is not spelled out. See you soon at 3Takeaways.com (https://www.3takeaways.com/)

This transcript was auto-generated. Please forgive any errors.