The Ethical Dilemma We face on AI and Autonomous Tech

**01:00** So, I wanted to start today with that video because I think it really captures the complexity surrounding technology today. Now I've looked at that video with a lot of colleagues and friends, and I find that people's reactions to that video follow along sort of three phases. So let me describe them and you can ask if your reactions followed along these phases.

Phase 1, Wow is that cool? Right? I mean is that cool, those guys are amazing! Then they start kicking it. That's kind of ,you know, that's kind of mean. Right, I don't know that they should be doing that. I don't think they should be allowed to kick them like that. And then, there's phase 3, a kind of walking up the hill and looking around. That big guy starts going up the hill. it's kind of, hmm, you know they're a little creepy.

I don't know if I'd feel happy running into them walking in the woods, you know, what would they do? I'm not sure we should be building those things. I don't think they should be allowed to build them. So the question that I want to explore with you today is **who decides should**.

**02:17** Now, I've spent a lifetime studying military operations as you heard I was even in the Defense Department for a while, and I can tell you, the answer to this question for military operations and Technology is very clear. Who decides should for military operations? It's the government. Its policy. Its Defense Department leadership. One of the things we expect our military to be able to do is comply with something called the laws of armed conflict. Now, the laws of armed conflicts say many important things, but one of the things that says is that the military should do everything in its power to target combatants and protect civilians in conflict. Now, I can tell you the military tries very very hard to do this, studied operations for decades. They are the perfect? Absolutely not! Do they make mistakes? Yes! But they try very hard and they've gotten better and better and better at it, as technology has improved. One of the things that has improved our precision weapons. Precision weapons have made a big difference here in their ability to focus on combatants and protect civilians. Now more recently, we've put precision weapons on something called remotely piloted vehicles: drones. Now, I know drones are controversial, and I confess I'm not entirely sure I understand why they're so controversial and I wonder if people have the perception or misperception that they're actually autonomous. They are remotely piloted. The pilot is not in the aircraft, but there is a pilot on the ground in a distant location, but they have positive control of that aircraft from takeoff to landing, and certainly for any employment of a weapon. In fact, today, it is DoD policy that any employment of lethal capability have a human in the loop. And when it comes to autonomous systems, there's a special directive governing autonomous systems that specifically says lethal autonomous capability is not allowed on the battlefield today. And one of the reasons for that is a concern that lethal autonomous systems or technology is not able to discriminate between combatants and civilians, and so therefore, it would not be in compliance with the laws of armed conflict. Now let me tell you. There are a lot of people who think that this policy is overly constraining. I've heard a lot about it. Right. It's holding us back. It's putting men and women in uniform in harm's way needlessly. Our adversaries are going to get advanced ahead of us. So, it's, you know, there's a lot of pushback on this directive. I've thought a lot about this directive. I personally think it's just right. And I want to ask you to think along with me using the example of our friend spellchecker. Okay, I don't know about you, but I often hit Send on a text just after I noticed, that my helpful smartphone changed the spelling and the meaning of what I wanted to send. I hope that happens to you and I'm not the only one, right? Okay. So, let's ask ourselves, if technology today can't understand the intent of a few simple lines of text, how could we count on it to discriminate between a combatant and a civilian in environments so complex such as warfare? So I personally think the policy has it right for now. But the technology is evolving. And I also think it will probably evolve to a point where it can discriminate, and I am confident that the policy will evolve along with the technology. And so I think this is a nice example of the relationship between technology and policy: one evolves, the other evolves. They stay together. And that happens because technology that I'm talking about here is under the purview of the government.

Now there was a time when advanced technology was developed under the purview of the government in most cases. That is not true today. Today, advanced technology is developed in the commercial world, far more than in government. Consider for a moment the fact that if you add up the research and development budgets of the top 5 defense contractors, talking the big ones, Lockheed Martin and Boeing etc.. The top 5, add them all up, that comes to less than half of Microsoft's research and development budget in the year, less than half. That doesn't even begin to consider the R&D budgets of Apple and Google and so many others, right? So, advanced technology is no longer the purview of the government. It's the purview of the commercial world. And policy makers are not governing advanced technology in the commercial world. I think that technology is amazing today and it is going so fast that policymakers have fallen behind. And I 1personally am very dubious that they can catch up. So where does that leave us? I think it leaves us with a bit of a toxic brew, where we have advanced powerful technologies that are available to anyone who wants to buy them with few if any constraints over their development and accessibility.

**07:44** Now, technology is making our world better. Ok. We do not want to regulate technology. We want technology to continue going forward. Take, for example, this very simple but lovely application of technology today. A non-profit foundation has developed this app. It pairs people with CPR skills with people in need having acute cardiac arrest. This simple application of technology is saving lives today. Awesome! Right? That is awesome! We don't want to do anything to stop that. But there are other applications of technology that we didn't anticipate that aren't making our world a better place and that we aren't able to govern.

**08:28** So I want to continue to explore this a little bit and I want to work through an example with you. And I want to work through the example of the internet. So, I remember when the internet was first conceived of and introduced, right, was developed by DARPA, the Defense Department, but rapidly went out into the public, and the promise was this amazing notion. I still remember how I felt about it at the time that we would be able to instantly communicate with people all over the world, and any information you wanted would be at your fingertips, coming right up on your computer. Really? Really? Wow, that's amazing! Look! What's happened? That promise has been realized! It's amazing. Right? I think the promise of the Internet has really changed our lives forever. There's no going back. I don't know about you, but I'm really cranky when my Wi-Fi is down for just a little bit. Right? We expect it now. It's part of life. But when the internet was first introduced, I don't think we really envisioned this future, world of cyber, or cyber warfare. Or the fact that the Defense Department would one day need a cyber command whose whole mission is to operate in cyberspace and defend it and defend our networks. So, that's an unanticipated consequence of the internet. The internet has introduced a whole set of vulnerabilities. And those vulnerabilities are also impacting our lives.

**09:56** Take for example. this, um…this group of people that worked at Sony. So, you're at Sony, you come to work, you turn on your computer, you expect your calendar and your email to come up, and instead, this is what you see. That happened to every employee in Sony in November two years ago. That was happening right before Sony was due to release a parody, a movie about the North Korean leader Kim Jong-un. That was terrible, okay, terrible. It was their biggest concerns at Sony before this happened, was how big a flop it was going to be. Okay. But North Korea didn't think it was so funny, they didn't understand that it might be a flop. They didn't care. And this is what happened to Sony. They were hacked. Information was stolen embarrassing information and they got this threat and multiple threats for the next several weeks. And it had a very devastating effect on Sony and the people who work there. Or perhaps consider that you're one of the twenty one point nine million people, I bet some of you here are, I certainly a.m., who had their personal identifying information stolen out of the Office of Personnel management's database, okay, stolen by a hacker .What's that hacker going to do with my data? I don't know. You know, it's unnerving. It's really unnerving. And then, take a much more recent example where we have a nation hacking into our political process, the very foundation of our government, a very big deal indeed. So, that's where we've been. That's the story of the internet, great promise. All achieved even more so, but also simultaneously introduced some vulnerabilities that we hadn't envisioned, hadn't counted on, and hadn't prepared for.

**11:53** All right, so let's stop and ask ourselves going forward. What's next? Where are we now? And what do we see going forward where maybe we can anticipate a little better, do a little bit better job? So I want to talk a little bit about the Internet of Things. Now, the Internet of Things is already here in some ways. Right? You can already turn the lights on in your home before you get there. So it's nice and friendly when you arrive, turn the heat up, so it's warm and toasty, get your coffee done in the morning, from just to click on your smartphone. Wonderful things! All aimed at making our lives easier, making our lives better! Internet of Things manufacturers now are preparing on to include something like this little guy into our homes. This little guy is going to be your friendly household servant. He's going to help wake you up in the morning. He's going to get the kids ready for school. He's going to get breakfast ready. He's going to have your favorite news program on your radio. Who wouldn't want him? I'm a guy I want him. Right. My life is hard. I really would love to have some help, right. I think we all feel the press of time today, and we, this is a very appealing concept. It would make our lives better, or would it?

**13:09** Who decides the cyber protection in our friend, the robot? Who decides? f you're sitting out there thinking that there's some government policy want somewhere that's going to make this guy safe, I'm afraid that you're wrong. There is no policy governing it. There is no sense of what's the appropriate cyber protection to have in this robot before it goes to market. There's nobody deciding when it's ready to go to market. This robot is going to come to market and we're going to decide to buy it. And so, who decides? We do! And the company that builds it. Okay, so I am concerned about this. And I wish I had an answer, because I really don't think that policy makers are going to be able to keep up with this explosion of technology. And you know, this has been the study of academics for a long time. Morality affects legality of technical new technologies and the introduction of technology. And I want to suggest to you today that the time for study is over. These things are here today. This technology is exploding today.

**14:22** Now, you may think that that robot is way far off, and we have a lot of time to get it right. But the manufacturers are predicting that they'll be ready for market in less than two years. Now, you may be skeptical. I can tell you, I'm skeptical. But I'm a terrible predictor of these things. Two years ago, I said no way are we going to have autonomous cars on the road in any near time, right. There's going to take a long time. Well, if you're driving on the Beltway today, there's a very good chance that there's an autonomous vehicle driving somewhere in your midst, right now today. So, you know, autonomous cars is another really interesting thing, right, because autonomous cars are also going to have to make some difficult choices. What if a car, autonomous car gets into a situation where it has to choose whether to protect the occupants of the car or the people outside the car. Does the car drive into the pedestrians or the brick wall? It seems to me that's a little bit like trying to choose between a combatant and a civilian. Yet, the cars are on the road now. And what about their hack ability? If you don't think you can hack these cars, I encourage you to Google Jeep hack. Many of you have, I can hear that. You know, it's and that's not even an autonomous car, right. It's just driving down the road, its radio goes crazy, air conditioning goes crazy. But its transmission is cut, and it’s caused to a stop in the midst of traffic, a very dangerous situation.

**15:58** So, these things are with us now, and I think there's a sense of urgency we have to start really grappling with this problem. And I don't think policymakers are going to solve it. So, I think that maybe it calls for some new form of corporate social responsibility. I think companies need to take this on. It's a little different version of it, but it you could see how it might apply where companies think about the vulnerabilities of their technology as well as the promise and the coolness of it. And perhaps we could get companies to adopt something we do in the Defense Department called red teaming. When the Defense Department comes up with new plans or new technologies that we want to introduce, we bring people in who are very imaginative and expert on what adversaries might do, and we say okay, have at it. give us your all, use your imaginations, take our plans, tell us how this technology can be turned against us. And they do that, and that gives us a chance to fix our plans, and to change the way we use the technology. So we have a much better chance of success when we need to use these things. Perhaps we could Red Team, new technologies before they go on the street. We could get somebody to use their imagination to ask: how that robot might end up being used against us instead of helpfully helping us, make our lives better. Now, in order for a new form of corporate social responsibility to work, there has to be a reward system, right? so they have to invest in more protection. It's going to take a more time, cost a little more money. So the reward is that you preferentially buy the products that come with security that means consumers need to be educated, and they need to ask the hard questions before they just buy the device. And so, that means the answer to who decides should, is us. I really think all of us. We have to decide should. Corporations need to decide should, and consumers need to decide should.

Now listen, I don't want to be debbie downer. I love technology, okay. I work at the Applied Physics lab. I am surrounded by cool technology every day, and I love to buy it, okay. But I want a world where the technology is taking us to that great place that makes our lives better. And I want to avoid a world where we're surprised by unintended events and vulnerabilities that we just didn't take a moment to anticipate. But I'm confident that if we all pull together and we start asking the hard questions, we can drive our world in that good direction, not just for us today, but for generations to come. Thank you!