Audio file

119039-IHuman.mp3

Transcript

Intelligence is the ability to understand. We passed on what we know to machines.

The rise of artificial intelligence is happening fast, but some fear the new technology might have more problems than anticipated.

We will not control it. Artificial intelligence is simply non-biological intelligence. And intelligence itself is simply the ability to accomplish goals. I'm convinced that AI will ultimately be either the best thing ever to happen to humanity or the worst thing ever to happen. We can use it to solve all of today's and tomorrow's greatest problems. Cure diseases, deal with climate change, lift everybody out of poverty. But we could use exactly the same technology to create a brutal global dictatorship with unprecedented surveillance and inequality and suffering. That's why this is the most important conversation of our time.

Artificial intelligence is everywhere because we now have thinking machines If you go on social media or online, there's an artificial intelligence engine that decides what to recommend. If you go on Facebook and you're just scrolling through your friend's posts, there's an artificial intelligence engine that's picking which one to show you first and which one to bury. If you try to get insurance, there's an AI engine trying to figure out how risky you are. And if you apply for a job, It's quite possible that an AI engine looks at the resume.

We are made of data. Every one of us is made of data. In terms of how we behave, how we talk, how we love, what we do every day. So computer scientists are developing deep learning algorithms that can learn to identify, classify, and predict patterns within massive amounts of data. We are facing a form of precision surveillance. You could call it algorithmic surveillance, and it means that you cannot go unrecognized. You are always under the watch of algorithms.

Almost all the AI development on the planet today is done by a handful of big technology companies or by a few large governments. If we look at what AI is mostly being developed for, I would say it's killing, spying, and brainwashing. So, I mean, we have military AI, we have a whole surveillance apparatus being built using AI by major governments, and we have an advertising industry which is oriented toward recognizing what ads to try to sell to someone.

We humans have come to a fork in the road now. The AI we have today is very narrow. The holy grail of AI research ever since the beginning is to make AI that can do everything better than us. We've basically built a god. It's going to revolutionize life as we know it. It's incredibly important to take a step back and think carefully about this. What sort of society do we want?

So we're in this historic transformation. Like we're raising this new creature. You have a new offspring of sorts. But just like actual offspring, you don't get to control everything it's going to do.

We are living at this privileged moment where, for the first time, we will see, probably, that AI is really going to out-compete humans in many, many, if not all, important humans. Everything is going to change. A new form of life is emerging. When I was a boy, I thought, how can I maximize my impact? And then it was clear that I have to build something that learns to become smarter than myself, such that I can retire and the smarter thing can further self-improve and solve all the problems that I cannot solve. multiplying that tiny little bit of creativity that I have into infinity. And that's what has been driving me since then. How am I trying to build a general-purpose artificial intelligence? If you want to be intelligent, you have to recognize speech and video and... handwriting and faces and all kinds of things. And there we have made a lot of progress. The LSTM neural networks, which we developed in our labs in Munich and in Switzerland. It's now used for speech recognition and translation and video recognition. They are now in everybody's smartphone. Almost 1 billion iPhones and then over 2 billion Android phones. So we are generating all kinds of useful byproducts on the way to the general goal. The main goal is an artificial general intelligence, an AGI, that can learn to improve the learning algorithm itself. So it basically can learn to improve the way it learns. And it can also recursively improve the way it learns the way it learns without any limitations except for the basic fundamental limitations of computability. One of my favorite robots is this one here. We use this robot for our studies of artificial curiosity, where we are trying to teach this robot to teach itself. What is a baby doing? A baby is curiously exploring its world. That's how it learns how gravity works, and how certain things topple, and so on. And as it learns to ask questions about the world, and as it learns to answer these questions, it becomes a more and more general problem solver. And so our artificial systems are also learning to ask all kinds of questions, not just slavishly try to answer the questions given to them by humans. You have to give AI the freedom to invent its own tasks. If you don't do that, it's not going to become very smart. On the other hand, it's really hard to predict what they are going to do.

I feel that technology is a force of nature. I feel like there is a lot of similarity between technology and biological evolution. Playing God. Scientists have been accused of playing God for a while. But there is a real sense in which we are creating something. very different from anything we've created so far. I was interested in the concept of AI

from a relatively early age. At some point, I got especially interested in machine learning. What is experience? What is learning? What is thinking? How does the brain work? These questions are philosophical, but it looks like we can come up with algorithms that both do useful things and help us answer these questions. It's almost like applied philosophy. Artificial General Intelligence, AGI, a computer system that can do any job or any task that a human does, but only better. Yeah, I mean, we definitely will be able to create completely autonomous beings with their own goals. And it will be very important, especially as these beings become much smarter than humans, it's going to be important to have these beings, that the goals of these beings be aligned with our goals. That's what we're trying to do at OpenAI. Be at the forefront of research and steer the research, steer the initial conditions, so to maximize the chance that the future will be good for humans. Now, AI is a great thing, because AI will solve all the problems that we have today. It will solve employment, it will solve disease, it will solve poverty, but it will also create new problems. I think that… The problem of fake news is going to be 1000, a million times worse. Cyber attacks will become much more extreme. We will have totally automated AI weapons. I think AI has the potential to create infinitely stable dictatorships. We're going to see dramatically more intelligent systems in 10 or 15 years from now. And I think it's highly likely that those systems will have completely astronomical impact on society. Will humans actually benefit? And who will benefit, who will not?

In 2012, IBM estimated that an average person is generating 500 megabytes of digital footprints every single day. Imagine that you wanted to back up one day worth of data that humanity is leaving behind on paper. How tall will be the stack of paper that contains just one day worth of data that humanity is producing? It's like from the earth to the sun, four times over. In 2025, we'll be generating 62 gigabytes of data per person per day. We are living a ton of digital footprints while going through our lives. They provide computer algorithms with a fairly good idea about who we are, what we want, what we are doing. In my work, I looked at different types of digital footprints. I looked at Facebook likes. I looked at language, credit card records, web browsing histories, search records. And each time I found that if you get enough of this data, you can accurately predict future behavior and reveal important intimate traits. This can be used in great ways, but it can also be used to manipulate people. Facebook is delivering daily information to 2 billion people or more. If you slightly change the functioning of Facebook engine, you can move the opinions and hence the votes of millions of people. A politician wouldn't be able to figure out which message each one of his or her voters would like, but a computer can see what political message would be particularly convincing. for you.

Ladies and gentlemen, it's my privilege to speak to you today about the power of big data and psychographics in the electoral process.

Data firm Cambridge Analytica secretly harvested the personal information of 50 million unsuspecting Facebook users.

USA.

USA! USA! The data firm hired by Donald Trump's presidential election campaign used secretly obtained information to directly target potential American voters.

With that, they say they can predict the personality of every single adult in the United States.

Tonight, we're hearing from Cambridge Analytica whistleblower Christopher Wiley.

What we worked on was data harvesting programs where we would pull data and run that data through algorithms that could profile their personality traits and other psychological attributes.

To exploit mental vulnerabilities that our algorithms showed that they had.

Cambridge Analytica mentioned once or said that their models were based on my work. But Cambridge Analytica is just one of the hundreds of companies that are using such methods to target voters. You know, I would be asked questions by journalists such as, you know, so how do you feel about electing Trump and supporting Brexit? You know, how do you answer to such question? I guess that I have to deal with being blamed for all of it.

How tech started was as a democratizing force, as a force for good, as an ability for humans to interact with each other without gatekeepers. There's never been a bigger experiment in communications for the human race. What happens when everybody gets to have their say? You would assume that it would be for the better, that there'd be more democracy, there'd be more discussion, there'd be more tolerance. But what's happened is that these systems have been hijacked.

We stand for connecting every person, for a global community.

One company, Facebook, is responsible for the communications of a lot of the human race. Same thing with Google. Everything you want to know about the world comes from them. This is a global information economy that is controlled by a small group of people.

The world's richest companies are all technology companies. Google, Apple, Microsoft, Amazon, Facebook. It's staggering how, in probably just 10 years, that the entire corporate power structure are basically in the business of trading electrons. These little bits and bytes are really the new currency.

The way that data is monetized is happening all around us, even if it's invisible to us. Google has every amount of information available. They track people by their GPS

location. They know exactly what your search history has been. They know your political preferences. Your search history alone can tell you everything about an individual, from their health problems to their sexual preferences. So Google's reach is unlimited.

So we've seen Google and Facebook rise into these large surveillance machines, and they're both actually ad brokers. It sounds really mundane, but they're high-tech ad brokers. And the reason they're so profitable is that they're using artificial intelligence to process all this data about you, and then to match you with the advertiser that wants to reach people like you for whatever message.

One of the problems with technology is that it's been developed to be addictive. The way these companies design these things is in order to pull you in and engage you. They want to become essentially a slot machine of attention. So you're always paying attention. You're always jacked into the matrix. You're always checking.

When somebody controls what you read, they also control what you think. You get more of what you've seen before and liked before because this gives more traffic. And that gives more ads. But it also locks you into your echo chamber. And this is what leads to this polarization that we see today. Jair Bolsonaro, Brazil's right-wing populist candidate, sometimes likened to Donald Trump, winning the presidency Sunday night in that country's most polarizing election in decades.

Bolsonaro! What we are seeing around the world is upheaval in polarization and conflict that is partially pushed by algorithms that's figured out that political extremes, tribalism, and sort of shouting for your team and feeling good about it, is engaging.

Social media may be adding to the attention to hate crimes around the globe.

It's about how people can become radicalized by living in the fever swamps of the internet.

So is this a key moment for the tech giants? Are they now prepared to take responsibility as publishers for what they share with the world?

If you deploy a powerful, potent technology at scale, and if you're talking about Google and Facebook, you're deploying things at scale of billions. If your artificial intelligence is pushing polarization, you have global upheaval, potentially. White lives.

Matter! White lives matter! White lives matter! White lives matter! White lives matter! White lives matter.

Artificial General Intelligence, AGI. Imagine your smartest friend with 1000 friends just as smart. And then run them at 1000 times faster than real time. So it means that in every day of our time, they will do three years of thinking. Can you imagine how much you could do For every day, you could do three years' worth of work. It wouldn't be an unfair comparison to say that what we have right now is even more exciting than, you

know, the quantum physicists of the early 20th century that discovered nuclear power. I feel extremely lucky to be taking part in it. Many machine learning experts, people who are very knowledgeable and very experienced, have a lot of skepticism about HEI, about when it could happen and about whether it could happen at all. But right now this is something that just not that many people have realized yet, that the speed of computers for neural networks, for AI, are going to become maybe 100,000 times faster in a small number of years. The entire hardware industry, for a long time, didn't really know what to do next. But these artificial neural networks, now that they actually work, you have a reason to build huge computers. You can build a brain in silicon. It's possible. But the very first AGIs will be basically very, very large data centers. Compact with specialized neural network processors working in parallel. Compact, hot, power-hungry package, consuming like 10 million homes worth of energy. Do you have a roast beef sandwich? Yeah. Something slightly different. Just this one. But even the very first AGIs will be dramatically more capable than humans. Humans will no longer be economically useful for nearly any task. Why would you want to hire a human if you could just get a computer that's going to do it much better and much more cheaply? AGI is going to be like, without question, the most important technology in the history of the planet by a huge margin. It's going to be bigger than electricity, nuclear, the internet combined. In fact, you could say that the whole purpose of all human science, the purpose of computer science, the end game, this is the end game, to build this. And it's going to be built. It's going to be a new life form. It's going to be We're going to make us obsolete.

We got programmed to go down here where we... European manufacturers know the Americans have invested heavily in the necessary hardware.

To step into a brave new world of power, performance, and productivity. All.

Of the images you are about to see on the large screen will be generated by what's in that Macintosh.

Ladies and gentlemen, it's my honor and privilege to introduce to you the Windows 95 development team.

Human physical labor has been mostly obsolete for getting on for a century. Routine human mental labor is rapidly becoming obsolete. And that's why we're seeing a lot of middle-class jobs. disappearing.

Every once in a while, a revolutionary product comes along that changes everything. Today, Apple is reinventing the phone.

Machine intelligence is already all around us. The list of things that we humans can do better than machines is actually shrinking pretty fast.

Driverless cars are great. They probably will reduce accidents. Except, alongside with that, in the United States, you're gonna lose 10 million jobs. What are you going to do with 10 million unemployed people? The risk for social conflict and tensions If you exacerbate inequalities, it's very, very high.

AGI can, by definition, do all jobs better than we can do. People who are saying, Ah, there'll always be jobs that humans can do better than machines, are simply betting against science and saying they'll never be AGI.

What we're seeing now is like a train hurtling down a dark tunnel at breakneck speed, and it looks like we're sleeping at the wind.

A large fraction of the digital footprints we're leaving behind are digital images. And specifically, what's really interesting to me as a psychologist are digital images of our faces. Here you can see the difference in the facial outline of an average gay, an average straight face. And you can see that straight men have slightly broader jaws. Gay women have slightly larger chose compared with straight women. Computer algorithms can reveal our political views or sexual orientation or intelligence just based on the picture of our faces. Even a human brain can distinguish between gay and straight men with some accuracy. Again, now it turns out that the computer can do it with much higher accuracy. What you're seeing here is an accuracy of off-the-shelf facial recognition software. This is terrible news for gay men and women all around the world. And not only gay men and women, because the same algorithms can be used to detect other intimate traits, think being a member of the opposition, or being a liberal, or being an atheist. Being an atheist is also punishable by death in Saudi Arabia, for instance. My mission as an academic is to warn people about the dangers of algorithms being able to reveal our intimate traits. The problem is that when people receive bad news, they very often choose to dismiss them. Well, it's a bit scary when you start receiving death threats from one day to another, and I received quite a few death threats. But as a scientist, I have to basically show what is possible. So what I'm really interested in now is to try to see whether we can predict other traits from people's faces. Now, if you can detect depression from face or suicidal thoughts, maybe a CCTV system on the train station can save some lives. What if we could predict that someone is more prone to commit a crime? You probably had a school counselor, a psychologist hired there to identify children that potentially may have some behavioral problems. So now imagine if you could predict with higher accuracy that someone is likely to commit a crime in the future, from the language used, from their face, from their facial expressions, from the likes on Facebook, I'm not developing new methods, I'm just describing something or testing something in an academic environment. But there obviously is a chance that while warning people against risks of new technologies, I may also give some people new ideas.

We haven't yet seen the future in terms of the ways in which the new data-driven society is going to really evolve. The tech companies want to get every possible bit of information that they can collect on everyone to facilitate business. The police and the military want to do the same thing to facilitate security. The interests that the two have in common are immense, and so the extent of collaboration between what you might call a military tech complex is growing dramatically.

The CIA, for a very long time, has maintained a close connection with Silicon Valley. Their venture capital firm, known as In-Q-Tel, makes seed investments to startup companies developing breakthrough technology that the CIA hopes to deploy. Palantir, the big data analytics firm, one of their first seed investments was from In-Q-Tel.

In-Q-Tel struck gold in Palantir in helping to create a private vendor that has intelligence and artificial intelligence capabilities that the government can't even compete with.

Good evening. I'm Peter Thiel. I'm not a politician, but neither is Donald Trump.

He is a builder, and it's time to rebuild America.

Peter Thiel, the founder of Palantir, was a Donald Trump transition advisor and a close friend and donor. Trump was elected largely on the promise to deport millions of immigrants. The only way you can do that is with a lot of intelligence, and that's where Palantir comes in. They ingest huge troves of data, which include where you live, where you work, who you know, who your neighbors are, who your family is, where you have visited, where you stay, your social media profile. Palantir gets all of that and is remarkably good at structuring it in a way that helps law enforcement immigration authorities or intelligence agencies of any kind track you, find you, and learn everything there is to know about you.

We're putting AI in charge now of ever more important decisions that affect people's lives. Old-school AI used to have its intelligence programmed in by humans who understood how it worked. But today, powerful AI systems have just learned for themselves, and we have no clue really how they work, which makes it really hard to trust them.

This isn't some futuristic technology. This is now. AI might help determine where a fire department is built in a community or where a school is built. It might decide whether you get bail or whether you stay in jail. It might decide where the police are going to be. It might decide whether you're going to be under additional police scrutiny.

It's popular now in the U.S. to do predictive policing. So what they do is they use an algorithm to figure out where crime will be. And then they use that to tell where we should send police officers. So that's based on a measurement of crime rate. So we know that there is bias. Black people and Hispanic people are pulled over or stopped by the police officers more frequently than white people are. So we have this bias data

going in. And then what happens is you use that to say, oh, here's where the cops should go. Well, the cops go to those neighborhoods, and guess what they do? They arrest people. And then it feeds back bias data into the system. And that's called a feedback loop.

Predictive policing leads at the extremes to experts saying, Show me your baby, and I will tell you whether she's going to be a criminal. Now that we can predict it, we're going to then surveil those kids much more closely. and we're going to jump on them at the first sign of a problem, and that's going to make for more effective policing. It does, but it's going to make for a really grim society, and it's reinforcing dramatically existing injustices.

Imagine a world in which networks of CCTV cameras, drone surveillance cameras, have sophisticated face recognition technologies and are connected to other government surveillance databases. We will have the technology in place to have all of our movements comprehensively tracked and recorded. What that also means is that we will have created a surveillance time machine that will allow governments and powerful corporations to essentially hit rewind on our lives. We might not be under any suspicion now, and five years from now, they might want to know more about us and can then recreate granularly everything we've done, everyone we've seen, everyone we've been around over that entire period. That's an extraordinary amount of power for us to cede to anyone. And it's a world that I think has been difficult for people to imagine. But we've already built the architecture to enable that.

I'm a political reporter, and I'm very interested in the ways powerful industries use their political power to influence the public policy process. The large tech companies and their lobbyists get together behind closed doors and are able to craft policies that we all have to live under. That's true for surveillance policies, for policies in terms of data collection, but also increasingly important when it comes to military and foreign policy. Starting in 2016, the Defense Department formed the Defense Innovation Board. That's a special body created to bring top tech executives into closer contact with the military. Eric Schmidt, former chairman of Alphabet, the parent company of Google, became the chairman of the Defense Innovation Board. And one of their first priorities was to say, we need more artificial intelligence integrated into the military. I've worked with a group of volunteers over the last couple of years to take a look at innovation in the overall military. And my summary conclusion is that we have fantastic people who are trapped in a very bad system.

From the Department of Defense's perspective, where I really started to get interested in it, when we started to think about unmanned systems and how robotic and unmanned systems would start to change war. The smarter you made the unmanned systems and robots, the more powerful you might be able to make your military.

Under Secretary of Defense Robert Work put together a major memo known as the Algorithmic Warfare Cross-Functional Team, better known as Project Maven. Eric Schmidt gave a number of speeches and media appearances where he said this effort was designed to increase fuel efficiency in the Air Force, to help with the logistics. But behind closed doors, there was another parallel effort. Late in 2017, as part of Project Maven, Google, Eric Schmidt's firm, was tapped to secretly work on another part of Project Maven, and that was to take the vast volumes of image data vacuumed up by drones operating in Iraq and Afghanistan, and to teach an AI to quickly identify targets on the battlefield.

We have a sensor, and the sensor can do full motion video of an entire city. And we would have three seven-person teams working constantly, and they could process 15% of the information. The other 85% of the information was just left on the cutting room floor. So we said, hey, AI and machine learning would help us process 100% of the information.

Google has long had the motto, don't be evil. They have created a public image that they are devoted to public transparency. But for Google to slowly transform into a defense contractor, they maintained the utmost secrecy. You had Google entering into this contract with most of the employees, even employees who were working on the program, completely left in the dark.

Usually within Google, anyone in the company is allowed to know about any other project that's happening in some other part of the company. With Project Maven, the fact that it was kept secret, I think, was kind of alarming to people, because that's not the norm at Google.

When the story was first revealed, it set off a firestorm within Google. You had a number of employees quitting in protest, others signing a petition, objecting to this work.

You have to really say, like, I don't want to be part of this anymore. There are companies called defense contractors, and Google should just not be one of those companies, because people need to trust Google for Google to work.

Good morning.

Welcome to Google I/O.

We've seen emails that show that Google simply continued to mislead their employees, that the drone targeting program was only a minor effort that could at most be worth $9 million to the firm, which is, you know, drops in the bucket for a gigantic company like Google. But from internal emails that we obtained, Google was expecting Project Maven would ramp up to as much as $250 million, and that this entire effort would provide Google with special Defense Department certification to make them available for even bigger defense contracts, some worth as much as $10 billion. The pressure for Google

to compete for military contracts has come at a time when its competitors are also shifting their culture. Amazon similarly pitching the military and law enforcement. IBM and other leading firms, they're pitching law enforcement and military. To stay competitive, Google has slowly transformed.

The Defense Science Board said of all of the technological advances that are happening right now, the single most important thing was artificial intelligence and the autonomous operations that it would lead. Are we investing enough?

Once we develop what are known as autonomous, lethal weapons, in other words, weapons that are not controlled at all, they are genuinely autonomous. You've only got to get a president who says, the hell with international law, we've got these weapons, we're going to do what we want with them.

We're very close.

When you have the hardware already set up and all you have to.

Do is flip a switch to make it fully autonomous, what is it there that's stopping you from doing that? There's something really to be feared in war at machine speed.

What if you're a machine and you've run millions and millions of different war scenarios and you have a team of drones and you've delegated control to half of them and you're collaborating in real time? What happens when that swarm of drones is tasked with engaging a city?

How will they take over that city?

The answer is we won't know until it happens.

We do not want an AI system to decide what human it would attack, but we're going up against authoritarian competitors. So in my view, an authoritarian regime will have less problem delegating authority to a machine to make lethal decisions. So how that plays out remains to be seen.

Almost the gift of AI now is that it will force us collectively to think through at a very basic level what does it mean to be human, What do I do as a human better than a certain super smart machine can do? First we create our technology and then it recreates us. We need to make sure that we don't miss some of the things that make us so beautifully human.

Once we build intelligent machines, The philosophical vocabulary we have available to think about ourselves as human increasingly fails us. If I ask you to write up a list of all the terms you have available to describe yourself as human, there are not so many terms. Culture, history, sociality, maybe politics, civilization, subjectivity. All of these terms ground in two positions. That humans are more than mere animals and that

humans are more than mere machines. But if machines truly think... There is a large set of key philosophical questions in which what is at stake is. Who are we? What's our place in the world? What is the world? How is it structured? Do the categories that we have relied on? Do they still work? Are they wrong?

Many people think of intelligence as something mysterious that can only exist inside of biological organisms like us. But intelligence is all about information processing. It doesn't matter whether the intelligence is processed by carbon atoms inside of cells and brains and people, or by silicon atoms and computers. Part of the success of AI recently has come from stealing great ideas from evolution. We noticed that the brain, for example, has all these neurons inside connected in complicated ways. So we stole that idea and abstracted it into artificial neural networks in computers. And that's what's revolutionized machine intelligence. If we one day get artificial general intelligence, then by definition, AI can also do better the job of AI programming. And that means that further progress in making AI will be dominated not by human programmers, but by AI. Recursively self-improving AI could leave human intelligence far behind, creating super intelligence. It's going to be the last invention we ever need to make. because it can then invent everything else much faster than we could.

It's... it's...

It's...

It's...

There is a future that we all need to talk about. Some of the fundamental questions about the future of artificial intelligence. Not just where it's going, but what it means for society to go there. It is not what computers can do, but what computers should do. As the generation of people that is bringing AI to the future.

We are the generation that will answer this question first and foremost.

We haven't created the human-level thinking machine yet, but we get closer and closer. Maybe we'll get to human-level AI in five years from now, or maybe it'll take 50 or 100 years from now. It almost doesn't matter. Like, these are all really, really soon in terms of the overall history of humanity. Very nice. So the AI field is extremely international. China is up and coming and is starting to rival the US, Europe, and Japan in terms of putting a lot of processing power behind AI and gathering a lot of data to help AI learn. We have a young generation of Chinese researchers now. Nobody knows where the next revolution is going to come from.

China always want to become a superpower in the world. Chinese government think they gave them the chance to become one of the most advanced, technology-wise, business-wise. So Chinese government look at this as a huge opportunity. Like they raise the flag, I said, that's good field. The company should jump into it. Then China's

commercial world is a company just, okay, come and raise the flag. That's good. Let's put the money into it. Chinese tech giants like Baidu, like the Tencent, like Alibaba, they put a lot of the investment into the AI field. So we see that China's AI development is booming.

In China, everybody has Alipay and WeChat pay. So the mobile payments is everywhere. And with that, they can do a lot of AI analysis to know your spending habits, like your credit rating. Face recognition technology is widely adopted in China, in airport, in train station. So in the future, Maybe in just a few months, you don't need a paper ticket to board a train. Only your face.

We generate the world's biggest platform of facial recognition. We have 300,000 developers are using our platform. A lot of it is the selfish camera apps. It makes you look more beautiful. There is millions and millions of cameras in the world. Each camera, from my point, is about a data generator. In the machine eye, your face will change into the features, and it will turn your face into a paragraph of the code. So we can detect how old are you, you male or female, and your emotions. Shopping is about what kind of thing you are looking at, so that we can track in your eyeballs. So, if you are focusing on some product, we can track in that, so that we can... and know which kind of people like which kind of product. . . Don't should have the same way. Sam Guan. So that I did. Yeah, Sam Guan. So you should push your car. So you go in.

The Chinese government is using multiple different kinds of technologies, whether it's AI, whether it's big data platforms, facial recognition, voice recognition, essentially to monitor what the population is doing. I think the Chinese government has made very clear its intent to gather massive amounts of data about people to socially engineer a dissent-free society. The logic behind the Chinese government's social credit system, it's to take the idea that whether you are creditworthy for a financial loan and adding to it a very political dimension to say, are you a trustworthy human being? what you've said online. Have you ever been critical of the authorities? Do you have a criminal record? And all of that information is packaged up together to rate you in ways that if you have performed well, in their view, you'll have easier access to certain kinds of state services or benefits. But that if you haven't done very well, you're going to be penalized or restricted. There's no way for people to challenge those designations, or in some cases, even know that they've been put in that category. And it's not until they try to access some kind of state service, or buy a plane ticket, or get a passport, or enroll their kid in school, that they come to learn that they've been labeled in this way, and that there are negative consequences for them as a result. We've spent the better part of the last one or two years looking at abuses of surveillance technology across China. And a lot of that work has taken us to Xinjiang, the northwestern region of China that has more than half population of Turkic Muslims, Uyghurs, Kazakhs, Hui. This is a region and a population the Chinese government has long considered to be politically suspect

or disloyal. We came to find information about what's called the Integrated Joint Operations Platform, which is a predictive policing program. And that's one of the programs that has been spitting out lists of names of people to be subjected to political re-education. A number of our interviewees for the report we just released about the political education camps in Xinjiang just painted an extraordinary portrait of a surveillance state. A region awash in surveillance cameras for facial recognition purposes, checkpoints, body scanners, QR codes outside people's homes. It really is the stuff of dystopian movies that we've all gone to and thought, wow, that would be a creepy world to live in. Yeah, well, 13 million Turkic Muslims in China are living in that reality right now.

The Intercept reports that Google is planning to launch a censored version of its search engine in China.

Google's search for new markets leads it to China, despite Beijing's rules on censorship.

Tell us more about why you felt it was your ethical responsibility to resign, because you talk about being complicit in censorship and oppression and surveillance. There is a Chinese venture company that has to be set up for Google to operate in China. And the question is, to what degree did they get to control the blacklist, and to what degree would they have just unfettered access to surveilling Chinese citizens? And the fact that Google refuses to respond to human rights organizations on this, I think, should be extremely disturbing to everyone. Due to my conviction that dissent is fundamental to functioning democracies, I'm forced to resign in order to avoid contributing to or profiting from the erosion of protections for dissidents. The United Nations is currently reporting that between 200,000 and 1 million Uyghurs have been disappeared into re-education camps. And there is a serious argument that Google would be complicit should it launch a surveilled version of search in China. Dragonfly is a project meant to launch search in China under Chinese government regulations, which include censoring sensitive content, basic queries on human rights, information about political representatives is blocked. information about student protest is blocked. And that's one small part of it. Perhaps a deeper concern is the surveillance side of this. When I raised the issue with my managers, with my colleagues, there was a lot of concern, but everyone just said, I don't know anything. And then when there was a meeting, finally, there was essentially no addressing the serious concerns associated with it. So then I filed my formal resignation, not just to my manager, but I actually distributed it company-wide, and that's the letter that I was reading from. Personally, I haven't slept well. I've had pretty horrific headaches. I wake up in the middle of the night just sweating. With that said, what I've found since speaking out is just how positive the global response to this has been. Engineers should demand to know what the uses of their technical contributions are and to have a seat at the table in those ethical

decisions. Most citizens don't really understand what it means to be in a very large-scale prescriptive technology. Where someone has already pre-divided the work, And all you know about is your little piece. And almost certainly, you don't understand how it fits in. So I think it's worth drawing the analogy to physicists' work on the atomic bomb. In fact, that's actually the community I came out of. I wasn't a nuclear scientist by any means, but I was an applied mathematician. And my PhD program was largely funded to train people to work in weapons labs. One could certainly argue that there is an existential threat, and whoever is leading in AI will lead militarily.

China fully expects to pass the United States as the number one economy in the world, and it believes that AI will make that jump more quickly, and more dramatically. And they also see it as being able to leapfrog the United States in terms of military power. Their plan is very simple. We want to catch the United States in these technologies by 2020. We want to surpass the United States in these technologies by 2025. And we want to be the world leader in AI and autonomous technologies by 2030. It is a national plan. It is backed up by at least $150 billion in investments. So this is definitely a race.

AI is a little bit like fire. Fire was invented 700,000 years ago, and it has its pros and cons. People realized you can use fire to keep warm at night and to cook. But they also realized that you can kill other people with it. Fire also has this AI-like quality of growing in a wildfire without further human ado. But the advantages outweigh the disadvantages by so much that we are not going to stop its development. Europe is waking up. Lots of companies in Europe are realizing that the next wave of AI will be much bigger than the current wave. The next wave of AI will be about robots. All these machines that make things, that produce stuff, that build other machines, They are going to become smart. In the not-so-distant future, we will have robots that we can teach like we teach kids. For example, I will talk to a little robot and I will say, look here, robot.

Look here.

Let's assemble a smartphone. We take the slab of plastic like that and we take the screwdriver like that. And now we screw in everything like this. No, not like this. Like this. Look, robot, look. Like this. And he will fail a couple of times, but rather quickly he will learn to do the same thing much better than I could do it. And then we start the learning and we make a million copies and sell it. Regulation of AI sounds like an attractive idea, but I don't think it's possible. One of the reasons why it won't work is the sheer curiosity of scientists. They don't give a damn for regulation. Military powers won't give a damn for regulations either. They will say, if we, the Americans, don't do it, then the Chinese will do it. And the Chinese will say, oh, if we don't do it, then the Russians will do it. No matter what kind of political regulation is out there, all these military-industrial complexes, they will, almost by definition, have to ignore that, because they want to avoid falling behind.

The program developed by the company.

OpenAI can write coherent and credible stories just like human beings.

It's one small step for machine, one giant leap for machine kind. IBM's newest artificial intelligence system took on experienced human debaters and won a live debate.

Computer-generated videos known as deepfakes are being used to put women's faces on pornographic videos.

Artificial intelligence evolves at a very crazy pace. You know, it's, like, progressing so fast. In some ways, we're only at the beginning right now. You have so many potential applications. It's a goldmine. Since 2012, when deep learning became like a big game changer in the computer vision community, we were one of the first to actually adopt deep learning and apply it in the field of computer graphics. A lot of our research is funded by government, military, intelligence agencies. The way we create these photoreal mappings Usually the way it works is that we need two subjects, a source and a target, and I can do a face replacement. One of the applications is, for example, I want to manipulate someone's facing things that he did not. It can be used for creative things, for funny content, but obviously it can also be used for just simply manipulate videos and generate fake news. This can be very dangerous. if it gets into the wrong hands, it can get out of control very quickly.

We're entering an era in which our enemies can make it look like anyone is saying anything at any point in time, even if they would never say those things. Moving forward, we need to be more vigilant with what we trust from the Internet. It may sound basic, But how we move forward in the age of information is gonna be the difference between whether we survive or whether we become some kind of ********* dystopia.

One criticism that is frequently raised against my work is saying that, hey, there were stupid ideas in the past, like phrenology or physiognomy. There were people claiming that you can read a character of a person just based on their face. People would say, this is rubbish. We know it was just thinly veiled racism and superstition. But the fact that someone made a claim in the past and tried to support this claim with invalid reasoning doesn't automatically invalidate the claim. Of course, people should have rights to their privacy when it comes to sexual orientation, or political views. But I'm also afraid that in a current technological environment, this is essentially impossible. People should realize there's no going back, there's no running away from the algorithms. The sooner we accept the inevitable and inconvenient truth that Privacy is gone. The sooner we can actually start thinking about how to make sure that our societies are ready for the post-privacy age.

While speaking about facial recognition, in my deep thoughts I sometimes get to the very dark era of our history. When the people had to live in the system where some part

of the society was accepted and some part of the society was accused to death. What would Mengele do to have such an instrument in his hands? It would be very quick and efficient for selection. And this is the apocalyptic vision.

So in the near future, the entire story of you will exist in a vast array of connected databases of faces, genomes, behaviors, and emotion. So you will have a digital avatar of yourself online, which records how well you are doing as a citizen, What kind of relationship do you have? What kind of political orientation and sexual orientation? Based on all of those data, those algorithms will be able to manipulate your behavior with an extreme precision, changing how we think and probably in the future how we feel.

The beliefs and desires of the first AGIs will be extremely important. So it's important to program them correctly. I think that if this is not done, then the nature of evolution, of natural selection, will favor those systems, prioritize their own survival of other us. It's not that it's going to actively hate humans and want to harm them. But it's just going to be too powerful. And I think a good analogy would be the way humans treat animals. It's not that we hate animals. I think humans love animals and have a lot of affection for them. But when the time comes to build a highway between two cities, we are not asking the animals for permission. We just do it, because it's important for us. And I think by default, that's the kind of relationship that's going to be between us and AGIs which are truly autonomous and operating on their own behalf. If you have an arms race dynamics between multiple teams trying to build the AGI first, they will have less time to make sure that the AGI that they will build will care deeply for humans. Because the way I imagine it is that there is an avalanche. There is an avalanche of AGI development. Imagine you have this huge, unstoppable force. And I think it's pretty likely the entire surface of the earth will be covered with solar panels and data centers. Given these kinds of concerns, it will be important that AGI is somehow built as an operation within multiple countries. The future is going to be good for the AIs regardless. It would be nice if it were good for humans as well.

Is there a lot of responsibility weighing on my shoulders? Not really. Was there a lot of responsibility on the shoulders of the parents of Einstein? The parents somehow made him, but they had no way of predicting what he would do and how he would change the world. And so you can't really hold them responsible for that. So I'm not a very human-centric person. I think I'm a little stepping stone in the evolution of the universe towards higher complexity. But it's also clear to me that I'm not the crown of creation and that humankind as a whole is not the crown of creation. But we are setting the stage for something that is bigger than us, that transcends us. And that will go out there in a way where humans cannot follow and transform the entire universe, or at least the reachable universe. So I find beauty and awe in seeing myself as part of this much grander theme.

AI is inevitable. We need to make sure we have the necessary human regulation to prevent the weaponization of artificial intelligence. We don't need any more weaponization of such a powerful tool.

One of the most critical things, I think, is the need for international governance. We have an imbalance of power here. So now we have corporations with more power, might, and ability than entire countries. How do we make sure that people's voices are getting heard?

It can't be a law-free zone. It can't be a rights-free zone. We can't embrace all of these wonderful new technologies for the 21st century without trying to bring with us the package of human rights that we fought so hard to achieve and that remains so fragile.

AI isn't good, and it isn't evil either. It's just going to amplify the desires and goals of whoever controls it. And AI today is under the control of a very, very small group of people. The most important question that we humans have to ask ourselves at this point in history requires no technical knowledge. It's the question of what sort of future society do we want to create? with all this technology we're making. What do we want the role of humans to be in this world?