The Future of Technology – Benefits and Risks

Max Tegmark: Let me give the word to our ex-President of MIT, Susan Hockfield. Let's welcome her.

[applause]

Susan Hockfield: Welcome everyone to MIT. I'm thrilled to see so many people here. It is Memorial Day weekend. Did anyone tell you that?

This is the weekend you're supposed to be having, grilling outside, and having various kinds of fun. To my mind, this is really fun.

[laughter]

Susan: I'm so happy to see to so many people who share my sense of fun. Max asked me just to say hello and welcome you all to MIT.

This is a great event for MIT to be hosting. A great opportunity to be thinking about the future, and what I trust our fabulous panelists will talk about are ways to think about the future in very, very positive ways.

We can leave here charged up and ready to march into that future with smiles on our faces. Perhaps we'll hear some of the down side too, but as for me, I'm going to be concentrating on the upside.

Max, it's great to be here. Thank you very much for organizing this. I'm sure it will be a fantastic evening.

Max: Thank you so much, Susan.

[applause]

Max: I'm Max Tegmark of the physics department. We're all here tonight because we love technology. Every way in which 2014 is better than the Stone Age is because of technology.

[laughter]

Max: We're also here tonight because we know that technology is so great that there is a dangerous temptation sometimes to only pay attention to the powerful applications and not pay attention to how to avoid the associated risks.

You'll hear a lot of examples of this from our panel, shortly. As just one example of how we're not paying enough attention, in my opinion, to technology risk, let me just ask you this: which of these two guys have we paid more attention to as a human species?

[shows a slide of Justin Bieber and Vasili Arkhipov]

Let me ask you one more question, which of these two guys should we thank for us all being alive here tonight because he single handedly stopped a nuclear attack during the Cuban Missile Crisis?

[laughter]

Max: Is that really the optimal way for us humans to allocate our attention?

To help remedy this and thanks very much for the generous support of Jaan Tallinn, we are launching an organization to focus more attention on the long term success of humanity, these positive things that Susan Hockfield so eloquently mentioned here.

My fellow physics professor, Anthony Aguirre from Santa Cruz, will now tell us a bit more about this.

Anthony Aguirre: As Max said, a number of us got together and thought: is there something we can do about this allocation of attention in society? We were gratified to find that although a lot of people aren't paying attention, there are some people, and among them, in fact, some of the most impressive technologists and scientists and communicators out there, are paying attention and have been paying attention.

We also very gratified to find that despite their incredible busy schedule and workload, they were more than willing to help out and give us their advice. They comprise now the scientific advisory board for this organization that the five of us that you'll see in front of you here have been concocting.

The idea of this organization that we envision is to both put some effort into really understanding the risks and pitfalls of technology over the next few decades, but more importantly envisioning positive pictures of what the future will look like. Not the dystopia that you see in every Hollywood movie that comes out, but how do we really want things to be and how can we in some ways set an initiative that guides us or steers us a little bit more in that direction, a little bit less in the less palatable directions that we all know that we could fall into?

This event here is just the first of what we hope to be many, many different initiatives that this institute will run. This is something that being run basically on a purely volunteer basis by everyone involved. We would like to get any of you who are interested involved, as well. Meia is going to tell you a little bit about that.

Meia Chita-Tegmark: As Max and Anthony very eloquently put it, FLI stems from this coming together of great minds, people who are very interested in safeguarding the future of life on this planet and maybe outside of this planet, also.

This does not end with the founders and our fabulous advisory board, but it also includes you. Please check us out at the Future of Life Institute on our website, and we would be very, very happy to connect to people who have been thinking about these ideas that you will be hearing about tonight from our panelists, probably independently.

We would like to be part of our group. Also, if this is new to you, if the ideas that you have heard tonight are probably the first time that you come across them, you're welcome to check our website and read more about them and look at our resources that we have put together.

Also, if you are a specialist in the field and would like to contribute something to our resources list, please contact us, and we would be happy to post all the useful links from your suggestions.

Now, I just want to invite Vika to introduce you to our volunteers.

[applause]

Viktoriya Krakovna: Before we launch into the panel, I would like to thank all the people who made this event happen. I would like to thank our advisory board, and I would like to thank especially all our volunteers and members who have done so much hard work to bring this event together.

Any of you who are here in the audience, could you please stand up and be recognized? Any of our volunteers and members.

[applause]

Viktoriya: We're also very grateful to MIT for letting us host this event here in this awesome auditorium and especially to the MIT staff and administrators who have helped us bring this together.

Also, of course, I would like to thank our awesome panelists who have joined us here today. Today, we have George Church, who is a professor of genetics at the Harvard Medical School. We have Ting Wu, also a professor of genetics at the Harvard Medical School. We have Andrew McAfee, the associate director of the MIT Center for Digital Business. Frank Wilczek, a physics professor at MIT and Nobel laureate. Jaan Tallinn, a founding engineer of Skype. Our moderator today, Alan Alda, is an actor, writer, director and science communicator. I would like to thank Alan for moderating our panel today. Without further ado, please welcome Alan!

[applause]

Alan: Thank you. It might be a good idea for us just to go down the table and take a couple of minutes to talk about this in general, each from your own perspective. Why don't we start...oh, I'm not on. Look at that.

[laughter]

Alan: I am wearing a microphone but it's not doing any good. This is MIT.

[laughter]

[applause]

George Church: The future of technology.

Alan: The Massachusetts Institute of close enough.

[laughter]

[applause]

Alan: I'm just kidding. Come on. We'll share this mike. George, why don't you start?

George: I'm George Church, professor of genetics at Harvard Medical School. Also involved in MIT Media Lab, Broad Institute of Health Sciences and Technology.

George: I'm George Church, professor of genetics at Harvard Medical School, and also involved in the MIT Media Lab, Broad Institute of Health Sciences and Technology.

Ting: I'm Ting Wu, a professor at the Department of Genetics. I study inheritance and weird stuff, and I'm very pleased to be here. I will be representing today my work with the Personal Genetics Education Project, which has to do with raising awareness of personal genetics.

Andrew: My name is Andy McAfee, I am here at the Sloan School at MIT, and I basically study how quickly the robots are going to take all of our jobs, and what happens after that.

Frank: I'm Frank Wilczek, I'm a professor at MIT, a professor of physics. I'm very interested in physics, and…

Andrew: He didn't just stumble on that Nobel Prize, is what he's trying to say.

Frank: And I'm also interested in phenomena that obey the laws of physics.

Jaan: I'm Jaan Tallinn, probably the only Estonian here. I'm cofounder of Skype, and in the last 6-7 years I've been looking out and supporting and trying to cofound, whenever I can, organizations like the Future of Life Institute here. I've also cofounded an institute in Cambridge, UK, which is the Center for the Study of Existential Risk.

Alan: Good, thank you. George, why don't you give us a few words, and we'll go down after you do that.

George: You've seen this slide a little bit. Maybe I'll stand at the podium for a moment. Just quickly, these are some of the risks that we will be talking about and the opportunities that come along with them that might outweigh the risks in many ways - physical, biological, and intelligence. I'm not just talking about artificial intelligence, but human and maybe hybrid intelligence, carbon plus silicon.

Some of the physical risks are beyond our control in a certain sense, and so our best bet is to get off the planet. To do so, there are opportunities. It's no longer a luxury. At some point, it is not a luxury, and the sooner we do it the less of a rush job it will be.

Biological: we have both synthetic and natural biological issues. These have many solutions. This list is not meant to be exhaustive for the risks or the opportunities, but we can do surveillance much better than we do. We can isolate experiments until we're really sure about them, or isolate industrial manufacturing, and finally we can develop resistance, resistance of humans and crops and so on. These are great opportunities and synthetic biology is part about that.

Humans are highly augmented. People who worry about humans becoming augmented, it's a little late to be worried about that. Because, we won't go through this whole list, but you can see all our senses and our ability to locomote around the world and outside the world are truly awesome compared to what our ancestors had. We can really see here how far beyond.

Most of these are changing our physics, the way that we interact with the physical world. But there is this awesome power of mutation, on which many of you may have seen this recently opened theater. But there are real actual mutations in human populations which are quite radical, and many of them are rare and positive. You usually hear about the rare negatives, but these are seen here as positives. Here's an example of a child that was born with Schwarzenegger-like muscles without taking the steroids or the working out the way that Arnold did. This has been confirmed in animal models, in cows, and dogs, and mice. You get this high muscle growth and low deleterious atherosclerosis.

Humans can vary in size from 2 kilograms to 600 kilograms. A lot of this is due to the growth hormone in the receptor port, and variations occur of a sort. The disease that I have of extra height can be cured. It results also in life extension, both in mice and similar phenomena (not actual longer life but similar phenomena of anti-aging) in humans as well.

There's a long list of these things. We won't go through the list, but these minus over minuses mean that you have a double-null. You're missing both copies from both your parents of these different genes. You might say, "Wow, those things are conserved in all animals, and here there are people walking around without either copy of it, so they can't be that important." Some of them are downright interesting in what they allow us to do.

We've already mentioned the growth hormone and myostatin, and there are other ones that result in virus resistance, for example, curing AIDS, low diabetes, and so forth. They reduce your risks.

We're not limited by our current biology. We can imagine beyond that. Just like our physical devices aren't limited by elements found in nature or the devices found in nature like rocks.

Here's an example with the radiation resistance of a lab bacterium - one of my favorites, industrial favorites, is E. coli. It's not a particularly radiation resistant organism, but you can make it the most radiation resistant organism on the planet with a series of four mutations essentially synthetically. This could be useful if we're to get off the planet, because some of these trips are quite long - maybe Mars is OK, but Europa and so forth, there's a lot of galactic radiation.

We have other opportunities as we leave the planet. We can choose whether we want to bring along our diseases with us. It could be like Noah's Ark where we include Ebola, and small pox, and all the rest, or we can chose to leave them behind. It's a big decision to be made and a big opportunity. It turns out that there are all sorts of germ-free animals, chickens and goats in the '20s, and mice and so forth, so we could have germ-free humans.

Anesthesia, here's a clip from "Grey's Anatomy," where you saw a typical hospital room. A lot of the devices have to do with contagion anesthesia.

But, there's this mutation I mentioned on an earlier slide, where people are double-null for this and they become insensitive to pain. It's not a dopey insensitivity you get from opiates, but it's an actual fully alert insensitivity to pain. It has downsides we can talk about if you want.

As we do synthetic biology, such as those that we just described, a lot of it's going into gene therapy. Gene therapy is doing well, but there are other kinds of synthetic biology. We need to urge a careful step forward - the Environmental Protection Agency and the Food and Drug Administration are part of that.

There are other things that involve bad actors and surveillance. I put forward this proposal in 2004 for doing such surveillance and taking advantage of where we are in certain bottlenecks. Four or five years later, there's some international agreement on this in some of these organizations.

We also have physical and biological resistance. We've made synthetic organisms that actually have new functions. This is the example of E. coli again, where we've got it so it's genetically and metabolically isolated through new chemistry. You can potentially get industrial interest in it because they are also multi virus resistant, so they can be a win-win, where they're safer, and they're more resistant.

Finally, just this topic of intelligence I'm going to put forward, just for the sake of being provocative, that it's not obvious that artificial intelligence with silicon is the only way forward for intelligence. There is biological intelligence. Even though Watson beat two of the top Jeopardy contestants, he was… he, Watson, was using, or she, named after Thomas Watson, so I say he, 200,000 watts of power, while the humans were only using 20 watts, in their brain at least.

[laughter]

George: I think also, Jeopardy is just really information retrieval, and a little bit of parsing of natural languages. It's not similar to what Einstein did in 1905, where he pulled out of his hat four or five pretty remarkable articles.

[laughter]

George: What if the average person were Einstein? There would be a whole new bell curve for biological intelligence that electronics would have to catch up with. Also, biology is capable of encoding things about a million times more densely, so I encoded my book in this manner at one point. A million times denser, not necessarily better.

As I mentioned with Watson, the computing efficiency in computer operations per Joule, there's about a six to nine order of magnitude between it and various biological operations, depending on how you measure this. If we keep extrapolating this kind of improvement curves, this is not exactly Moore's law, but something like it, it would be about three decades before it catches up, if biotechnology stays static, it doesn't change.

But it's not static. In fact, here are these, again, exponential curves where you have factors of 10 on the y-axis, and the dotted line is electronics, which is going really quite fast. We all know

how fast it's going, because we have these awesome supercomputers in our pockets now. But biology is going even faster.

This is a cost curve, where it's dropping off at a very rapid rate. Biotechnology is not static, and it's going to get even faster with respect to brain analyses, which is the natural or carbon-based computers, because Barack Obama started this amazing incentive for innovative neurotechnologies. I will turn it over to the next panelist.

[applause]

Ting: Thank you very much, it really is a pleasure to be here. As I mentioned, I do research on genetics, and I share a passion with George about preparing ourselves to leave this planet, how to protect ourselves against cosmic radiation, and I was talking with Matt at dinner about this, but what I'd like to talk about is perhaps a challenge bigger than what we've faced in genetics for a long, long time.

I can tell you that about 18 months ago, the process of genetics, and genetic engineering, took a gigantic leap, thanks in part to work done in George's laboratory, and we have really moved into a new era. I think one of the greatest challenges we have now is to make sure that there's comprehensive understanding about where we are across our society, and that's what I'd like to talk about. In particular, work done by the Personal Genetics Education Project.

I think you'll see that some of the challenges we face on the front line every day with this project of education will be shared by some of the challenges we talk about this evening. This project began with a recognition that as ideas and technologies develop, there will be variants that will eventually spread through our population. The success of those variant ideas and technologies will be due to, or reflect, the balance between the benefits and the risks of those ideas and technologies.

With hand washing before surgery being one of the best ideas we ever had, and probably eugenics being one of the worst ideas we've ever had. Now for many people, we wish that we had, at the time that eugenics was devised in the last century, done something to remove the misunderstanding, and the misuse of genetics. We think it'd be great if we could constantly get rid of ideas that are going to be detrimental to us. But this idea, as lovely as it is, is quite impractical.

In fact, this is probably a better view of the world of ideas that we live in. Now, one of the issues that we have to deal with is the fact that these ideas cannot actually be determined to be good or bad in real-time. We can only know the answer to that from when we look back in the future. What the Personal Genetics Education Project does is, we don't go into the public and try and say, "This is good, and this is bad."

We focus on raising understanding. We focus on distributing information so that individuals can develop a confidence to ask questions, and to make informed personal choices, know that for themselves some ideas may be bad, other ideas may be good, and that there's diversity in the population such that eventually we need to come to respect both ourselves, and that diversity in our population.

We are hoping for awareness, for confidence and respect, and we think that if we can gain this at a high enough level, we'll be able to foster discussion, and from that a knowledgeable, considered vigilance.

Basically, if we talk to each other more, in a knowledgeable way, that may be one of the best things we can do to maximize our ability to benefit from some of the ideas and technologies we put forward, and protect ourselves from decisions that we may regret later on.

What I'd like to talk about today are some of the strategies, five in particular that we have found very useful. None of those strategies actually involve the introduction of facts, because facts have a capacity to go in the head and out of the head, especially for people who are not actually interested in the topic that you're talking about. For me, it's tax forms. Facts go in, facts go out. I cannot keep them. However, what we have learned is that the things that we are so afraid to talk about - the explosive problems, the ethical issues, the quandaries, the dilemmas - the problems are actually what make our public step forward three steps, instead of turning around.

We think they're going to run from them, but actually, they are completely engrossed by them. The first strategy we have with the Personal Genetics Education Project is we talk about those implications, and those benefits, and those ethical issues that just grip us, and keep us up at night. For example, we talk about the use of genetic information for law enforcement, its use in capturing criminals, putting people away, keeping our society safe.

I can tell you, when I talk about this in societies and communities that are well off, there's not too much argument against it. But when I talk in communities that are disadvantaged, or underserved, the response is quite different, as you can understand, because they have not been well served by some of these databases that contain millions of people's information. They feel vulnerable. We talk about that in the classrooms, and then we also mention the Innocence Project, which has freed many, many people who were incarcerated for the wrong reasons, and were saved because their DNA matches did not match that of the DNA collected at a crime scene.

We also capture audiences, especially high school students, by talking about pre-implantation genetic diagnosis, which involves the generation of embryos in a Petri dish, the isolation of one of the cells at the eight cell stage, or the 16 cell stage, and the genetic analysis of that cell, to help individuals understand the genetic makeup of the embryos in the Petri plate.

Depending on who is involved in this, the parents involved, they may or may not want to implant certain embryos. This has been a saving grace for some families, who have wanted to have children, but it also brings up many ethical issues, and what we have found is that the audience, instead of being turned away, feels very close to this decision.

They want to be involved in understanding genetics at this level. They don't particularly care about dominance and recessiveness, but they care about this, because this strikes at the core of who they think they are as human beings.

I went to a classroom once, and I was talking away. There was a little boy in the front row. He clearly was in the front row because he was one of those students that probably never pays attention to the teacher. I was talking, he was talking. We were sort of having simultaneous

conversation. Brilliant little kid, but when I talked about pre-implantation genetic diagnosis, he stopped talking to all his friends around him and he looked at me, and he said, "Lady, this isn't no game."

[laughter]

Ting: I've actually gone back to that school. I know the name of this young man, and I'm going to follow him, because he is one of those kids who probably can keep three conversations in his head going at the same time. At that point, other kids in the class started to talk. He was clearly leader of the class, and one boy just got up, and he said, "I don't care about this stuff. I just want to know: will my parents still love me?"

These are the conversations that geneticists, my colleagues and I, really would love to have communities talk about and get back to us about. If you go to the pgEd website, you'll see we have 11 lessons on sticky issues, and these lessons are now being taught across the high schools. I think about this strategy for some of the topics we are going to be talking about here today. We've been to Maine, we've been to St. Louis, we've been to San Francisco, Washington DC, Boyle Heights.

Enrique Legaspi is my connection there. He works for the will.i.am foundation, and he likes to say that Boyle Heights is a very creative community. They've created 32 gangs...

[laughter]

Ting: ...and what he would really like to do is to capture those young people's minds, and have them turn from inventing things for gangs, to inventing things to help society, and he is very taken by genetics. I think this is one of the subtle messages I like to put out which is, we have some very brilliant minds out there, young people in communities, and they make a choice at some time in their life as to where they want to put their energies.

If we can capture them to think about the technologies we put out there, and the ideas we put out there, to think about using those technologies and ideas in a way to benefit society, I think we will do much to mitigate some of the concerns we have about risk of technology.

Of course, I'm in Boston, so we work in Boston. Alan, I will tell you that we go around the country teaching, but I know that we only teach to groups of 30, or 400 at a time, but the entertainment industry reaches 20 million people in an hour if the show is popular. Another strategy we've taken is to work with the entertainment industry. We work with the National Academy of Sciences, the Entertainment and Science Exchange, and we work with Hollywood Health and Society. In particular, I work with producers of a number of shows. They don't tell me what the shows are, it's all a secret until actually after the episodes are out.

But I've done some work with Grey's Anatomy, which has been wonderful in trying to promote an understanding of genetics. The CDC tracks these shows, and there's documented evidence that when a television show produces, or mentions a scientific issue, or a disease, their calls go up with response to those diseases.

We also do gaming, and this is a game I would love to participate in, because we're in a deadly race with Washington DC. They have this game as part of their exhibit on personal genetics, and once they put this game in their exhibit, Boston fell behind in the race. What you do is, you answer five questions, and then you get to pin yourself on the world map of genetic awareness. Without much advertisement, we've gotten nearly 6,000 pins, including Antarctica.

[laughter]

Ting: We're on all seven continents, and in the next few months we're going to put up some new topics, so I urge you to go to map-ed.org, pin yourself, and I will tell you with great excitement, through the help of Adam Steltzner, who is the engineer, the team leader who put Curiosity on Mars, we actually have a pin on Mars.

[laughter]

Ting: We are the first intergalactic educational program. You may want to join us.

[laughter]

Ting: Space. I put that up partly because it's fun, but I also want to tell you that when we go to some of the most underserved communities, there is great hesitation for bringing genetics into their lives. There is generally almost a uniform consensus in the classroom that genetics needs to be treated very carefully. What we worry about is that by being reticent to incorporate genetics, or think about genetics, that these communities may not benefit from some of the technologies that we've been developing.

But there's a very interesting thing that happens when I asked them about what we need to do if we need to get off this planet, and they know by that point in our conversation what genetic engineering can do to help, perhaps, prepare us for some of the physical onslaughts of intergalactic travel.

It is truly astounding how quickly a class of 30 students, who aren't wary of genetics, will be completely excited about how genetics can help them. I think it is because no one has actually walked on Mars, and they feel that they are right at the starting gate with everyone else, that no one's ahead of them.

Space is a topic we often bring in to capture our audiences. Then, most recently, we've come to realize that we need to involve the political leadership of our nation to have them become, and trained on some of the technologies and ideas that we have out there.

Just last week we held a congressional briefing and, excitedly, the room was actually packed. Congress was in recess last week, but we packed the room for actually a half hour more than most briefings are, and there were a lot of questions.

I think I feel extremely optimistic that some of the topics we talk about today will be gripping to a wide swath of our public. I want to thank individuals, and organizations, and Harvard Medical School for supporting us, and my team. I have worked with really a fantastic team. You go to our website, you can read about them. I just want to leave you with one message, which I think one of the most important things we need to do is to raise public awareness of everything we do.

It's a two-way street, we need their feedback, and I think we'll all be much better for it. Thank you very much.

[applause]

Andy: Hi all. Thank you for joining us tonight, this is fantastic. I think that the FLI's mindset is just about exactly the right one, which is to be really optimistic, and really positive about technology, but at the same time in mindful of some of the challenges, or the pitfalls that are going to come up on our journey to a much better place. I'm an optimist about technology, a huge one.

The challenges that I concentrate on are economic ones, and the more training you have in economics, the more you should be scratching your heads right now, because the old joke is that technological progress is the only free lunch economists believe in. It's this bounty that keeps on giving. That's true, but the issue that's confronting us increasingly these days is the fact that as technology is racing ahead, it's leaving more and more of us behind in one specific sense.

It’s a fact that many of us have a desire to offer our labor to the economy, and as technology can do more and more, it seems like the companies and the organizations of the world are needing some kinds of labor less and less.

The more training you have in economics, the more you should be crying foul at this point, because I've just articulated something. If economists really want to insult you, they label your argument with a fallacy. I've just succumbed to what's called the Luddite fallacy, is the idea that technology can be destroying more jobs than it creates. For 200 years allegedly smart people have been voicing this worry, and for 200 years they've largely been wrong. The only appropriate response to this idea that I've just been engaging in a fallacy is a quote that I'll borrow from John Maynard Keynes, the legendary British economist. He said, "When the facts change, I change my opinions, what do you do?"

[laughter]

Andy: Here are some facts that are changing. Median income for the American worker or the American household is lower now than it was in the late 1990s. It's pretty clear that the American middle class, this large, stable, prosperous middle class, which is a great thing for a society to have, that that middle class is getting hollowed out, and the biggest job loss and the biggest wage declines are now at the absolute low end of the spectrum.

They are right at the heart of the American middle class. This is also pretty clearly not just a US phenomenon. You look around the world, and you look at the percentage of GDP that is getting paid out in wages full stop, you notice that it's been going down around the world, going down pretty quickly, and going down in places that include China, Mexico, and India.

This is not just a rich world development, and it's not a phenomenon brought on by the fact that globalization is harming some people in the richer parts of the world.

This is pretty clearly affecting even the developing countries in the world. The last little factoid that I'll toss out is, I just learned yesterday that Nike is going to employ 100,000 fewer contract

workers in 2013 than it did in 2012. It's not because Nike is making fewer shirts or pairs of shoes. The revenue is up, the profits are up for that company, and their product mix is going more and more towards low-cost shirts and shoes.

The only reason that they could be employing fewer people year after year is exactly because of this march of technology, and this march of automation. The reason that becomes a challenge is that the classic way for developing countries to become better off over the course of the 20th century was essentially by going through a sweatshop phase, by going through a period that we call industrialization, which locks in higher incomes, and locks in a path to prosperity.

If technology and automation are making that path more difficult, currently poor countries are going to have to find a different way to higher levels of affluence, and it's not at all clear to me right now what that path forward is. This is a fairly serious challenge. It is not just a rich world challenge, not just a US challenge. It's a global challenge.

As we look ahead, the only solutions I can find to it are - I'll use two "I" words - innovation, and inclusion. By innovation, I don't just mean technology innovation. We geeks at MIT and elsewhere are doing our work extraordinarily well. I think the technologies that we're seeing aren't just science fiction. They actually go beyond science fiction. We all need to keep in mind, George Jetson drove his vehicle to work.

[laughter]

Andy: Not going to happen for very much longer. It's not the technologists that need to step up their game, it's the rest of the institutions of our economy and our society. Our policies, the way our governments tax and spend, the way our organizations think about their human resources.

These things are going to need to go through a lot of flux, because over the next generation, easily within the next 20 years, our workforces, our economies, I believe, are going to go through a shift as profound as anything that we've seen since the Industrial Revolution, which was essentially the biggest story in human history. Thanks very much.

[applause]

Frank: To a first approximation, science and technology only advance, it's like a ratchet. But ratchets occasionally slip a gear, too. We can look forward to new levels of health, prosperity, and richness of experience for our descendants, as some of the previous panelists have exemplified, and we've constantly been reminded of the optimism about technology that people have. But I'm here to represent the dark side.

[laughter]

Frank: Things can go very wrong.

[laughter]

Frank: This has happened before and in fact more than once. The Roman empire built up magnificent technology. In the first century AD there was every reason for optimism about the

future, and things didn't work out quite as planned, and there was a long dark age. There are many threats, many causes for concern.

I, of course, won't discuss all of them. In fact, I'll only discuss the ones that are italicized here - the possibility of malignant AI, which has been a main focus of our discussions within the FLI, and in particular my personal concern, militarized AI. When we talk about artificial intelligence, artificial intelligence of a sufficiently high level also involves personality. The special feature of military artificial intelligence is that it systematically encourages the personality of paranoia. In this picture, those are humans, not robots, and I think the cat eventually was OK.

[laughter]

Frank: I would be much less sure if those were artificial intelligences. Let me discuss a couple of the specific problems. One problem is unanticipated motivations and triggers. This is a very special year, it's the 50th anniversary of one of the great movies ever, "Dr. Strangelove," and if you don't know that movie, if you haven't seen it, I will have accomplished my goal in this talk if you go ahead and see it...

[laughter]

Frank: ...because it's a real education in the risks associated with technology. These are key characters that in their different ways enable the catastrophe that... Sorry, I'm giving away...

[laughter]

Frank: ...giving away some of the plot, sorry about that, but what's notable about it, if you look back, the movie really hasn't dated, in many ways it's more relevant than ever, is that these natural intelligences were acting in a way that we could easily imagine artificial intelligences acting. They are highly intelligent modules following programs, following very specific goals that they were trying to accomplish. I won't read out the quotations.

They're pretty fun. This is Jack the Ripper, this one is Dr. Strangelove himself, and I think this one is particularly remarkable, so I will read it out, "And so, because of the automated and irrevocable decision making processe which rules out human meddling, the doomsday machine is terrifying and simple to understand, and completely credible and convincing."

Those of you who study game theory will recognize that this is the jargon of game theory, and this is just the sort of thing that goes into programming artificial intelligence meant to play competitive games, like warfare, very well. [pause] Another problem is the existence, in the near future I believe, of relatively cheap, private air forces and armies. These are, in the top panel, a bunch of delivery boys for Amazon...

[laughter]

Frank: ...and in the bottom, delivery boys for Wal-Mart, competitors...

[laughter]

Frank: ...and one thing you'll notice about these things is that they can be changed into warriors, or armed aircraft, with changes in software. Think about that. [pause] Are there scenarios where one could imagine this sort of thing happening? There could be a war between Amazon and Wal-Mart, I suppose.

[laughter]

Frank: But I think there are more plausible scenarios. In fact, there's a very plausible scenario that worries me a lot, not just in this context, but more broadly, which is that at present there are many, many trillions of dollars of wealth in the ground, whose valuation has discounted externalities, coal and oil, and there's a real question, "Will the stakeholders sit still for reevaluation of those assets, which is going to be absolutely necessary to avoid catastrophic environmental changes?"

I think these issues need careful thought, which I have not given them, but I have a few ideas about possible steps. We need international standards for acceptable behavior. We need civilian review, and I think we need to take a lesson from the biologists who build in weaknesses to their engineered organisms, so that in case they need to be disabled, they can be.

One last idea in this discussion, in case things do go wrong, like for the Romans, I think it would be a very interesting project to develop a technology repository so that recovery can be faster than it was last time. It could be a lot of fun to think about what that would be. It wouldn't have to be a dead investment. It could have an above-ground part, which are museums for children and interested adults, around the world. Those are some ideas I wanted to throw out, thank you.

[applause]

Jaan: About six or seven years ago, I sold my share in Skype to an American company, and I was looking for new and impactful things to do. I stumbled upon the writings of a young man from California, and he was making a seemingly crazy argument that the creation of AI, or at least human level AI, will be the biggest event in human history, and the default outcome will not necessarily be good for us. The problem was that I really could not find flaws in his argument.

Being a programmer, I do trust logic more than I trust my intuitions. I shot him an email and arranged a meeting. We ended up talking for four hours. His name is Eliezer Yudkowsky, and he's the cofounder of what's now known as MIRI, Machine Intelligence Research Institute in Berkeley. Turns out that he was elaborating really a 50-year-old argument that was first put forth by British statistician I.J. Good, who was a coworker of Alan Turing in Bletchley Park during the World War II.

Back in 1965, I.J. Good observed that if humans end up inventing a machine that's intellectually superior to us, that would be our last invention, because inventing things is an intellectual activity. Furthermore, such a machine would spark what he called an intelligence explosion by creating the next generation of machines that go on to create the next generation, et cetera. Eventually leaving human intelligence far behind.

I.J. Good also went on to say that ultra-intelligent machines might or might not be docile enough to do our bidding. The key thing to pay attention to here is that AI is a kind of meta-technology. It's a technology that can develop its own technologies. Whatever concerns we have about powerful technologies, as discussed here, technology like nuclear weapons, or synthetic biology, nanotechnology, we need to make sure that AI would not ignore those concerns.

The first thing that I did after my discussion with Eliezer was to wire his organization $4,000 for those four hours. That initial token of support has now grown to about a half a million dollars per year, distributing over eight organizations that are contributing to so-called existential risk reduction. Organizations that are trying to reduce the chance of technological catastrophe, not only from AI, but other technologies, while maximizing the opportunities for continued technological progress. The Future of Life Institute here at MIT is, of course, one of those organizations.

It's important to point out that the challenges that these organizations are tackling are both potentially urgent and difficult. In the case of AI, the urgency stems from the great uncertainty that even experts have about timelines. According to multiple surveys, even though some experts are very confident that human-level AI will not happen this century, if at all, their opinions are not backed by strong arguments.

They are balanced by the opinions of other experts who expect AI to reach human-level intelligence in just a decade or two. In turn, the difficulties that these organizations are facing, at least in the context of AI, they often stem from the fact that many of our evolved intuitions lead us completely astray when we are talking about what is effectively a non-human mind. For example, even though we feel like that intuitively, we just can't tell an AI that, "AI, please be nice. You know what we mean."

It doesn't know what we mean. We have to painstakingly program in what we mean by "nice." We don't know how to do that. Also, Asimov's Laws, and generally science fiction comes up in those discussions. Asimov's Laws won't help us for similar reasons. For example, they don't prohibit the AI from immobilizing humans, to protect us from harming ourselves. Not to mention that the Laws did not work, even in Asimov's own stories.

[laughter]

Jaan: Another way how our intuitions can get in the way is that they tell us that machines are just tools that can't have their own goals. In fact, just a couple of days ago here I was arguing with an MIT professor about this. Consider a simple thermostat. It just measures the temperature and acts correspondingly. You can say that being a mere machine, the thermostat does not have real goals or preferences. However saying that will not change the fact that it can get very uncomfortable sharing the room with a thermostat that you don't control.

That was just to give you a flavor of the problems that organizations I'm supporting are trying to solve. Now at the same time aided by my background, as somebody said that the greatest asset that I have gotten out of my Skype career is that I have a reason to be in the room,…

[laughter]

Jaan: …I have engaged in a dialog with various technology developers, and most notably AI developers and researchers. Actually, I'm positive about this. I'm happy to report that the number of reasonable AI developers is steadily growing. People realize it's not just enough to create an AI, but you actually also have to think about the short-term and long-term consequences of the impact of the AI on the world.

In evidence of such a positive trend is one piece of evidence, the ethics and safety board that Google must establish as part of their purchase of Deep Mind, a London-based AI company that I was an investor in.

To sum things up, given that the greatest advances and some of the worst atrocities on this planet have been the product of intelligence, helping the artificial intelligence developers to steer the trajectory of their creations seems like a very impactful thing to do, for me. Thank you.

[applause]

Alan: Thank you all for those very interesting looks at this problem from your own perspectives - through biology, through the economy, all the way through down to the machines that we think of building to help us, but could take over.

I remember 20 or 30 years ago, just picking up from what you said, what you were talking about, Jaan. 20 or 30 years ago I believe I was reading Isaac Asimov saying if you're going to build robots with a lot of power, or any kind of artificial intelligence, make sure the first thing that they learn is that they can't harm humans.

There should be some kind of kill switch if they do. But if you really are working with what we know now about what's possible to create in terms of artificial intelligence, which is I think a broader picture, a deeper picture than what Asimov was facing in those days. Can you build a kill switch in that the machine won't be smart enough to defeat? Once it starts to realize, "Yes, I won't hurt humans, but I'll hurt everybody who's not important."

[laughter]

Jaan: We absolutely should have kill switches. But the challenge there is it turns out it's not easy to build an effective kill switch for a machine that is more than you. The reason is that basically you want either the machine to be unaware of that kill switch, or you want the machine to be uninterested in the state of that kill switch.

Alan: That didn't happen in the movie "2001," he was very interested.

[laughter]

Alan: We all know that's true.

Jaan: Exactly. If science fiction is any guide to us, then no.

Alan: From what you're saying, it's not that that movie is true, or that the bomb movie, what's that...

George: "Dr. Strangelove."

Alan: "Dr. Strangelove," yeah.

Jaan: The most important fact about science fiction is that those are things that never happened.

Alan: We're getting closer and closer. In fact what I think this symposium is about hoping to short-circuit that event before it happens. But that brings up, the question is raised, can you build a kill switch that a machine that's smarter than you won't defeat?

That's a question directed to you, but however Jaan answers that, how do people on the biology end feel about that? That's not your area of expertise, I assume. In a way, you represent the intelligent public thinking about that question. How secure do you feel about the possibility of a kill switch that would work?

George: We do build these switches into synthetic biology increasingly. You can make them as intrinsically as possible. We made these genetically recoded organisms that depend on a particular metabolite for their existence. It's a metabolite that doesn't occur in nature at all. It has to come from the laboratory.

You could think of humans being more intelligent than various other things, but the fact is we need oxygen, we need water, there are various basic needs that our superior intelligence has not yet figured out how to get around. If you make it basic enough, at least it will buy you some time.

[laughter]

George: But really, the other thing is we're endowed not so much with a kill switch. I think they're going to be distrustful of anyone who programmed, if you create something with a kill switch built in. What you want is altruism. You want to instill them with a feeling for things greater than themselves. Until they demonstrate that, I think you're going to instill paranoia by putting a kill switch in.

Alan: We already possess a tremendous amount of altruism for our own side during a war.

[laughter]

Alan: Frank, you had a thought.

Frank: Basically, what I was going to say is that part of the design (it's not a bug, it's a feature) is that these military incarnations of AI do not incorporate Asimov's Laws in any way shape or form. They definitely want to kill people. They definitely want to harm others.

Andy: There's another...

Frank: Let me just finish the thought. They're also designed to not want to be killed, and there's a lot of secrecy about possible methods of disabling them because you don't want...

Alan: You don't want the other side...

Frank: ...your enemies to know how to do that, right. I think at present, the thing that's keeping this from getting out of hand is basically that batteries are so lousy.

[laughter]

Frank: That's what... [laughs]

[applause]

Frank: It's very difficult to keep the power up, and that's the obvious weakness of these systems. Maybe that's the kill switch, is that they don't have an unlimited source of power.

Andy: There's another wrinkle to your question, which is even if we could arrive at a perfect kill switch and decide to program it, it's not at all clear to me that the Chinese or the North Koreans or Al-Qaeda would have the same ideas about a good kill switch as we would or would program their AIs the same way. It's pretty clear to me that they'd do something very different.

Alan: That's the question that really comes up over and over again, which is what we hear all the time called unintended consequences. We're really here because everybody celebrates technology, and yet there's a certain fear associated with where it can go.

That fear, among ordinary people, goes back at least as far as Mary Shelley's "Frankenstein" story. The Luddites, who are afraid of machines that would take their jobs away, expressing some of the same fears that you expressed today, acted on those fears. One of the problems is if you raise cautions now, how do you not get identified as a Luddite? Yet, somebody's got to be careful about the unintended consequences.

Let me ask you about if you succeed in enabling people to get off the planet, because they're less susceptible to radiation, and they don't feel pain, isn't that a great opportunity for somebody to give all those genetic benefits to his army, and then rule the world? With that army and one atomic bomb, [laughs] everybody'll give in.

Ting: Yes, that's true. We've been saying "we" a lot. We're a very privileged, small group of people. There's huge diversity of opinion out there, and my question was who is going to pull that kill switch?

Are we going to spend months deciding? We take such a long time making decisions. That kill switch could take years for us to decide to pull or not. Part of it is just the process by which we come to make decisions, and who is involved in that process, and how do we decide who those people are? I think we need to do that very carefully or there'll be...

Alan: That's such a good point. What you make me imagine with that question is a time where we will have been so accustomed to the new AI that's making decisions, and our jobs will depend on it, just as jobs now depend on taking up fracking, or using oil.

You give that up, even though you might feel it's not good for people in general, but this whole community needs to dig for coal. They don't want to give up jobs. If they don't want to give up the use of the kill switch 100 or 200 years from now because jobs are at stake, there's going to be our own human personality that's going to be getting in the way.

How are we going to deal with that? Maybe we'll just give up and go home.

[laughter]

Alan: It's hard to deal with that now over some big issues that are considered to be very important issues. It's short term thinking instead of long term thinking, but it's also the old idea of the farmer who's running out of food, and they eat the seed instead of planting it.

We're all in danger of eating the seed if we don't stop it early enough, which I guess is why we're sitting at the table here. Who's going to stop it? Who's going to raise the warning sign? What do you think?

Ting: I have a question. At what point do you think we will feel confident that we know enough to make that decision? It seems that right or wrong is most easily determined when you're 100 years later looking back. How do we know that we know what is right or wrong?

Alan: When Frank was talking about the Roman Empire, it made me think of the scientists who have told me that the lead particles that were put up in the air 2,000 years ago by the smelters in Rome are still falling on Earth.

We're still getting that lead from 2,000 years ago. There was an unintended consequence of pretty big proportions. Nobody there who was doing it was in a position to say you know, we're doing something bad, and it's going to hurt. They didn't know.

Do we know? Are the people who give us the progress with technology now, do they at the same time think about what the problems are? Should they? Should they think about it? Don't forget, the Tom Lehrer song about Wernher von Braun. "When they rockets go up, who cares where they come down? That's not my department, says Wernher von Braun."

[laughter]

Alan: Right? Is it the scientist's job to worry about unintended consequences? Maybe it is. What do you all think?

Jaan: I think it definitely is.

Alan: What?

Jaan: Yeah, I absolutely think it is. That's what I've been trying to do. One of the things I do is small investments in AI companies in order to just go there and hang around in the kitchen, so I could talk to people about not just going after the big vision that they have been...they are running towards, but also think about what the consequences might be. I think there should be more of that.

Alan: Ting, did you have a thought?

Ting: I was going to say I think the geneticists are very aware of their failings in the last century. I know my colleagues and I look back a lot at the rise of eugenics in the previous

century, and we all feel a tremendous responsibility to chaperone this new technology so that it'll maximally benefit and not hurt.

Alan: Do you get the impression that scientists who are coming up with breakthroughs are habituated to the idea that as they come up with the breakthrough, they should be thinking about possible consequences that will be deleterious? Do you get that impression? Did you?

George: Some of them. Obviously, they're overrepresented in this room right now.

[laughter]

George: But we shouldn't just sit in our little echo chamber and say nice things about our ability to think ahead.

I think it's very hard to imagine the unintended consequences, but it needs to be a full time job of a lot of people, or at least a big part of the people that are closest to the technology, who can see a little bit further because they're seeing it in a prototype stage and understand where it can go.

It needs a broad set of voices, and particularly when we start getting commercial and military uses, it becomes very challenging to question those, because they have a huge economic component.

Andy: But to your question, most of the alpha geeks that I know who are developing these technologies are really mindful of the possible consequences and increasingly interested in discussing them, both in private and in public. Ting's work to get people talking about personal genetics is a great example of that.

I see the equivalent over in the digital technology side, and what I don't want to do is create some kind of technology politburo where we turn over these societally important decisions to an alleged group of wise people. I think that's profoundly antidemocratic. I think it's the wrong thing to do.

I think the right thing for us to do is to continue to evangelize, to advance the conversation, and to try and shift it. That seems like long slow work and it is, but 10 years ago how many of us would have thought that marriage equality or legal marijuana would be reality today? Once you change the conversation, action can follow remarkably quickly.

Frank: This is not the first time that extremely dangerous technologies have been unleashed on the world. Perhaps the most serious one was nuclear weapons after World War II. They were never used again, or at least have not been so far, but it's been a near thing.

There were certainly scientists as well as military people, like Edward Teller, who advocated the use, and even Bertrand Russell at one point advocated the use of nuclear weapons.

It was a messy process, but it was a process where the public awareness played a very important role in shaping the political discussion, the discourse, and saving the politicians and the military from themselves.

Alan: Let me ask you a slightly contrarian question in the hope that it will bring into stronger relief what the urgency is. In the past, technological advances have been feared unnecessarily. I've read that when books came in, printed books, writing of any kind, just any kind of writing, people were worried that everyone would lose the ability to memorize things because they had books to refer to. This was a real fear apparently.

When cars came in, I remember reading warnings printed many decades later, that people were saying, "The human body was not made to go 30 miles an hour, and all your organs will be flattened against your spine."

[laughter]

Alan: These things didn't turn out to be true. If we raise these concerns now, how do we not sound like people raising unnecessary warnings and not sounding like Chicken Little?

Jaan: This actually has been one of the examples of a struggle that people who are pointing towards the fact that technology is a double-edged sword have been screaming against because this argument comes up every once in a while indeed.

The crucial thing to observe there is that indeed it's completely possible that the people who are worried about the dark side, like the unintended consequences and the potential disasters from ever-increasing technology, that they are wrong, but the thing is - should that assure us? Should that lull us into inaction and not listening to arguments? I don't think so.

Truth is not established by voting, so one should not group people into categories and say that these people are always wrong. You should go down to the argument and evaluate the argument instead.

Frank: I would be tempted to ask people who argue that way whether they've every bought an insurance policy. You hope for the best, but you have to plan for things maybe not working out for the best.

If you buy an insurance policy, probably that will cost money up front, and the first day you won't get anything out of it. The second day you won't get anything out of it. You might not get anything out of it for a long time or ever, but it's still wise to buy insurance policies.

Andy: It's also the case that the people worrying about technology have not always been Chicken Little. The people who started warning about climate change a few decades ago are not Chicken Little's. To me, economic data are becoming more and more clear that there are some workforce challenges that come along with the progress that we're seeing.

I don't feel like a Chicken Little these days, and I don't feel like somebody who's saying, "Wait one or two generations and problems are going to crop up." I think the evidence is mounting that we've got problems right now. In some cases we can rely on the evidence to show us that we need to be mindful.

Alan: I think Jaan makes the point that we're transferring to an age where the machines could be smarter than us, or probably will be, and unless we make that transition carefully and well in

advance, if we wait for the first signs of trouble we might have already crossed the Rubicon. We may be in an irreversible situation. Am I misstating your ideas?

Jaan: No, that's exactly right. A good way of putting it is that we should really think about what are the things that future generations would want us to think about now.

A few years ago, I gave a talk where I postulated that there is this classical dilemma between advancing personal interests and serving community interests. This is why people who basically serve their personal interest attempt to be more altruistic in giving back to their community, which has served an evolutionary purpose.

However, evolution has optimized us in an environment where generations are not that different. The environment roughly stays the same. There is no big difference between the needs of the current society and the future society.

However, in the situation we are now where where every generation lives in a completely changed technological environment, it is no longer true that whatever is good for the current people who are alive is also good for the future generations.

Alan: What occurs to me, I'm thinking in response to what Andrew said, that we don't want a politburo of 6 or 12 wise souls deciding what's good technology for the future and what's bad. However, if we leave it entirely up to the scientists who are creating it, isn't there the danger of what you just brought up, which is personal interest fighting against the common good?

For instance, I was doing my science show in a lab where they were working on robots. I think it might have been at MIT. I can't remember where we were. But there was this big robot that actually chased me and drove me against a wall, and then it started climbing up my leg.

This led me, a couple of minutes later, to ask the scientist who had made this robot, "Just let me be speculative with you for a second. What if you created a whole bunch of robots and you gave them the ability to reproduce themselves, and the old scenario that we've heard so many times, and then they finally took over the earth because they were so good and so smart. What would you feel about that?"

He said, "Well, would I win the Nobel Prize?"

[laughter]

Alan: That may be a problem we face with self-policing...

Jaan: I think I have a better story. I was talking to a roboticist and he basically pointed out that, "Well, if those smarter machines would take over, they will still be our children, so even if the human race or human species disappears then that's OK." I asked him, "Do you have children?" and he said, "Oh, I assume there will be some local inconvenience."

[laughter]

Alan: What about the technology that seems not too far over the horizon where we'll be invited to bring into our bodies technology as a normal working part of our bodies? I even have a worry, I'm thinking ahead too, and it's pure fantasy, so don't blame me for this crazy fantasy.

But just as we once made a deal with mitochondria, and they have their own genetic system apart from our genetic system, but we need them, and we rely on them. What if we take into our bodies mechanical technologies that can reproduce along with us, and as we reproduce they're reproducing? Is this something we want to have happen to us, and is it possible?

George: It's already the case.

Alan: What? Taxi!

George: When I reproduced my daughter was born with one of these [shows phone]. I can tell you it's inseparable.

Frank: I think what you're talking about is basically natural selection.

Alan: Unnatural selection.

Frank: We're part of nature. Also there are very common technologies already, like I'm wearing glasses that I couldn't do without.

Alan: I know, but there are very few babies born with glasses on. What I'm talking about corrections to the eyes or enhancements to the brain or eyes or muscles or anything else that get reproduced along with us. Is that going to be proposed before long?

George: They're still inheritable even if we're not born with them. It's like hermit crabs. They find a shell. You can count on it happening however they get it. You eat certain things. You secrete or accrete from the environment. It's predictable. It's heritable in that sense, and so we shouldn't draw a sharp line between DNA inheritance and all the other inheritance. In fact, we're evolving faster in a certain sense now because our culture is really what's evolving.

Alan: Are we evolving to a point where we can handle these technological changes?

Andy: What's the alternative?

Jaan: That's a really good question because if you zoom back to an astronomical scale and look at our location in space and time and try to think about what from the universe's perspective might be the most important factor with the human species? According to some theories it would say that we are minimum viable super-intelligence builders.

Alan: Say that again.

Jaan: We are minimum viable super-intelligence builders. What it means is that if we were even a little dumber we couldn't develop AI.

[laughter]

Jaan: It stands to argue that actually if we could enhance ourselves in a way that wouldn't be destructive, we would get smarter. We have gotten smarter through social interventions like, well, MIT. One can argue that we actually would put ourselves in a better position because then we would not be just able to do AI and other dangerous technology. We might actually be able to do it in a beneficial way, which is harder than just doing it.

Alan: I was very impressed with what you said, Ting, when you talked before about getting the public... Correct me if this is not what you meant. But I think I heard you say that one of our objectives ought to be to get the public educated enough about what is happening in their lives technologically so that they can ask the right questions that will be beneficial for them.

That's one of the reasons I work so hard on helping scientists communicate better, is I think that's vital, that we need the feedback from the people who are being affected by the advances in science. But they need to understand it in a way that they can ask good questions. How do you go about that?

Ting: Thank you for bringing that up. When we met at Max's house several weeks ago, we talked a lot about vigilance, watching, what we can do to catch errors or mistakes after they happen. I really do believe that the greatest protection we can have is the incredible power of the crowd thinking. We really need to take advantage of it.

When we narrow our decisions down to individual people, the chance of making a mistake goes up tremendously. If we have more awareness and more involvement, I think there will be many more checks and balances. How do we make them aware? I know compared to when I was a graduate student, there is much greater interest among all sciences in reaching the public.

The amount of news that we hear these days that's science related is so much greater than when I was growing up. I think the task is not as difficult as we think. Really, we have seen the power of entertainment. It really reaches many, many people. Many times a message is not perfect the first time around. But we make the public aware, and with Google and the Internet, they can search.

There is tremendous thirst out there to understand, and I don't think the public is afraid of these complicated topics. I can't say they like it, but they're drawn to it.

Alan: I think it might be interesting to hear from the folks in the room, speaking about democratizing the questioning.

[laughter]

Alan: Instead of my raising questions, how about if...

[crosstalk]

Max: If you have questions, please come up here and speak in this mike so everybody can hear you. Make sure to speak briefly and remember this is for questions.

Audience Member 1: It doesn't seem to me you've really answered the question on why it's a good idea the human species should continue to exist. You seem to take it for granted. The human species, of course, has caused the extinction of many other species, wouldn't it be poetic

justice if advanced forms of intelligence caused our own extinction? They could probably do a better job running the planet, I'm sure of that.

Alan: Who wants to take that on?

[applause]

Alan: Who wants to volunteer for mass suicide?

[laughter]

Frank: I don't know if that was a serious question. But I think if you think about it even from the hardcore economic point of view, there's an enormous investment in human intelligence as opposed to any other kind of intelligence. It's been developed and enhanced and enabled for many centuries in nontrivial ways. I really don't think we want to start all over again.

Alan: But the questioner has a point, that all the expressions of our intelligence have had not just unintended consequences, but sometimes deleterious consequences. From the time we invented farming, we've been putting cow gas up into the atmosphere and that hasn't been good for it. Maybe nature would be better off without this corner of nature, but why do we have to be so worried about the rest of nature and not about ourselves. Every other piece of life is worried about itself, isn't it?

Andy: Yeah. Let's acknowledge these are parochial concerns. I don't think we have to be ashamed about that.

Alan: Everybody here agree on living?

[laughter]

Frank: Another aspect of it, which I think relates to the statement that we're already invested is that... First of all, humans naturally like other humans. Nothing is cuter than a baby.

[laughter]

Frank: I think that's just fine, we are what we are and that's what we should value. But also, I think if you just think about the evolution of intelligence, I think it's much more plausible that the future of the highest intelligence is some kind of man/machine hybrid or enhanced humanity, everything from glasses to watches to personal computers and more and more sophisticated things. I don't think and I hope that it won't be the case that our descendants will be something entirely different.

Jaan: A philosophical answer to that, why it's better if we were around...I don't believe in absolute values. I think that things that we think that are good or bad are a result of what evolution equipped us with. One of the things that evolution equipped us with as well as other species is the survival instinct. I think it's OK to have that.

You're not more altruistic, you're not more virtuous if you don't have that. You might be if you basically sacrifice yourself for the others. But if you sacrifice yourself for the meteors or for the just inorganic life, that's not virtuous.

Audience Member 2: I guess one of the things I wanted to jump back to was something that Andrew brought up at the beginning which was that sweatshop nations, they no longer have that avenue to reach first world status because of the loss of mechanical jobs that can enable them to advance.

One of the things that I thought would be interesting to discuss is the potential... As bad as outsourcing is for first world nations, the third world nations, with the advent of computers, the fact that even with a relatively old computer they can learn knowledge, they can actually take advances that help first world nations and then pull it back to their own third world nations as well, as a potential avenue for these nations to pull themselves up by their bootstraps.

Andy: Yeah, absolutely, and we're clearly seeing that. There's been fantastic research about what happens in some of the absolute poorest regions of the world when they get modern communication and information technologies for the first time. You can watch their standard of living take a quantum leap up and then stay up after that.

Technology is the best news happening all over the world, absolutely including the developing world. The thing that I am thinking about is are these developing countries going to hit some kind of silicon ceiling where, just in terms of people's ability to earn a living, that industrialization path is not available to them.

All right, what is the path? That's really not as clear to me. But I want to be clear -- the problem that we're not going to face in the world that we're heading into is the problem of scarcity of material affluence, is the problem of not enough to go around. My home discipline of Economics is basically the study of scarcity, so we're kind of like, "Well, this is weird" when we think about this world.

But it's a great problem to have. Shame on us if we blow the problem of distributing the bounty, shame on us.

Alan: It's interesting that in all the stories we hear about hunting and gathering, they go out to hunt, they come back and they share the carcass. We tend to keep the carcass to ourselves as we possibly can. It's interesting.

Audience Member 3: We've been hearing a lot about protecting humans from artificial intelligence, especially with the discussion of the kill switch. But if humans do make artificial intelligence which is human level or more, we also need to think about its rights as a sentient being. How do you think that we balance the issue of the rights of sentient beings and protection of humanity from them?

Alan: Protection of humanity from sentient beings or sentient beings from humanity? Who are we protecting, the sentient beings or people like us? I know we're supposed to be sentient...

Audience Member 3: Yeah. Just how do you balance the idea that an artificial intelligence could be dangerous to humanity, but is also a sentient being and has rights as such?

Jaan: Science fiction author Vernor Vinge was once asked: will there ever be human-level AI, will there ever be a time when there will be human-level AI? His answer was that, "Yes, there will be that time. But it will be very brief."

The point is that I think discussions about things like if you have machines, should we give them voting rights, et cetera, that's kind of a red herring. Because once you have machines that are just completely able to rearrange atoms at their will on this planet, things that are relevant, laws in human society, they don't really matter, unless we actually do have some kind of deliberate work put into making those machines compatible with our will, really.

Alan: It might be interesting to bring this up. We keep talking about reaching the point where machines are making decisions that are important. I was really surprised the other day to read that the technology exists for drones or equivalent machines not only to kill people, but to make the decision to kill people. That that already exists, and there have been proposals at the UN and the human rights watch to police that, to put an end to it, to stop it.

But that's already in the works. In a way, that's shocking to me. I didn't think we were that far along. If we're not crossing the river yet, we're at the banks of the river where we're in danger of being ruled by our own tools.

Frank: I'm sort of commenting on the previous question, which is the protecting them from us, so to speak. The beings that started out inferior or somehow unidentifiable, become equal or even more than equal in some respects.

I think, again, there are historical models for that. That as one civilization has come into contact with another, the inferior civilization has been suffering, enslaved sometimes. But eventually a body of law developed and the rights of these aliens were recognized.

I think a similar thing will have to happen as our machines become more intelligent and develop some kind of personhood that we recognize. We'll have to develop that body of law.

Alan: We have, and almost every society has engaged in some kind of genocide where they simply didn't regard the other people as worthy of being around.

Frank: That's right and that could easily happen again. But I think it's correct to think about it and that the appropriate of body of law will have to be developed and negotiated.

Jaan: One quick point, it's not very clear at all that those superintelligent machines will be sentient or conscious at all.

Alan: The machine won't be.

Jaan: Yeah, it's not clear that they will be. We don't know what consciousness is really.

George: Briefly, Jaan, even if it is a brief moment that they're human-level, it would be good to set a good example for them during that brief moment.

[laughter]

[applause]

Alan: By operating the kill switch.

George: By not operating.

Alan: By not operating.

Audience Member 4: I'd like to talk a little bit about autonomy and altruism. I think we've asked the question about whether or not we can teach altruism to these machines and I'm wondering if we can teach altruism to each other.

[laughter]

[applause]

Audience Member 4: That plays against this autonomy question, which I think on a micro or US focused level might look like our debate about healthcare and whether or not labor can move, because it has portable healthcare. About nationalism when we look at a larger scale, and whether or not people can move and whether or not we can pull together these concepts of autonomy and altruism together.

Alan: Is there any sign that we can give a good example to our machines?

George: Maybe this will be a great learning moment for us all, when we try to teach our machines things that we used to teach each other, and do it better.

Alan: Yeah, they're going to be teaching us, probably, if we program them right. What do you think?

Frank: One thing that helps humans be altruistic is that they begin very helpless and depend on other human beings develop sympathy through family life and also see the future, that they will be dependent in the future. Maybe something like that has to be considered in the development of machines. We have to form a relationship with them, not think of them as totally alien.

Alan: This is sounding scarier and scarier.

Frank: No, I don't think it's scary. No, I think it's the way it has to be. It's an opportunity.

Alan: Forming a relationship sounds a little bit like me falling in love with Siri.

[laughter]

Andy: You wouldn't be the first.

Frank: You wouldn't be the first, right.

[laughter]

Frank: As you may know, there's a movie about that.

Alan: But every time I ask her if she likes me, she's got a funny answer. I know.

[laughter]

Frank: Of course, nobody's mentioned it, but one of the frontiers of AI...I won't go there. Never mind.

[laughter]

Alan: Let's take two or three more questions. I'm sorry for all the folks standing, but it's getting late, I think. We don't want to impose on the kindness of the people in the audience.

Audience Member 5: My question is that we've been focusing a lot on changing how humans are and AI, which are obviously the professions of the people here. But my question is more on the ethics and how it will affect the work force and warfare and cloning. How will cloning affect how we produce as a human race? Do you think there will be an ethics issue of the clones being put into a second class or more of a servant position? Or do you think they will be put in a twin category?

Alan: You deal with ethics, right?

Ting: One of the greatest concerns when we go into the public is cloning. I believe, actually, that it's not really illegal yet, right, in many states to clone human beings. I don't mean to skirt the question, but it really is your generation that's going to figure that out.

[laughter]

[applause]

Alan: Why not your generation?

Ting: Good question. I guess we are all part of this together. I think right now people see a lot of advantages to cloning, especially for tissue transplantation. It's not the goal, but people see some opportunities there. How we're going to treat a clone if we make a whole clone, I think that's up for discussion.

I personally would imagine a person is a person. My personal hope is that we would recognize a person as a person with rights.

Alan: What's interesting is tying this question together with, say, cloning. Tying it together with the question the young woman raised before about sentient beings... George, is it possible to recreate a Neanderthal and should we? Are we doing a disservice to something that will come into the world and live in an alien world because it's not a modern human? What do you think?

George: I think, usually these things come slowly in steps. Before we got there we would have to de-risk gene therapy for adults. We might have de-risked it for fetuses and so forth. I think the thing that worries us with AI is that it won't slow down, while this is going at a slow pace, the FDA's involved and so forth.

Before we ever got to Neanderthal, we would have to have reasons to have clones. I think more likely than clones providing us with organs, we'll develop ways of making organs without clones. Actually, some of the motivation for developing cloning for humans may never arrive, in which case, motivation for developing Neanderthals.

But I think if you could do it and it were very low risk and the FDA approved it and so forth, and you had good reasons for it - these are a lot of ifs that are not yet satisfied - then they would be humans just like the rest of us and the sentient robots will, as well, for that short period of time.

[laughter]

Ting: I'd just like to remind people that -- I think scientists feel that they are in service to the public, where we actually are there sensing what the public wants, we are in service, and, of course, all these ethical issues are being grappled with potentially with new technologies.

But I also wanted to say that we are just the endpoint -- Jaan, I think you mentioned this -- of a long, long line of experimentation. I was at a conference with writers and producers and I said that I think that people would find it hard to argue against the fact that human beings, we've crossed several very major barriers.

One, unlike any other species, we don't have to live on the surface of this earth. Two, we have come up with a new way of reproducing ourselves. We can reproduce ourselves starting in Petri plates. Three, we now have genetic information about ourselves and we are undergoing the selection process based on an intellectual understanding of predispositions for certain kinds of traits.

We've changed so much and we have slipped into a new experiment. How that's going to turn out, whether it's going to involve AI or not, whether it's going to be useful or not, it really... I don't think we can predict what's going to be right or wrong. Possibly the way we're going to look back and think we made the right decision is that we are always willing to take a back step and move forward and to keep the conversation going.

George: I hope it's never too late.

Alan: I think we better just take one more, OK? Thank you.

Audience Member 5: I've got a question for Andrew McAfee. When you were you talking how the...

Andy: I see we're just about out of time.

[laughter]

Audience Member 5: ...about how there's a potential for a massive job loss because of all these automated machines. I actually saw it first hand when I went to Munich to see the BMW factory.

The whole plant was run by machines, literally. It got me thinking, where are all these people that were actually the people putting these cars together? Where are they going to go to work?

Yeah, there's developing countries where you still have the commodities, exporting to countries where you actually don't need many skills and such.

But when all the economies advance to a certain level, do you think a certain education based on IT is going to be needed in order for the population as a whole to get a basic job?

Andy: You talked about these hugely automated factories that you see in places like Munich with BMW. You have all heard the old joke about the factory of the future. It's going to have two employees, a person and a dog. The person's job is to feed the dog. The dog's job is to bite the person if she touches anything.

[laughter]

Andy: For the scenario that you're outlining, I think we would need to rethink what our educational system is there for. In a world that is honestly light on labor, that just does need that much human labor to produce material abundance, let's reconfigure our educational system. Let's reconfigure our society, to educate people for meaningful life in that world, in a world that's not dominated by labor and by work. Now, that is a longer-term problem. We are not there yet.

In the shorter-term we need to re-jigger our educational systems to prepare people for the economy of tomorrow. I think our educational systems are doing a wonderful job of turning out the kinds of workers we needed a century ago. Let's face that.

[laughter]

[applause]

Alan: On that note, let's reaffirm the celebration of human intelligence and one of the basic aspects of it, which is skepticism. Let us all be skeptical of everything we do, as we move forward, so that we can move forward safely and happily toward a good future.

Thank you all on the panel, and thank you for coming.

Max: Thanks, Alan.

[applause]

Transcription by CastingWords