

**New America
Resilience Audio Interview**

FEI FEI LI

AI and the future of work, policy and geopolitics.

Anne-Marie Slaughter: I want to start by asking you about human-centered AI, or human-centered artificial intelligence. You founded Stanford's Human-Centered AI Institute in 2016, and that's not a term a lot of people have heard of. In fact, when people think artificial intelligence they think machine learning, they think something that maybe is simulating humans, but not something that would be human-centered. How do you think about that?

Fei-Fei Li: Yeah, thank you for asking that question. First, a small clarification: the Institute officially launched in the spring of 2019, but the work the ground work began in early 2017 at Stanford and a lot of colleagues were involved. It's just precisely as you said, when you think about AI as a discipline, and now as a widespread tool and application, it's about machines, it's about computers. As an AI technologist for the past 20 years, most of my work is on the technical side. But around 2016 and 2017, our society woke up to this maturing of this technology that is starting to impact real lives, such as the products we use on the internet, on our devices, the potential applications in health care, in self-driving cars, in a lot of business decision-making. So as this is turning real and touching human lives, I began thinking about the meaning of this technology, and more importantly, how do we harness the future of this technology so it can serve humanity in a positive and benevolent way? This is when we began to formulate this new framework of AI, which we call human centered AI, and it's centered around three principles. It's the three founding principles of Stanford's Human-Centered AI Institute, but I advocate this to all approaches or efforts of AI. The first founding principle is centered around understanding, predicting, and guiding AI's human and societal impact. It's fundamentally recognizing that this is no longer a niche, not only just a technical and computer science field. Its impact is so profound and prevalent that we need to turn this field into an interdisciplinary study, and put just as much effort and resources into the understanding of the human impact as we do into improving the technology itself.

AMS: So in other words, it's too important and too central to leave only to the computer scientists or the AI specialists.

Fei-Fei: Absolutely. I say that as an AI technologist, and I totally agree with you. We need to welcome economists, policy thinkers, legal scholars, historians, political scientists, philosophers, and all other disciplines to work together and to understand that some of the implications are huge. AI touches on the future of work, it touches on policy dealing with machine learning fairness, on ethical codes, and geopolitics. It's very prevalent. So the first founding principle is

recognizing and turning this field into an interdisciplinary field in order to really understand, predict, and guide human impact. The second founding principle of AI is an effort to change a verb for AI. Let me quiz you Anne-Marie. What is the verb you think is associated with AI today in the public mind?

AMS: That's so interesting. I would have said for me it's "learning," but I'm not sure that's what the public thinks. The public probably thinks "computing."

Fei-Fei: Right. Actually what I'm getting at is: What does the public think about the consequence of AI in terms of a definition.

AMS: So the consequences would be "displacing," as in displacing humans.

Fei-Fei: Yes. Displacing humans or replacing them. We want to change that to "augment and enhance" because this technology has so much more potential to be collaborative, interactive, and ultimately augmentative to human capabilities, instead of replacing or displacing. So this is why at Stanford and in my other efforts involving AI we're really promoting and investing in research and technology focusing on human-centered augmentative AI technology applications such as healthcare, education, sustainability, manufacturing. These are all amazing, and there are amazing opportunities here to really use this technology to the benefit of human workers, human consumers, and individuals and communities.

AMS: This is a view that is deeply humanist in that it rejects what for many of us are science fiction nightmares of computers replacing humans; of cyborgs or some kind of integrated human computer machine circuitry that paves over the human world with a world of perfectly rational computational machines. At least that's the nightmare I think of.

Fei-Fei: Yes. Whether it's sci-fi or Hollywood, plenty of people have this idea. I don't want to be naive here. Like we said in the first principle, the shifting of jobs, the changes of culture *will* happen because of technological changes, and humanity and civilization has seen this through over and over again, so it will have that effect. I think what's really important is that we spend far more resources and thoughts in creating the technology that can augment and collaborate with humans, because there are many scenarios that we need machines to help us. For example, in disaster relief situations, machines can get to places faster without the consequences of danger for human rescuers. They can map out the situation faster in healthcare, which has been part of the research I've been doing for the past eight years. We see how American clinicians are overworked and over-fatigued and experience information overload. And this is where machines can come to help because the bottom line is, we need to care for our patients. We want all the help we can get to give us back the time and effort to do the human caring. So yes, that is the second principle: it is about promoting human-centered AI technology to augment humans. The third founding principle is more foundational to the science. If we believe we can use machines to help humans to assist and collaborate, then today's AI technology is not sufficient. There's a

lot of hype, but truthfully this technology is still very narrow and it's very limited in its ability. At Stanford we see this amazing convergence of two of the most interesting sciences coming out of 21st century: one is brain size and the other is artificial intelligence. So we are promoting that interdisciplinary approach to combine the understanding of human neuroscience, cognition, and psychology into the next generation AI technology so it can be more empathetic, it can be more contextually aware, it can be more multi-sensory. There are many dimensions that are missing in today's technology and we have to get there so that our first two goals can be realized.

AMS: So would it be right to say that if the aim is to have AI that augments humans rather than replaces or displaces humans, then we need to understand more about how humans actually function, human intelligence and human capabilities, in order to mimic that with machines. Is that an accurate statement?

Fei-Fei: Almost. I agree with the first half of the statement. But I would say that this is an open area of debate and research. Personally, I don't think it's necessarily mimicking, it could be understanding that machines do what machines are best at, and humans do what humans are best at, but there is a layer of understanding that can bridge humans and machines. There are also situations where machines can do better than humans, and we should promote those situations. For example, back to healthcare, machines— compared to humans— could potentially have much better memory, and better compute capacity to integrate data in a way that humans cannot, and we should promote that. We also, in the meantime, need machines to have enough sensitivity and understanding of a human situation that either recognize we are at the limit of the machines or collaboratively remind the humans or assist the humans to get the task done. I'm thinking about a rescue situation. Knowing the condition of the human who needs to be rescued or helped is so important: where the pain is, or what their emotional state is, or the kind of movements humans can do, and the environment that can impact the humans. If we want machines to help, they need to have the kind of intelligence that could comprehend that situation.

AMS: Yes. One of the things that interests me most about AI is that figuring out what machines can do best often requires us to ask what they cannot do, or on the flip side of that, what is the essence of what makes us human? Because it's not the ability to calculate. That's very important but it's not unique to us. So it does require us to think about what is it that only humans can do, or humans can do far better than anything else.

Fei-Fei: Exactly. This is where I think the communication of today's AI is largely deficient. There is sometimes way too much hype about what this technology can do and there's fear or there's sometimes underestimation for misrepresentation of what this technology is.

AMS: Well, it does very quickly either fire the imagination in sort of positive directions or negative directions, or plugs into fears. I think it's hard to be accurate. So we are talking about resilience and you have already identified one way in which AI can make human society much more resilient: in the face of disasters. If we can send in machines in many situations where we

require search and rescue or other kinds of disaster relief that can work better than sending in humans and that then makes us more resilient, more able to recover from a disaster and thus to rebuild. But I'd be interested in other ways that AI can strengthen the resilience of our societies. You mentioned that it can help, for instance, sustainability. So talk to us a little bit about that.

Fei-Fei: Sure. First of all, I think you're right. There are so many ways AI or technology can help in general. I do want to put a caveat on the disaster relief situation. We are talking about the future of AI, I don't want it people to misunderstand that it's already happening because we still have many steps and much work to do to make that kind of AI assistance happen in a prevalent way for human disaster relief. So, in terms of AI coming to help in sustainability(and I can also give other examples) a lot of today's effort in helping the environment, understanding climate change, optimizing our energy usage, involves data and understanding the patterns of data, understanding where are our environment is affected in both prevalent spatial way but also in temporal longitudinal way. Here AI can help in tremendous ways because machine learning computing is a great tool for model prediction, whether it's for climate or environmental data. I have a colleague here at Stanford who uses satellite imagery to help assess crop information in regions, especially poor regions of the world where it is very hard to put resources, especially human resources, on the ground to go collect this kind of data for the region. This kind of analytics that uses satellite imagery and machine learning techniques is tremendously useful for policymakers. So this is just one example. Another example is energy uses, whether we're talking about a home or a big building or an entire city, we could go a long way in better optimizing and saving our energy if we knew how to save electricity or had a smart way to modulate our temperatures through a thermostat in our buildings and many other ways. And again, this is where tools like machine learning optimization can play a big role. So that's just a couple of examples in the area of sustainability. There are many other examples of machines helping humans. My own favorite topic these days and also my research topic has to do with aging. Our world is aging, a lot of societies including our country are going to have a growing aging population. If used correctly and smartly, the technologies were developing whether it's AI or robotics can have a profound impact in helping our elders to live better, work better, communicate better, and be more mobile, more independent. These are all possibilities, whether we're talking about driver assistance in the self-driving car revolution or healthcare improvements in early dementia detection or well-being management, we can hope that machines can play a role in helping us.

AMS: I love that. I certainly think often about aging and and the points at which people's lives change dramatically as they age, and one of them absolutely has traditionally been when you can no longer drive, when the family takes the keys away or you have some kind of accident. So the idea that you can stay mobile and autonomously mobile as opposed to having to depend on someone else to drive you is a tremendous extension of agency and the kind of autonomy that really defines the difference between being a child and an adult, so that is a wonderful example. You mentioned that AI can help with communication as people age, but it can also help with communication across language lines. And one of the things that I think many people are

already using rudimentary forms of AI in its translator programs. I know that when my children travel, they count on being able to use their phones to help translate when they are talking to someone from another country and I wondered if you would talk about that. You came to the United States when you were 16, and so you obviously learned English and are fluent in English. How do you see translation apps being used? They can't replace learning a language if you're going to live in another country, but they hold out the promise of communication in a way that we've never been able to have for many millions more people.

Fei-Fei: Absolutely. Well, first let me say that I wish I had an AI to help me when I was learning English. I was carrying a physical dictionary that was almost my size. That has really changed indeed. I myself rely on translation apps to understand menus and so on. So absolutely, you're totally right Anne-Marie. There are many translation applications happening, most of our apps or websites today, whether it's Google or other websites, already have that button that can translate a Spanish language document into English. But they're not perfect. So we as humans need to know how to use it, and watch out for the pitfalls. But there's also video conference applications or products that can almost real-time translate the speaker speeches from one language to the other and this is just so helpful. In the entertainment industry, there is a huge need for translating all sorts of programs from all sorts of languages. My own favorite example is in medicine. Imagine a world where medical translation is seamless, and patients and doctors around the world can exchange information and experiences because of that availability of translation. It would help so many patients and also help advance medicine because we all know that the collective data and cases that we learn from around the world will help all of us. This kind of example goes on in education, in law, in government communication, and of course, business, in so many cases.

AMS: You're right, and I hadn't really thought about the medical implications. Although it's also true even in many cities around the world where doctors have patients that they can't communicate with unless they have translators, and of course in many of our schools we have tens of languages being spoken by immigrant children who often are shut out from learning because we don't have enough teachers who can be dual language. So in that sense, when you think about AI augmenting human capacity, it's very direct because it will actually help human beings learn in multiple languages. I hadn't fully appreciated that use. Let me shift gears a little bit and ask you about what makes AI systems themselves more resilient, and how do you even think about resilience when you're thinking about how to build an AI system as effectively as possible?

Fei-Fei: Yeah, great question, Anne-Marie. I think I'll answer in two layers. The more standard, technical answer would definitely talk about resilience and robustness of computing, since there are some standard dimensions that the field of computer science has studied for decades, from security, from robustness, and from privacy, and all these points of view and we can get into the technical details. But I think on a meta level, what is very important and what we are advocating, is to recognize: What is the value for machines AI? Resilience is one of them, and to give credit

to a technology ethicist Shannon Baylor who said there are no independent machine values, because machine values are human values. So when we think about the important values of resilience in an AI system, from the design of that AI system to the development to the deployment, we should think about what resilience means to the humans who will be impacted, who will be using this AI system. Once we think from this point of view, a lot of things start to come into the picture. We want a robust system. We want a fair system. We want it to be truthful. We wanted it to respect privacy, and the list would go on to reflect what our human community and human society cares about in this definition of resilience.

AMS: That is fascinating, and it helps us think through the difference between resilience and robustness. One of the things we've been exploring throughout this series is that resilience does not just mean capacity to withstand external forces, to persist in the face of negative forces, though I think many people do think of resilience as a synonym for endurance. But actually, it's more than that. It's a capacity to adapt, to change, even to improve, in the face of challenges in various ways.

Fei-Fei: I would add that resilience is also an insistence on the inner value that whether there are external factors or not, persistence and insistence on the inner value is part of that resilience.

AMS: I agree with that. That may be something you see specifically because you think about human values and machine capabilities together, so the question of persisting with the value that we have to design into the technology is a part of resilience. But one of the other things we think about often is the way in which diversity contributes to resilience. So the most obvious way is that if you have a business and you have a board that is more diverse, that reflects people of many different types who have many different life experiences, and their collective judgment will be better than a more homogeneous board. We have plenty of research to show that, but it's also common sense that you'll have people who can bring different factors to the table because of their different perspectives and different life experiences. One of the problems with AI is that it reflects the biases of those who are designing it. I know this is something you have spent a lot of time thinking about and have founded an organization called AI For All. I'd love for you to talk about how we build diversity into AI or the flip side, how do we counter the biases that are often there?

Fei-Fei: Yeah, thank you for talking about this. I can't agree more with everything you've said. The origin of AI For All is very much connected to the human-centered AI approach. It started with the recognition that this technology, after so many decades of niche laboratory work, is finally taking off and it's going to impact human life and human society. So around 2014 I thought I was living in a very interesting time where there were two kinds of crisis that were simultaneously happening around AI. One of the crises is this media talk about Terminators coming next door and we have very prominent people stoking that fear and talking about the doom of humanity because of AI. I was living a daily life of a different crisis, which is the total

lack of diversity and inclusion in the field of AI. I was, for a long time, the only woman faculty member in Stanford's AI lab. I happen to be the director of that lab for five years, but I was the only woman faculty. We would go to learning and AI conferences, and the technical women population was at best 10% and, in companies that were advancing this technology rapidly, you just didn't see women in any rank, especially leadership ranks. So this was really concerning me because I really started to draw that line between the two dots. If we truly are worried about Terminators or to put it in a more positive way, if we want AI to be more like Baymax the nice robot this week rather than Terminators, we really need to think about the creators of AI, because it's the creators that would be imposing the values or at least would play a huge role in designing and developing this technology. This connected me to the real crisis that I was thinking about: the utter lack of diversity. I only talk about women as an example, but it's even worse when it comes to underrepresented racial minorities and other aspects of human inclusion and diversity. So long story short, during that moment of realization, I also was very lucky that at the time my former PhD student Olga Russakovsky, who is now a Princeton University assistant professor in the area of AI and computer science, was also thinking exactly the same thing. She was very passionate and committed to bringing more diversity into the field of AI. So Olga and I immediately hit off this idea and decided we need to start a high school student program at Stanford to invite more high school young women to get exposed to the field of AI, the technology of AI, and also very importantly, to the human-centered application cases of AI through research projects. So we started a pilot summer camp at Stanford in 2015 for two weeks for just two dozen high school students to come and join the AI lab and learn about AI and do research. It was so successful that by 2016 we had hundreds of applications at Stanford applying for these 24 slots. And we knew that this needed to be scaled up because the demand is huge and we want us to reach all of America eventually, not just the Bay Area. In early 2017 we became a national nonprofit organization called AI For All, co-founded by three Stanford colleagues, Olga Russakovsky, me, and Dr. Rick Summer, also a long-term stem educator at Stanford. In 2019 we had 11 summer camps throughout the country plus one in Canada. We have students who come from racial minorities, rural regions, low-income families, and who are women, and they even come from all over the world, even though we're focusing on American programs right now.

AMS: That is terrific and it shows that there is plenty of talent that many of our most advanced tech companies say they so desperately need. It's often right under their noses, but it doesn't look like what they expect talent to look like. So they perpetuate off of themselves. It's partly human nature to hire yourself because you much more easily recognized talent where you've already seen it. But it's extraordinary that you've scaled so fast. Do you imagine a world of AI that really does include people from all over the place?

Fei-Fei: Yes. Absolutely. First of all, it's not only imagining—it's already happening. As I told you earlier, quite a few years ago I was the only woman faculty at Stanford AI lab. Now, we have five women faculty in the lab of about 20 faculty. So we're not at the eventual goal, but we've made huge progress. Also, if you look at the students and alums of AI For All and what they're

doing, it's just phenomenal what these young students coming from all kinds of backgrounds do. We had one student who grew up in an agriculture community in a strawberry field in a trailer home, with a single mother from Mexico, and she attended our AI For All program and went back to her community and used machine learning and data science tools to analyze water quality for her community. We have students from African American communities attending our Princeton chapter visiting Washington, D.C. who start to connect AI technology and policy making issues with their local Congressmen and women. We have one young woman who is using AI to do creative arts and she has formed a community around her to show how AI can help human expression. So we are seeing this incredible creativity and talent from all over the map to change this field. I'm very confident that through programs like AI For All and collective efforts, we will see a different future.

AMS: And a brighter future. I must say when we think about what those systems are going to look like if they're going to augment human capabilities then all different kinds of humans certainly have to be involved in designing them and thinking through where they can best serve us. We have time for one more question and I want to ask you about a subject that, for many people, reflects a fear of AI, which is about the future of work. And I know you are serving on Governor Gavin Newsom's Future of Work commission for the State of California and you all are charged with thinking about the future of work and making predictions and recommendations. Many people really do hear that machines will replace them in their jobs. Other people talk about how machine learning can augment those jobs. I'd love for you to tell us how you think about that question.

Fei-Fei: Yes Anne-Marie. I think this is a really important topic as we talked about earlier, even though we are committed to creating the technology that augments humans, we cannot pretend that today's jobs will not change, and that it will have a profound human impact on our labor market. I think there's a McKinsey report a couple of years ago already saying that, even if we freeze all technology advances today, 50% of the tasks on various jobs can potentially be automated by existing technology. So the impact is going to be profound. Keep in mind. We're talking about tasks, not jobs. Typically a job involves many different tasks. So in terms of my role in a Governor Newsom's Future of Work Commission, it's quite an honor to be on it. We have a commission of many experts from California. It's a learning process for me to be on the same commission with labor union leaders, with business executives, with government representatives. We're still early in this year long process. You participated in the second of our 11 listening tours, and your speech was extremely inspiring and informative. It's incredible, pulling this multi-dimensional thinking and voices into this issue of future of work and I think this is the right approach. This is a very complex issue, and we should look at it from the technology side. How is it impacting tasks? How is it potentially replacing labor? How is it monitoring labor? We should also look at how it is impacting the consumer who's at the receiving end of people's work. But we also are looking at it from a policy point of view, from economics policy, from ethical legal policy, from incentives, from government. We're looking at it from an education point of view. More and more we recognize that there's a need to continue to learn, to re-skill.

This is also part of human resilience. How do we, throughout our lifetime, participate in continual learning and re-skilling as technology moves the nature of jobs in a different or dynamic way. So I believe this should be a multi-dimensional, multidisciplinary approach. I believe that technology plays a huge role, but the technology should be embedded in a greater context and greater conversation around the societal efforts in tackling the issue of the future of work.

AMS: Well, that sounds like just the right note on which to end. I like very much the vision of technology that is embedded in a human value system in human systems generally—economic system, social systems—in ways that help us both in concrete tasks that we need to carry out but also in our health, our well-being, our education, and the quality of our jobs. As much as people do not want to lose their sources of income, many people are actually working in jobs that could be far more interesting and tap much more of what they can bring to those jobs as multi-dimensional humans. So we are grateful to have had the chance to hear about human-centered AI and AI for All and we thank you very much for the conversation.

Fei-Fei: Thank you Anne-Marie, and thank you for all your leadership and work for bettering our society.