AI pioneer Fei-Fei Li: 'I'm more concerned about the risks that are here and now'

Fei-Fei Li is a pioneer of modern artificial intelligence (AI). Her work provided a crucial ingredient – big data – for the deep learning breakthroughs that occurred in the early 2010s. Li's new memoir, The Worlds I See, tells her story of finding her calling at the vanguard of the AI revolution and charts the development of the field from the inside. Li, 47, is a professor of computer science at Stanford University, where she specialises in computer vision. She is also a founding co-director of Stanford's Institute for Human-Centered Artificial Intelligence (HAI), which focuses on AI research, education and policy to improve the human condition, and a founder of the nonprofit AI4ALL, which aims to increase the diversity of people building AI systems. AI is promising to transform the world in ways that don't necessarily seem for the better: killing jobs, supercharging disinformation and surveillance, and causing harm through biased algorithms. Do you take any responsibility for how AI is being used? First, to be clear, AI is "promising" nothing. It is people who are promising – or not promising. AI is a piece of software. It is made by people, deployed by people and governed by people. Second, of course I don't take responsibility for how all of AI is being used. Should Maxwell take responsibility for how electricity is used because he developed a set of equations to describe it? But I am a person who has a voice and I feel I have a responsibility to raise important issues – which is why I created Stanford HAI. We cannot pretend AI is just a bunch of math equations and that's it. I view AI as a tool. And like other tools our relationship with it is messy. Tools are invented by and large to deliver good but there are unintended consequences and we have to understand and mitigate their risks well. You were born in China, the only child of a middle-class family that emigrated to the US when you were 15. You faced perilous economic circumstances, your mother was in poor health and you spoke little English. How did you get from there into AI research? You laid out all the challenges, but I was also very fortunate. My parents were supportive: irrespective of our financial situation and our immigrant status, they supported that nerdy sciencey kid. Because of that, I found physics in high school and I was determined to major in it [at university]. Then, also luckily, I was awarded a nearly full scholarship to attend Princeton. There I found fascination in audacious questions around what intelligence is, and what it means for a computational machine to be intelligent. That led me to my PhD studying AI and specifically computer vision. Your breakthrough contribution to the development of contemporary AI was ImageNet, which first came to fruition in 2009. It was a huge dataset to train and test the efficacy of AI object-recognition algorithms: more than 14m images, scraped from the web, and manually labelled into more than 20,000 noun categories thanks to crowd workers. Where did the idea come from and why was it so important? ImageNet departed from previous thinking because it was built on a very large amount of data, which is exactly what the deep learning family of algorithms [which

attempt to mimic the way the human brain signals, but had been dismissed by most as impractical] needed. The world came to know ImageNet in 2012 when it powered a deep learning neural network algorithm called AlexNet [developed by Geoffrey Hinton's group at the University of Toronto]. It was a watershed moment for AI because the combination gave machines reliable visual recognition ability, really for the first time. Today when you look at ChatGPT and large language model breakthroughs, they too are built upon a large amount of data. The lineage of that approach is ImageNet. Prior to ImageNet, I had created a far smaller dataset. But my idea to massively scale that up was discouraged by most and initially received little interest. It was only when [Hinton's] group – which had also been relatively overlooked – started to use it that the tide turned. **Your mother inspired you to think about the practical applications of AI in caring for patients. Where has that led?** Caring for my mom has been my life for decades and one thing I've come to realise is that between me, the nurses and the doctors we don't have enough help. There's not enough pairs of eyes. For example, my mom is a cardio patient and you need to be aware of these patients' condition in a continuous way. She's also elderly and at risk of falling. A pillar of my lab's research is augmenting the work of human carers with non-invasive smart cameras and smart sensors that use AI to alert and predict. **To what extent do you worry about the existential risk of AI systems – that they could gain unanticipated powers and destroy humanity – as some high-profile tech leaders and researchers have sounded the alarm about, and which was a large focus of last week's UK AI Safety Summit?** I respect the existential concern. I'm not saying it is silly and we should never worry about it. But, in terms of urgency, I'm more concerned about ameliorating the risks that are here and now. **Where do you stand on the regulation of AI, which is currently lacking?** Policymakers are now engaging in conversation, which is good. But there's a lot of hyperbole and extreme rhetoric on both sides. What's important is that we're nuanced and thoughtful. **What's the balance between regulation and innovation?** Are we trying to regulate writing a piece of AI code or [downstream] where the rubber meets the road? Do we create a separate agency, or go through existing ones? **Problems of bias being baked into AI technology have been well documented and ImageNet is no exception. It has been criticised for the use of misogynist, racist, ableist, and judgmental classificatory terms, matching pictures of people to words such as alcoholic, bad person, call girl and worse. How did you feel about your system being called out and how did you address it?** The process of making science is a collective one. It is important that it continues to be critiqued and iterated and I welcome honest intellectual discussion. ImageNet is built upon human language. Its backbone is a large lexical database of English called WordNet, created decades ago. And human language contains some harsh unfair terms. Despite the fact that we tried to filter out derogatory terms we did not do the perfect job. And that was why, around 2017, we went back and did more to debias it. **Should we, as some have argued, just outright reject some AI-based technology – such as facial recognition in policing – because it ends up being too harmful?** I think we need nuance, especially about how, specifically, it is being used. I would love for facial recognition technology to be used to augment and improve the work of police in appropriate ways. But we know the algorithms have limitations – [racial] bias has been an issue – and we shouldn't, intentionally or unintentionally, harm people and especially specific groups. It is a multistakeholder problem. **Disinformation – the creation and spread of false news and images – is in the spotlight particularly with the Israel-Hamas war. Could AI, which has proved startlingly good at creating fake content, also help combat it?** Disinformation is a profound problem and I think we should all be concerned about it. I think AI as a piece of technology could help. One area is in digital authentication of content: whether it is videos, images or written documents, can we find ways to authenticate it using AI? Or ways to watermark AI-generated content so it is distinguishable? AI might be better at calling out disinformation than humans in the future. **What do you think will be the next AI breakthrough?** I'm passionate about embodied AI [AI-powered robots that can interact with and learn from a physical environment]. It is a few years away, but it is something my lab is working on. I am also looking forward to the applications built upon the large language models of today that can truly be helpful to people's lives and work. One small but real example is using ChatGPT-like technology to help doctors write medical summaries, which can take a long time and be very mechanical. I hope that any time saved is time back to patients. **Some have called you the "godmother" or "mother" of AI – how do you feel about that?** My own true nature would never give myself such a title. But sometimes you have to take a relative view, and we have so few moments where women are given credit. If I contextualise it this way, I am OK with it. Only I don't want it to be singular: we should recognise more women for their contributions.