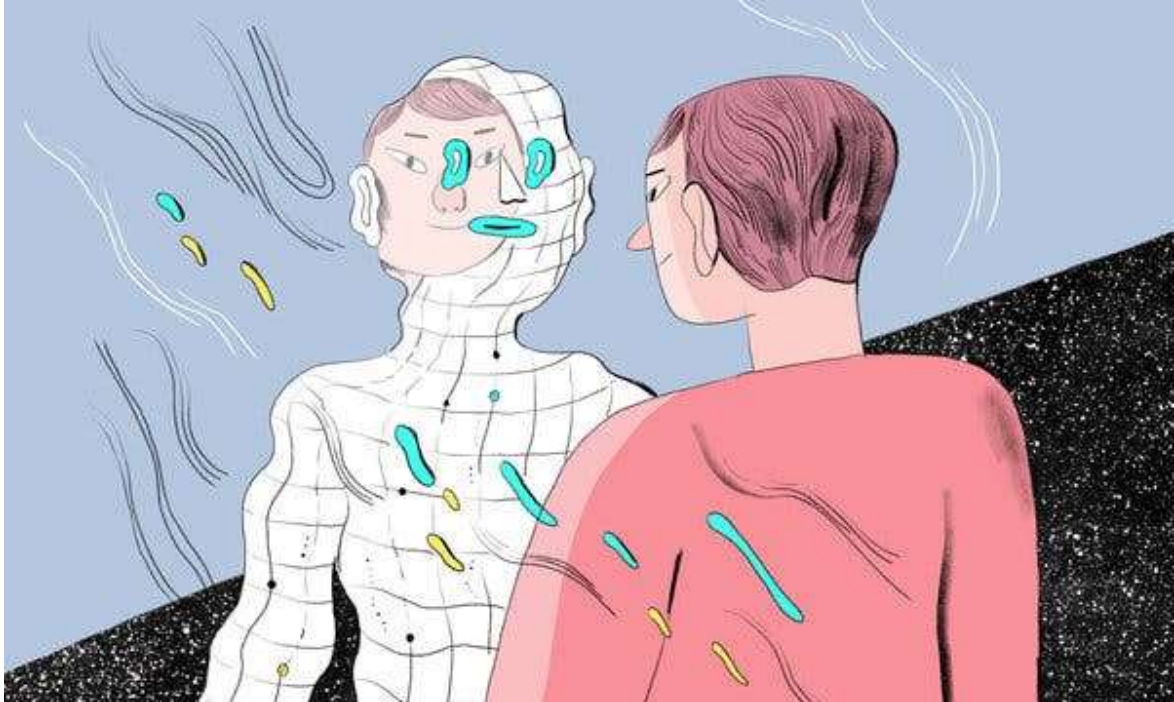Opinion

OP-ED CONTRIBUTOR

How to Make A.I. That's Good for People

By Fei-Fei Li

- March 7, 2018

Credit: Elisa Macellari

For a field that was not well known outside of academia a decade ago, artificial intelligence has grown dizzyingly fast. Tech companies from Silicon Valley to Beijing are betting everything on it, venture capitalists are pouring billions into research and development, and start-ups are being created on what seems like a daily basis. If our era is the next Industrial Revolution, as many claim, A.I. is surely one of its driving forces.

It is an especially exciting time for a researcher like me. When I was a graduate student in computer science in the early 2000s, computers were barely able to detect sharp edges in photographs, let alone recognize something as loosely defined as a human face. But thanks to the growth of big data, advances in algorithms like neural networks and an abundance of powerful computer hardware, something momentous has occurred: A.I. has gone from an academic niche to the leading differentiator in a wide range of industries, including manufacturing, health care, transportation and retail.

I worry, however, that enthusiasm for A.I. is preventing us from reckoning with its looming effects on society. Despite its name, there is nothing "artificial" about this technology — it is made by humans, intended to behave like humans and affects humans. So if we want it to play a positive role in tomorrow's world, it must be guided by human concerns.

I call this approach "human-centered A.I." It consists of three goals that can help responsibly guide the development of intelligent machines.

First, A.I. needs to reflect more of the depth that characterizes our own intelligence. Consider the richness of human visual perception. It's complex and deeply contextual, and naturally balances our awareness of the obvious with a sensitivity to nuance. By comparison, machine perception remains strikingly narrow.

Sometimes this difference is trivial. For instance, in my lab, an image-captioning algorithm once fairly summarized a photo as "a man riding a horse" but failed to note the fact that both were bronze sculptures. Other times, the difference is more profound, as when the same algorithm described an image of zebras grazing on a savanna beneath a rainbow. While the summary was technically correct, it was entirely devoid of aesthetic awareness, failing to detect any of the vibrancy or depth a human would naturally appreciate.

That may seem like a subjective or inconsequential critique, but it points to a major aspect of human perception beyond the grasp of our algorithms. How can we expect machines to anticipate our needs — much less contribute to our well-being — without insight into these "fuzzier" dimensions of our experience?

Making A.I. more sensitive to the full scope of human thought is no simple task. The solutions are likely to require insights derived from fields beyond computer science, which means programmers will have to learn to collaborate more often with experts in other domains.

Such collaboration would represent a return to the roots of our field, not a departure from it. Younger A.I. enthusiasts may be surprised to learn that the principles of today's deep-learning algorithms stretch back more than 60 years to the neuroscientific researchers David Hubel and Torsten Wiesel, who discovered how the hierarchy of neurons in a cat's visual cortex responds to stimuli.

Likewise, ImageNet, a data set of millions of training photographs that helped to advance computer vision, is based on a project called WordNet, created in 1995 by the cognitive scientist and linguist George Miller. WordNet was intended to organize the semantic concepts of English.

Reconnecting A.I. with fields like cognitive science, psychology and even sociology will give us a far richer foundation on which to base the development of machine intelligence. And we can expect the resulting technology to collaborate and communicate more naturally, which will help us approach the second goal of human-centered A.I.: enhancing us, not replacing us.

Imagine the role that A.I. might play during surgery. The goal need not be to automate the process entirely. Instead, a combination of smart software and specialized hardware could help surgeons focus on their strengths — traits like dexterity and adaptability — while keeping tabs on more mundane tasks and protecting against human error, fatigue and distraction.

Or consider senior care. Robots may never be the ideal custodians of the elderly, but intelligent sensors are already showing promise in helping human caretakers focus more on their relationships with those they provide care for by automatically monitoring drug dosages and going through safety checklists.

These are examples of a trend toward automating those elements of jobs that are repetitive, error-prone and even dangerous. What's left are the creative, intellectual and emotional roles for which humans are still best suited.

No amount of ingenuity, however, will fully eliminate the threat of job displacement. Addressing this concern is the third goal of human-centered A.I.: ensuring that the development of this technology is guided, at each step, by concern for its effect on humans.

Today's anxieties over labor are just the start. Additional pitfalls include bias against underrepresented communities in machine learning, the tension between A.I.'s appetite for data and the privacy rights of individuals and the geopolitical implications of a global intelligence race.

Adequately facing these challenges will require commitments from many of our largest institutions. Universities are uniquely positioned to foster connections between computer science and traditionally unrelated departments like the social sciences and even humanities, through

interdisciplinary projects, courses and seminars. Governments can make a greater effort to encourage computer science education, especially among young girls, racial minorities and other groups whose perspectives have been underrepresented in A.I. And corporations should combine their aggressive investment in intelligent algorithms with ethical A.I. policies that temper ambition with responsibility.

No technology is more reflective of its creators than A.I. It has been said that there are no "machine" values at all, in fact; machine values *are* human values. A human-centered approach to A.I. means these machines don't have to be our competitors, but partners in securing our well-being. However autonomous our technology becomes, its impact on the world — for better or worse — will always be our responsibility.

Fei-Fei Li is a professor of computer science at Stanford, where she directs the Stanford Artificial Intelligence Lab, and the chief scientist for A.I. research at Google Cloud.