



I4ALL
AI4ALL
2148 Broadway
Oakland, CA 94612
contact@ai-4-all.org

House Science Committee Testimony

Dr. Fei-Fei Li

Introduction

As I often say to my students, there’s nothing “artificial” about AI. It’s inspired by people, it’s created by people, and—most importantly—it impacts people. It’s a powerful tool we’re only just beginning to understand, and that’s a profound responsibility. I’d like to talk about what this technology looks like today, the challenges we’ll face in the coming years, and what we can do now to ensure we’re building a future worth living in.

Artificial intelligence emerged in the mid 20th century as a quest to build machines with intellectual capabilities similar to those of the human mind. As a science, it draws on fields like cognitive science, neuroscience, statistics and mathematics. As a technology, it represents some of the most active developments in computer science and engineering, including machine learning (ML)—a family of techniques that use statistical modeling to learn from data. Today, AI and ML are part of a vibrant, interdisciplinary pursuit, with fields like robotics, natural language processing, computer vision, speech recognition, and even philosophy playing ever-growing roles.

It’s also changing the world. Thanks to the recent convergence of three key factors—powerful algorithms, fast computing hardware, and the era of big data—AI has rapidly become a driving force in what some are calling the fourth industrial revolution. It can help drive our cars, assist with radiology diagnoses, optimize energy consumption, and even track deforestation from satellites. All told, AI spending is expected to rise from 2017’s \$12B to \$57.6B by 2021, according to International Data Corporation (IDC).

We have good reason to be excited about AI. But this is a nascent field, and as a scientist, that humbles me. After all, we've never created a technology to mimic human qualities so closely, and we know little about the impact it will have on the world. Guiding its development will be an ethical, philosophical and humanistic challenge, and it will require a diverse community of contributors. I call this "Human-Centered AI", and it consists of three simple ideas.

Three Pillars of Human-Centered AI

Pillar I: AI must be more inspired by human intelligence

Much of AI's recent success lies in a narrow range of problems associated with an abundance of annotated example data—such as accurately labeled photos or text—analyzed by algorithms in process called *supervised learning*. But this approach has its limits. 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 statues. Similar lapses are found in speech recognition, natural language processing, robotics and more.

Simply put, AI technologies such as machine perception remains shallow compared to our own visual intelligence. Algorithms lack our talent for extrapolating from few examples, they tend to miss the big picture when given disorganized information, and they're devoid of our remarkable capacity for context and emotion. Researchers are working hard to address these shortcomings, and progress will deliver tools that understand our needs with newfound depth. But their work underscores the growing need for collaboration between AI and neuroscience, cognitive science and the behavioral sciences.

Pillar II: AI should strive to enhance us, not replace us

A recent McKinsey report found that half of all current work activities can be theoretically automated by technology that *already* exists. This is an alarming fact, and there's no denying that job displacement will be a central challenge of the coming decade. But what if we imagine AI as a tool for *augmenting* human capabilities, rather than replacing them?

For example, consider the problem of hospital-acquired infections. Although they claim over 90,000 lives each year—many times more than traffic fatalities—their cause is often simple human error. In response, my Stanford colleagues and

collaborators are partnering with Stanford Children's Hospital to explore the use of smart sensors to track hand hygiene automatically, with the goal of improving health outcomes and giving caregivers greater peace of mind. So while a nurse or doctor's true skills are unlikely to be replaced by machines—complex problem solving in uncertain situations, bedside manner, and an emotional connection with patients—this research demonstrates how technology can enhance them.

Similar opportunities to intelligently augment human capabilities abound, from healthcare to education, manufacturing to search and rescue, and city planning to entertainment. It's true that automation is a challenge to certain forms of human labor, but it's also an opportunity to elevate our notion of work.

Pillar III: AI must be guided by a concern for its human impact

The sheer power of AI means we have an obligation to consider its unintended consequences as thoughtfully as its desired effects. This begins with the technology itself, where fairness is among the most pressing concerns. For example, a recent study revealed that widely used facial recognition software recognizes white, male faces with considerably higher accuracy than darker-skinned and/or female faces.

Addressing such inequities will force us to confront a host of related problems. One is the issue of *interpretability*, or an AI system's capacity to explain its reasoning. As they play a growing role in deliberations ranging from parole hearings to loan applications, this ability will be critical. And because AI relies on data to learn about the world through large quantities of images, text, video and personal records, privacy and security will pose major challenges as well.

The picture only gets more complex at the societal level. First, some degree of labor disruption is inevitable, and now is the time to prepare for its effects with proactive measures like retraining. Next, licensing and certification will have to expand to include technology that can perform medical, legal and other highly regulated tasks without human involvement. And our ethics, not to mention the laws that codify them, will have to evolve to answer questions of accountability as their role in our lives grows more consequential.

Society will soon face changes we can't even imagine, and these predictions are only the beginning. We can expect an exciting but uncertain future, and a sense of humility will serve us well.

What Can Be Done?

In Government

Like nuclear technology, biotech and energy security, AI should be an active topic of discussion among policymakers.

This will require a far greater baseline of AI proficiency in all three branches of government, however, as well as ongoing dialogues with industry and academia. This technology is simply too important to be owned by private interests alone, and publicly-funded research can provide a transparent, ethical foundation for its development.

However, both public and private research will depend on a growing field of technologists with cutting-edge expertise in subjects like machine learning and data science. To ensure we can meet demand, it's vital we provide funding for education and diversity outreach in the STEM fields, with an emphasis on AI, starting as early as possible.

In Academia

Academia has a unique opportunity to bring clarity and rigor to our understanding of AI. The rare mix of intellectual resources found in our universities is a perfect environment for studying the effects of AI on our world, as well as supporting the cross-disciplinary demands of the next generation of research. Current campuses often separate computer science from other fields like the humanities and social sciences, but the time has come for this practice to change.

Academic leaders can encourage discussions across departments in the form of classes, seminars, institutions, symposiums and joint projects.

It's encouraging to note that academic venues are already playing a unique role in combating bias. Organizations like FAT/ML (Fairness, Accountability, and Transparency in Machine Learning), IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, the Partnership on AI and AI Now Institute are active sources of thought leadership in the ethical implementation of AI. I hope to see this trend continue.

In Industry

Finally, businesses must develop a better balance between their responsibility to shareholders and their obligations to their users, community and society at large. After all, commercial products tend to change the world fastest. So while the tech industry's aggressive investment in AI promises rapid advances, the time has come to complement that ambition with ethical, responsible policies.

I'm happy to say some companies are already taking the first steps, with published guidelines intended to establish ethical boundaries and best practices for the field. But this is just a start. Across the tech industry and beyond, much more can be done to promote the responsible development of AI, including relationships with academia, nonprofits and governments to encourage the study of its human impact.

The Common Thread: More Humanity

Even after nearly two decades in this field, I'm still heartened by the creativity, intelligence and diversity of the people driving it. I introduce high school students to AI research through AI4ALL, a non-profit I co-founded, advise doctoral students at Stanford, and work with Google researchers worldwide. I've learned that no age, ethnicity or gender has a monopoly on innovation. In fact, more and more of the best work in our field comes from women, people of color, and other groups that have traditionally been underrepresented in STEM.

Unfortunately, a lack of representation remains a crisis in AI. A 2014 report revealed that women in tech feel stalled, encounter persistent bias, and are statistically more likely to leave their jobs. As a result, females hold only a third of all high-tech positions, and even fewer at the executive level. In fact, only 13% of AI companies have female CEOs. For people of color, the numbers are even worse.

The tech world has a long way to go to address its lack of diversity, starting with hiring practices. Improvements are likely to yield immediate benefit, as research¹ from McKinsey and Credit Suisse to the Journal of Personality and Social Psychology have quantified the benefits of diversity on workplace performance. But studies also suggests this problem

¹ <https://hbr.org/2016/11/why-diverse-teams-are-smarter>

has far earlier origins; girls, for instance, tend to lose interest in STEM curricula around age 15, usually due to a lack of peers and role models that look like them. How should we expect the world's most powerful algorithms to behave if they remain ignorant to the diversity of human experience?

Organizations like AI4ALL were founded to reverse this trend. Our summer program gives students from underrepresented communities first-hand experience in real AI labs at Stanford and UC Berkeley, and we're expanding across North America this year. 91% of our graduates feel confident they'll pursue a career in the field, and 97% feel they're part of a community in AI and computer science. In the words of one recent graduate, "I used to think I wasn't smart enough to do computer science and AI. But now I've gained so much confidence because of all the support, and being around other girls who are into the same thing."

Conclusion

As computational ethicist Shannon Vallor said, there are no independent "machine" values; machine values are *human* values. A human-centered approach to AI means these machines don't have to be our competitors, but partners in securing our wellbeing. However autonomous our technology becomes, its impact on the world will always be our responsibility. Now's the time to start making sure its an impact we can be proud of.