WORLD ECONOMIC FORUM

Top 9 ethical issues in artificial intelligence

Faced with an automated future, what moral framework should guide us? Image: Matthew Wiebe

21 Oct 2016

Julia Bossmann
Global Shaper, San Francisco Hub, Fathom Computing

Optimizing logistics, detecting fraud, composing art, conducting research, providing translations: intelligent machine systems are transforming our lives for the better. As these systems become more capable, our world becomes more efficient and consequently richer.

Tech giants such as Alphabet, Amazon, Facebook, IBM and Microsoft – as well as individuals like Stephen Hawking and Elon Musk – believe that now is the right time to talk about the nearly boundless landscape of artificial intelligence. In many ways, this is just as much a new frontier for ethics and risk assessment as it is for emerging technology. So which issues and conversations keep AI experts up at night?

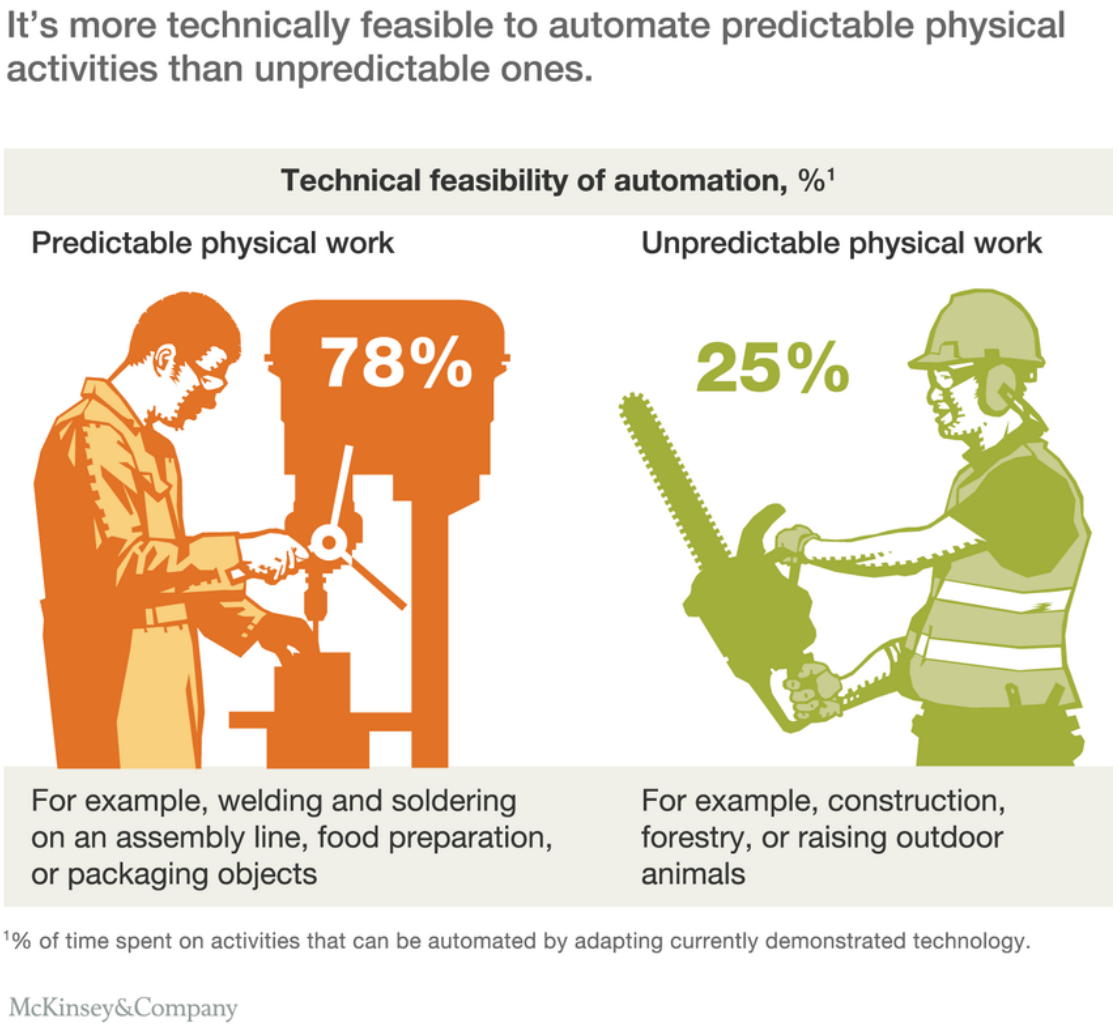
1. Unemployment. What happens after the end of jobs?

The hierarchy of labour is concerned primarily with automation. As we've invented ways to automate jobs, we could create room for people to assume more complex roles, moving from the physical work that dominated the pre-industrial globe to the cognitive labour that characterizes strategic and administrative work in our globalized society.

Look at trucking: it currently employs millions of individuals in the United States alone. What will happen to them if the self-driving trucks promised by Tesla's Elon Musk become widely available in the next decade? But on the other hand, if we consider the lower risk of accidents, self-driving trucks seem like an ethical choice. The same scenario could happen to office workers, as well as to the majority of the workforce in developed countries.

This is where we come to the question of how we are going to spend our time. Most people still rely on selling their time to have enough income to sustain themselves and their families. We can only hope that this opportunity will enable people to find meaning in non-labour activities, such as caring for their families, engaging with their communities and learning new ways to contribute to human society.

If we succeed with the transition, one day we might look back and think that it was barbaric that human beings were required to sell the majority of their waking time just to be able to live.



2. Inequality. How do we distribute the wealth created by machines?

Our economic system is based on compensation for contribution to the economy, often assessed using an hourly wage. The majority of companies are still dependent on hourly work when it comes to products and services. But by using artificial intelligence, a company can drastically cut down on relying on the human workforce, and this means that revenues will go to fewer people. Consequently, individuals who have ownership in AI-driven companies will make all the money.

We are already seeing a widening wealth gap, where start-up founders take home a large portion of the economic surplus they create. In 2014, roughly the same revenues were generated by the three biggest companies in Detroit and the three biggest companies in Silicon Valley ... only in Silicon Valley there were 10 times fewer employees.

If we're truly imagining a post-work society, how do we structure a fair post-labour economy?

3. Humanity. How do machines affect our behaviour and interaction?

Artificially intelligent bots are becoming better and better at modelling human conversation and relationships. In 2015, a bot named Eugene Goostman won the Turing Challenge for the first time. In this challenge, human raters used text input to chat with an unknown entity, then guessed whether they had been chatting with a human or a machine. Eugene Goostman fooled more than half of the human raters into thinking they had been talking to a human being.

This milestone is only the start of an age where we will frequently interact with machines as if they are humans; whether in customer service or sales. While humans are limited in the attention and kindness that they can expend on another person, artificial bots can channel virtually unlimited resources into building relationships.

Even though not many of us are aware of this, we are already witnesses to how machines can trigger the reward centres in the human brain. Just look at click-bait headlines and video games. These headlines are often optimized with A/B testing, a rudimentary form of algorithmic optimization for content to capture our attention. This and other methods are used to make numerous video and mobile games become addictive. Tech addiction is the new frontier of human dependency.

On the other hand, maybe we can think of a different use for software, which has already become effective at directing human attention and triggering certain actions. When used right, this could evolve into an opportunity to nudge society towards more beneficial behavior. However, in the wrong hands it could prove detrimental.

4. Artificial stupidity. How can we guard against mistakes?

Intelligence comes from learning, whether you're human or machine. Systems usually have a training phase in which they "learn" to detect the right patterns and act according to their input. Once a system is fully trained, it can then go into test phase, where it is hit with more examples and we see how it performs.

Obviously, the training phase cannot cover all possible examples that a system may deal with in the real world. These systems can be fooled in ways that humans wouldn't be. For example, random dot patterns can lead a machine to "see" things that aren't there. If we rely on AI to bring us into a new world of labour, security and efficiency, we need to ensure that the machine performs as planned, and that people can't overpower it to use it for their own ends.

5. Racist robots. How do we eliminate AI bias?

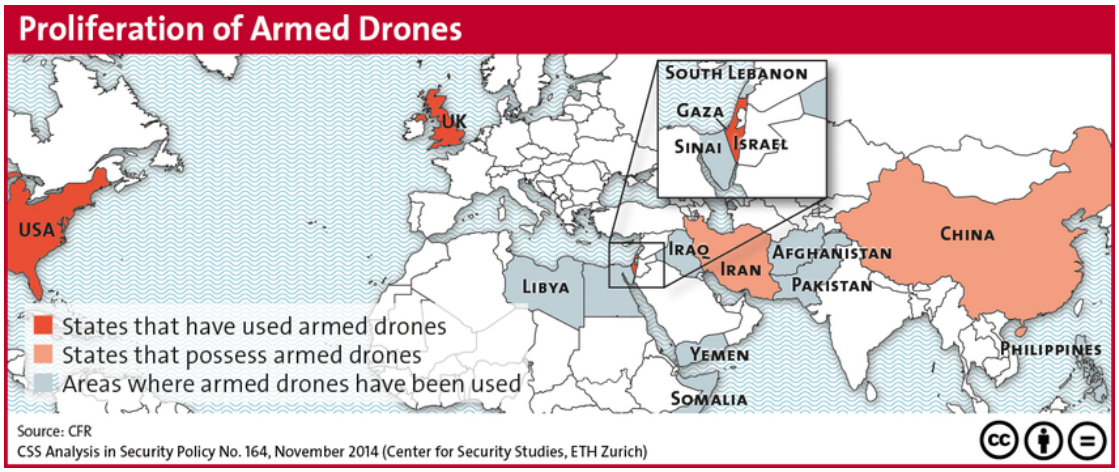
Though artificial intelligence is capable of a speed and capacity of processing that's far beyond that of humans, it cannot always be trusted to be fair and neutral. Google and its parent company

Alphabet are one of the leaders when it comes to artificial intelligence, as seen in Google's Photos service, where AI is used to identify people, objects and scenes. But it can go wrong, such as when a camera missed the mark on racial sensitivity, or when a software used to predict future criminals showed bias against black people.

We shouldn't forget that AI systems are created by humans, who can be biased and judgemental. Once again, if used right, or if used by those who strive for social progress, artificial intelligence can become a catalyst for positive change.

6. Security. How do we keep AI safe from adversaries?

The more powerful a technology becomes, the more can it be used for nefarious reasons as well as good. This applies not only to robots produced to replace human soldiers, or autonomous weapons, but to AI systems that can cause damage if used maliciously. Because these fights won't be fought on the battleground only, cybersecurity will become even more important. After all, we're dealing with a system that is faster and more capable than us by orders of magnitude.



7. Evil genies. How do we protect against unintended consequences?

It's not just adversaries we have to worry about. What if artificial intelligence itself turned against us? This doesn't mean by turning "evil" in the way a human might, or the way AI disasters are depicted in Hollywood movies. Rather, we can imagine an advanced AI system as a "genie in a bottle" that can fulfill wishes, but with terrible unforeseen consequences.

In the case of a machine, there is unlikely to be malice at play, only a lack of understanding of the full context in which the wish was made. Imagine an AI system that is asked to eradicate cancer in the world. After a lot of computing, it spits out a formula that does, in fact, bring about the end of cancer – by killing everyone on the planet. The computer would have achieved its goal of "no more cancer" very efficiently, but not in the way humans intended it.

8. Singularity. How do we stay in control of a complex intelligent system?

The reason humans are on top of the food chain is not down to sharp teeth or strong muscles. Human dominance is almost entirely due to our ingenuity and intelligence. We can get the better of bigger, faster, stronger animals because we can create and use tools to control them: both physical tools such as cages and weapons, and cognitive tools like training and conditioning.

This poses a serious question about artificial intelligence: will it, one day, have the same advantage over us? We can't rely on just "pulling the plug" either, because a sufficiently advanced machine may anticipate this move and defend itself. This is what some call the “singularity”: the point in time when human beings are no longer the most intelligent beings on earth.

9. Robot rights. How do we define the humane treatment of AI?

While neuroscientists are still working on unlocking the secrets of conscious experience, we understand more about the basic mechanisms of reward and aversion. We share these mechanisms with even simple animals. In a way, we are building similar mechanisms of reward and aversion in systems of artificial intelligence. For example, reinforcement learning is similar to training a dog: improved performance is reinforced with a virtual reward.

Right now, these systems are fairly superficial, but they are becoming more complex and life-like. Could we consider a system to be suffering when its reward functions give it negative input? What's more, so-called genetic algorithms work by creating many instances of a system at once, of which only the most successful "survive" and combine to form the next generation of instances. This happens over many generations and is a way of improving a system. The unsuccessful instances are deleted. At what point might we consider genetic algorithms a form of mass murder?

Once we consider machines as entities that can perceive, feel and act, it's not a huge leap to ponder their legal status. Should they be treated like animals of comparable intelligence? Will we consider the suffering of "feeling" machines?

Some ethical questions are about mitigating suffering, some about risking negative outcomes. While we consider these risks, we should also keep in mind that, on the whole, this technological progress means better lives for everyone. Artificial intelligence has vast potential, and its responsible implementation is up to us.
