Artificial Intelligence: Are We Prepared for the Breakthrough?

Yijun Jiang (10/19/2015, STS.034, second draft of op-ed)

This summer, the fifth episode of the movie *the Terminator* was released and has earned nearly $450 million worldwide at box offices. In the movie, scientists created Skynet, an artificial intelligence (AI), or simply speaking a smart computer system, which is so advanced that it has gained self-awareness and treats humans as its threat. Led by the savior John Conner, humans must fight against their own machines to prevent themselves from being extinct.

The evil depiction in the film has not at all bothered people's interest in AI. Several weeks ago, Apple bought VocalIQ, a UK-based startup who develops softwares to help computers, such as Siri, better understand human languages. Since its release in 2011, Siri has undergone a series of improvements. Compared with its first version which only responded to a few basic instructions, Siri has grown into a multifunctional intelligent personal assistant. It understands not only your orders to call or text a friend, but also requests for almost any information available, from flight schedule to odds of a royal flush.

Siri, though significantly different from Skynet, is a typical example of AI. Over the past few years, research results in AI exploded. Anti-spam filters, spelling checks, online translation and face detection jump to us one after another and make our life much easier. Technology giants are also intensively competing against each other in AIs that interact more with the physical world. Google, for example, is actively engaged in its Driverless Car project, where a computer instead of a driver analyzes what is happening on the road and tells the car what to do. It can be imagined that in the future, AIs will take over most of our tedious work and complete tasks beyond our current abilities.

As the fanatic enthusiasm in AI research carries on, more and more people start questioning whether AIs are trustworthy. What if somebody hacks into a driverless car and tries to kidnap the passengers? What if my Siri has a bug and deletes all my important files? However, many AI scientists and software companies claim that these concerns are unnecessary as long as software developers pay enough attention, do enough testing and make their codes robust enough.

Well, yes and no. Those specialists are right that such problems can be easily fixed by clearing the bugs and releasing software patches. Even if an AI program goes crazy, we can always turn off the emergency switch to prevent further damage. But there is another hidden risk of AI that is not mentioned that much – the risk of creating a monster like Skynet in the movie.

I admit that it *sounds* ridiculous to relate current research on AI with a movie where human civilization is endangered by intelligent machines. The main reason is, despite of the rapid development in AI studies, the AIs that we are working with are still in an elementary stage. They are usually programmed to be specialized on only one job. Although they are much more efficient than us in this particular job, they lack the ability to “think”, that is, to learn from experiences, make judgments and self-improve beyond what software engineers have taught them. Such an AI cannot *deliberately* do us harm. All we need to be aware of is a sudden crash of codes that malfunctions the machine.

But Skynet is much more complicated. It actively learns from the outside world and improves itself. This behavior is exactly like the evolution of ourselves. The only difference is, the evolution of a powerful AI takes much less time than us due to its hardware advantage that we organisms do not possess. In an epoch when supercomputers can simulate the evolution of a galaxy over billions of years within days, tracing the history of life and biological intelligence in a short time is no longer unthinkable, once the AI knows how to start. However, our technology is still unable to create such a starting point. So why do we bother worrying about something that is not known yet?

Because this day may not be far away. The fact is, the pace of human development is accelerating over the history, and this trend will keep on into the future. People in the 1500s would be surprised by the development in classical physics and maritime technology if they had time traveled to 200 years later. However, the same kind of surprise can be experienced merely over the recent 20 years, when inventions like the Internet, smart phones and new materials came into being. The important lesson is, it is misleading to predict the future based on our current pace. Breakthroughs come much faster than we have often expected.

This accelerating trend applies equally well to the study of AI. As we gain more and more experience on elementary AIs, we are accelerating the advent of the breakthrough, where AIs acquire the general ability to evolve and become far superior than us in all aspects. It is like mixing amino acids on early earth and suddenly creating life among them.

In a survey by the Swedish philosopher Nick Bostrom in 2013, hundreds of AI experts are asked to estimate the year of this breakthrough. Statistics show that, by the year 2050, the breakthrough is believed to come with a probability of 50%. Not surprisingly, some specialists hold different views. Paul Allen, the founder of Microsoft, believes that we may have underestimated the potential difficulties in AI research.

No matter who makes the correct prediction, it is necessary to be prepared for this super AI. We cannot reject it because it shows such great potential in dealing with some of the biggest challenges of mankind such as global warming and energy crisis. But on the other hand, we must make sure that it does not turn into an evil Skynet. Unfortunately, driven by profits, most AI studies are completely focused on the development of better AIs, without paying any attention to this possible risk. And as I have said, the risk is not that far away.

Admittedly, precaution of such a risk is not an easy job and demands much serious effort, possibly even more than the development of AI itself. This is because, to prevent a Skynet we have to teach AIs to respect the values of human society, a task much easier said than done. Undoubtedly, technical as well as ethical issues will emerge and entangle. But this is not an excuse to refuse relevant research. Since a super AI will eventually come to us, probably the wise thing to do is to start thinking of it right now instead of continuing our aggressive effort to improve current AIs and hitting the breakthrough point unprepared.

As a big fan of *the Terminator*, I am looking forward to its sixth episode. However, in real life there is no John Conner for us to put faith in. We need to be responsible for every step we make pushing forward artificial intelligence. And it is never too early to ask ourselves if we are prepared for the breakthrough.