

Artificial Intelligence: Are We Prepared for a God?

Yijun Jiang (10/07/2015, STS.034, first draft of op-ed)

Several days 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 instructions to call or text a friend, but also requests for almost any information available, from flight schedule to odds of a royal flush.

It is not just Apple who cares about intelligent machines. Technology giants such as Microsoft and Google are fully engaged in the game. In the famous Google Driverless Car project, a computer instead of a driver analyzes what is happening on the road and tells the car what to do. Other examples include the online conversation maker Cleverbot, the song composer MySong and automated drones.

Computer scientists call these smart computers, technically speaking the algorithms inside them, artificial intelligence (AI). As its name suggests, AI is the science of simulating human intelligence on a machine. Its history dates back to 1950, when Alan Turing published his famous paper *Computer Machinery and Intelligence*, in which the question “can machines think” was raised. Since then, more and more studies were made to answer this question. In 1956, the first AI software was demonstrated at Carnegie Mellon University. In 1997, IBM's chess machine Deep Blue defeated the world champion Garry Kasparov. In 2014, the computer program Eugene convinced 33% of human judges after a 5-minute conversation that it was a 13-year-old Ukrainian boy instead of a machine. It is thus regarded as the first AI to pass the Turing Test, a test on a machine's ability to exhibit human intelligence.

As we are progressing towards better AIs, hot debates are taking place on their possible impacts over human beings. On the one hand, many AI scientists and software companies believe (or at least claim) that AIs will do us good by taking care of the jobs that we are not good at or even unable to do such as stock market prediction and drug development. On the other hand, there are increasing worries in the general public about machines overtaking the control of human society, and even destroying the entire human civilization.

A major cause of public concern comes from fictions and movies, where AIs, like Skynet in *the Terminator*, are often depicted as so well-developed that they gain self-awareness and treat the human race as their threat. However, even the state-of-art AIs today are still far from being that powerful. These AIs are usually programmed to be specialized on only one job and are thus known as artificial

narrow intelligence (ANI). An ANI cannot do us harm *deliberately*. An intelligent personal assistant might delete all your important files and make your life in a mess; a driverless car might try to run over you and put you in danger. But these unfortunate events are usually caused by either someone hacking into the system or a bug ignored by software engineers. The fix to these problems is easy. We can improve the codes so that they are more reliable. We can also just power off the machines when they go wrong.

The fact that our abilities are limited to ANI today is a main reason why optimists feel safe. Skynet seems too far from us. Even if it does come into being one day, why can't we still simply unplug it when it declares war against us?

Such thoughts may be completely wrong. Looking back on our history, we see an exponentially accelerating trend of human development, a phenomenon described by Ray Kurzweil as *the Law of Accelerating Returns*. This means that breakthroughs will come to us faster than we have ever expected. Moreover, once a breakthrough is made, subsequent developments will happen at an astonishing speed. Just think of how short it takes our society to be informationalized after the emergence of the Internet! As a result, the advent of an AI that is intellectually comparable to human in all aspects, or an artificial general intelligence (AGI), may happen within decades.

Once AGI comes into being, because of its ability to learn from the outside world, it will almost undoubtedly evolve with time. Due to the advantage over human beings in hardware, an AGI will evolve much faster than us. It may take just a few days for a machine to finish the evolution that took us millions of years to go through. Moreover, it can only evolve faster and faster, again by *the Law of Accelerating Returns*, such that it will quickly become far superior than us. By that time, any human control over this machine, including modifying its code or cutting off its power, will be as ridiculous as an ant trying to control us. The machine will find its way around, quite possibly not at all understandable to us, to protect itself from being intervened by human effort.

If a godlike AI is inevitable, all that we can do is to ensure that it respects the values of human society, so that it is a benevolent God. This is definitely not an easy task and requires more serious research. Unfortunately, as the fierce competition for better AIs goes on, necessary awareness of the risks of creating a real Skynet is still insufficient in the society. Blind and aggressive study may bring us to the critical point from ANI to AGI so suddenly that we are not yet fully prepared.

Today, AIs are our helpers that make our lives easier. Tomorrow, AIs will be our friends that we rely on. But eventually, they might become our Gods, who have the power to bring us to either eternity or

extinction. There is, however, no John Conner for us to put faith in. We need to be responsible for every step we make pushing forward artificial intelligence.