

Artificial Intelligence and Consciousness – IONS

Blog

Artificial Intelligence and Consciousness

April 23, 2019

Arnaud Delorme, Scientist

There is a lot of hype in the media about artificial intelligence with new so-called deep learning AI able to beat people at the Go and Chess Master games, and play video games, including some complex social games such as Warcraft and Dota 2. Some would think that the era of true sentient artificial intelligence is upon us – an intelligence that rivals the intelligence of humans – and this could be a threat or a blessing.

The threat is real, but not in the way we usually see it in the movies, where a sentient AI might decide to suppress the suboptimal human race, as cleverly pictured by numerous movies and series. The risk of sentient AI taking over humanity is not real at this point. Sentient AI is still science fiction as, in the materialist paradigm, the process through which sentience emerges from matter is as mysterious as it was in the Greek era, two thousand years ago. The main explanation – which is truly unsatisfying to many – is that consciousness is a property of nature, like one of the fundamental forces in physics. This is what we are working on at IONS.

The real threat of AI is if its used as a tool in human conflicts and in warfare. Technology has always been at the core of human warfare. Autonomous AI provides a cost-effective solution for engaging an enemy while helping to minimize casualties. For example, imagine a small autonomous vehicle that could patrol the streets of rebel Middle Eastern cities and eliminate any human who is holding any kind of weapon using face recognition and all the bells and whistles that modern AI are now capable of. The system would not require human intervention and could decide to act on his own. Such systems probably already exist, although they have not been deployed for fear of public opinion backlash. The problem with such an approach is that because there is no risk of human casualties for the party using the autonomous AI drone or robot, it decreases the threshold at which military are willing to engage the enemy, therefore multiplying conflict. There is also the risk of sparking a technological race for designing such systems and its potential misuse. This is well pictured in the conspiracy film *Captain America: The Winter Soldier* where a corrupt ruler decides to become dictator. The self-appointed dictator finds the few million people who were most likely to rebel against such a coup using social media personality profiles – which, in real life, have been shown to be extremely accurate – then sent autonomous drones to eliminate such persons who may poses a potential threat (most likely using a combination of face recognition technology, cell phone identification, etc.) The autonomous drones would troll the country and simply eliminate potential dissidents. Of course, Captain America saves the day, but such a scenario is actually possible using existing technology.

On the other hand, one could look at AI and see it as part of the future evolution of humanity.

We are not separate bodies interacting with a world that is separate from us. For example, glasses are a piece of technology that is an extension of our body, in the same way a fire protection suit is for firefighters or cars are for taxi drivers. Smartphones have changed our brain wiring and how we interact with the world. The Internet has already changed our cognition, which is becoming more procedural and less information oriented, since information is readily available within a click or two.

In a similar way, AI will integrate slowly into our daily lives, with augmented reality – a system that super-imposes over our vision, through glasses or contact lenses, additional information – such as the rating of each business as we stroll down a street or the ability to buy online with a blink, etc. In the future, one may easily imagine a power boost processor that is implanted in the brain and able to communicate with neurons (something already possible with hearing implants supporting thousands of people in the world). Such systems could, for example, allow us to understand any language – something that is probably only a few years away if it doesn't already exist (verbal translation apps exist and hearing implants are already connected to smartphones). Implanted chips could help us access any kind of information on the Internet from within our brain, just by thinking about it.

The science fiction writer Greg Egan describes such an implanted chip in the short story *The Jewel*, but goes further. In his novel, a chip is implanted in infancy and it learns over several decades exactly how the brain of its host works. At about age 30, when the brain starts to show signs of aging, people undergo a surgery where the brain is removed and all cognition and sentience is transferred to the “Jewel” (the implanted chip), which is at this point perfectly mimicking the functioning of the brain in all aspects. The story describes the existential crisis individuals undergo as they go through this transition. As crazy as it seems, consciousness researcher Francisco Varela argues that physical dynamical systems should be defined by their function and not their parts – this is the autopoiesis hypothesis. This is why we are able to build artificial hearts and maybe one day, artificial brains. Note that this would not make consciousness any less mysterious or show that consciousness emerges from matter. It would just show that the brain, like any other organ, may be replaced. This might be in humanity's future and our future evolution as well.

Going further, one may even imagine a consciousness not bound to a human body. Consciousness content made of information may travel at the speed of light. For example, following the analogy above, if the implanted chip is now the seat of sentience, why not transfer its information through the Internet to another chip on another body on the other side of the planet. Almost instant Star Trek-like teleportation on Earth and beyond. This type of scenario is much more aligned with the universal consciousness we are working on at IONS than with the limited self-centered consciousness contained in the brain that is the current view of mainstream science.

If humanity survives its own insanity, this could be the bright future that lies ahead of us. Our silicon brain might still wonder what consciousness is, just faster.



Arnaud Delorme, PhD, is a research scientist at the Institute of Noetic

Sciences, a CNRS principal investigator in Toulouse, France, and on the faculty at the University of California, San Diego.

[Read Bio](#)