28/07/2008
News

Did a philosopher kill WALL−E?

WALL−E movie poster from Pixar.

Disney Pixar have just released the movie, WALL−E. A bleak, post−apocalyptic tour−de−force, the movie depicts the gentle romance between two robots of the future: WALL−E, the not−so−bright and not−so−attractive "guy" with the big heart and sweet personality, and EVE, the sleek, sexy, totally out−of−his−league babe.

The story goes like this: a hundred years into the future, Earth over−polluted and overtaken by garbage can no longer sustain life. So we flee to outer space, leaving the planet's clean−up in the mechanical pincers of an army of stout, capable robots.

Seven hundred, entirely uneventful years pass and now pillars of compacted trash line the city skies like

towering skyscrapers. One day, WALL−E now the sole surviving creature of his kind meets EVE, a visitor from outer space with a mysterious mission.

Pixar designed these robots so that they're, well, human. We see them as human. We see them communicate, we see them think, act, understand, love. And we accept this. By the end of the movie, we've accepted WALL−E and EVE as equals and we may even shed a tear here and there for our new−found friends.

But what exactly is WALL−E? Is he pure fantasy and speculative fiction?

Or is he is artificial intelligence simply the way of the future?

Alan Turing's Vision

"I believe that in about fifty years' time it will be possible to programme computers [...] so well, that an average interrogator will not have more than 70 per cent chance of making the right identification [between human and machine] after five minutes of questioning."

Alan Turing, *Computing Machinery and Intelligence*, Mind (1950), 442.

This prophecy, published in 1950 by English mathematician Alan Turing was a bold statement indeed. Remember, in that day and age, computers weren't sleek, glossy, or available in a variety of neat colours; no, they where clunky, they weighed nearly 30 tons, and they took gaggles of people to operate. Turing, however, saw past all that. He envisioned a day when digital computers programmed with rules and facts would possess the intelligence of man. (To read more on Turing and his work with early computers and artificial intelligence, see the *Plus* article *What computers can't do*.)

This boldness and guiding confidence was exactly what researchers needed and thus was born the field of artificial intelligence (AI). In the 1950s and 1960s, the field would see enormous growth and popularity. It became the *topic du jour* of students, researchers, writers, and even the movies.

In the 1960s, when Stanley Kubrick directed his 2001: A Space Odyssey, starring HAL, the omniscient and omnipotent robot, he had taken care to directly consult MIT Professor and AI expert Marvin Minsky, who assured him that, yes, by the end of the 20th century, robots like HAL would not only live among us, but they would exceed us in many capacities.

It was no longer a question of if machines would become intelligent, but when.

A Philosophical Fork in the Toaster

What computers can't do book cover.

"At a time when researchers were proposing grand plans for general problem solvers and automatic translation machines, Dreyfus predicted that they would fail because their conception of mental functioning was naive, and he suggested that they would do well to acquaint themselves with modern philosophical approaches to human being."

What computers still can't do (abstract), Hubert Dreyfus, 1992

In 1973, Berkeley philosophy professor Hubert Dreyfus published his book, *What computers can't do*, in which he proposed the exact opposite of what was on everybody's minds: machines, he reasoned as they were progressing now would never, ever, reach the same intellectual capacities as a human.

There is a passage in Dreyfus' book in which he recounts the results of a meeting among the top minds in computer science; there, his (early) report of AI was deemed to be "sinister", "dishonest", "hilariously funny", and an "incredible misrepresentation of history".

But of course, researchers in the AI community would be incensed. They would be, in fact, deeply, unapologetically annoyed.

After all, they'd just spent the last two decades of their lives telling the world what computers could and would do only to have their fundamental beliefs and dreams attacked by of all people a philosophy scholar?

The core of Dreyfus' critique was about rules. You see, a conventional machine is programmed to accept an input and apply a set of rules to produce an output. The idea is that any intellectual activity, whether it be adding numbers, playing chess, translating languages, or disposing of garbage, can be mimicked using a set of rules.

Dreyfus, however, argued that rules by themselves did not contain the necessary information for their application. Suppose we were to design a robot to process the following phrase:

Mary saw a puppy in the window. She wanted it.

What does "it" refer to, the puppy or the window?

But of course, even a child could tell you that it refers to the puppy. But how does a computer know? Does the computer know that puppies are furry, cute, and love to be hugged and touched by children? Can the computer understand that Mary probably doesn't want a silly windowpane?

What if instead the phrase was:

Mary saw a dog in window. She pressed her nose up against it.

Now, "it" refers to the window. But does the computer know that children enjoy pressing their noses against windows? Does the computer know that the puppy is out of Mary's reach, separated by a layer of glass?

Not only does understanding the nature of the word "it" in these sentences require such obvious facts about dogs and windows, but it also requires a certain human element. It requires us to empathise with how Mary may feel. It requires us to understand the physical nature of Mary's body and how she interacts with her environment.

Previously, many AI researchers believed that programming an understanding of language could be done syntactically that is, by appealing only to the rules of grammar and dictionary definitions. But Dreyfus (and linguists such as Noam Chomsky) pointed out that the issue was much, much more complex.

And they were right. AI researchers were already having difficulties producing machines with the common−sense understanding of a four−year old. There were simply too many rules too many rules and each rule leading to more and more rules.

Even the most basic statements and stories simply could not be understood.

So ... is WALL−E dead?

But what does this all mean for poor WALL−E? Did Hubert Dreyfus destroy the dream of ever producing a WALL−E? Is true artificial intelligence unlikely to ever happen?

No, no, and no!

Dreyfus never intended his original critique to be a crushing blow to artificial intelligence. The dream continues to live on, but today, researchers are older and wizen by his words. The field is no longer as naive and wide−eyed as it was half a century ago.

For example, one possible avenue for modern AI research is provided by our own brains: Instead of programming a computer to abide by the traditional step−by−step rules approach, we model it like the neurons in the human brain where the results of the program depend on the "strengths" of each particular neuron.

This radically different method of computing not only combines the work of psychologists and cognitive scientists in understanding how the human mind works, but also biologists and neuroscientists who study the physical brain, and finally, mathematicians and computer scientists, who work to develop the models for artificial neural networks.

Thus, if artificial intelligence is to succeed if WALL−E is to ever exist we know now that it is going to take the work of all of us of mathematicians, computer scientists, cognitive scientists, philosophers, and psychologists. The dream of imbuing a machine with an intellect if it is ever to happen will be the crowning achievement of not any one discipline, but of humankind as a whole.

Post a comment on this story.

About the author:

Phil Trinh was once asked by his English teacher what he'd be doing after high school.

"I want a PhD in maths," he said, nodding vigorously.

"Really?" the teacher asked incredulously.

"People actually do that? PhDs in maths? Dear Lord, what on earth for?"

Phil is now a doctoral student at Oxford's Mathematical Institute. He remembers this particular exchange fondly. So fondly, in fact, that even now − years later − he feels like slapping his forehead and muttering, "Gagh!"

Phil won the university student category of the *Plus* new writers award 2008 with his article *Maths on a plane*.