THE CASE AGAINST A GENERAL AI IN 2019

Byron Reese

2019 will be a big year for AI.

The technology has finally reached a point where it both works well and is accessible to a wide range of people. We have three things to thank for this: fast computers, lots of data (provided via cheap sensors) and perhaps most importantly, toolkits which make the basic tasks of AI vastly easier. All of this has converged in a relatively short time to bring us to where we are: A blue ocean of opportunity to apply this new technology to all kinds of problems and questions. I don't think it's an exaggeration to say that if all AI advances were to suddenly stop forever, it would take us decades to apply what we presently know to all of the places it could be used. That's why 2019 will be such a big year.

However, 2019 will also be a disappointing year for AI. All of the things we see it do in movies are still far away. 2019 won't see an AI pass the Turing Test, that is, trick a person chatting with it into thinking it is a human. In fact, no Turing Test candidate I have encountered gets past my first question, one any four-year-old can answer: "What is bigger, a nickel or the sun?"

Why this disconnect? Why will 2019 be a big exciting year for AI and a disappointing one as well? Because we use the term AI to describe two completely different things. The first is narrow AI. That is a computer program that can do one simple thing, like play chess or identify spam. That is what we will see make huge strides in 2019.

The other thing we mean by AI is often called artificial general intelligence, or AGI. That is an AI as versatile as a human being. It is creative and can teach itself any new skill. That's what we see in the movies, and that is the technology that almost seems to be moving farther away as we discover just how difficult it is to make a device that can only do simple math be as smart and creative as a human. Estimates on when we might get an AGI vary tremendously, from five to 500 years, which isn't all that helpful. Some people even say that it is impossible—literally impossible.

Impossible? That's a pretty bold claim. Few things are thought to be truly impossible in our modern world of technological marvels. The use of the word in books has fallen steadily over the last century as what was once thought impossible became commonplace and easy. But some things probably are impossible, such as traveling back in time. But what about AGI? Is it possible? Let's examine the case both for an against AGI.

The Case for AGI

Those who believe we can build an AGI operate from a single core assumption. While granting that no one understands how the brain works, they firmly believe that it is a machine, and therefore our mind must be a machine as well. Thus, ever more powerful computers eventually will duplicate the capabilities of the brain and yield intelligence. As Stephen Hawking explains:

"I believe there is no deep difference between what can be achieved by a biological brain and what can be achieved by a computer. It therefore follows that computers can, in theory, emulate human intelligence—and exceed it."

If nothing happens in the universe outside the laws of physics, then whatever makes us intelligent must obey the laws of physics. And if that is the case, we can eventually build something that does the same thing.

Consider this thought experiment: What if we built a mechanical neuron that worked exactly like the organic kind. And what if we then duplicated all the other parts of the brain mechanically as well. This isn't a stretch, given that we can make other artificial organs. Then, if you had a scanner of incredible power, it could make a synthetic copy of your brain right down to the atomic level. How in the world can you argue that won't have your intelligence?

The only way you get away from AGI being possible is by invoking some mystical, magical feature of the brain that we have no proof exists. In fact, we have a mountain of evidence that it doesn't. Every day we learn more and more about the brain, and not once have the scientists returned and said, "Guess what! We discovered a magical part of the brain that defies all laws of physics, and which therefore requires us to throw out all the science we have based on that physics for the last four hundred years." No, one by one, the inner workings of the brain are revealed. And yes, the brain is a fantastic organ, but there is nothing magical about it. It is just another device.

The only way you get away from AGI being possible is by invoking some mystical, magical feature of the brain that we have no proof exists.

Since the beginning of the computer age, people have come up with lists of things that computers will supposedly never be able to do. One by one, computers have done them. And even if there were some magical part of the brain (which there isn't), there would be no reason to assume that it is the mechanism by which we are intelligent. Even if you proved that this magical part is the secret sauce in our intelligence (which it isn't), there would be no reason to assume we can't find another way to achieve intelligence.

Thus, this argument concludes, of course we can build an AGI. Only mystics and spiritualists would say otherwise.

Let's now explore the other side.

The Case Against AGI

A brain contains a hundred billion neurons with a hundred trillion connections among them. But just as music is the space between the notes, you exist not in those neurons, but in the space between them. Somehow, your intelligence emerges from these connections.

We don't know how the mind comes into being, but we do know that computers don't operate anything at all like a mind, or even a brain for that matter. They simply do what they have been programmed to do. The words they output mean nothing to them. They have no idea if they are talking about coffee beans or cholera. They know nothing, they think nothing, they are as dead as fried chicken.

A computer can do only one simple thing: manipulate abstract symbols in memory. So what is incumbent on the "for AGI" camp is to explain how such a device, no matter how fast it can operate, could, in fact, "think."

We casually use language about computers as if they are creatures like us. We say things like, "When the computer sees someone repeatedly type in the wrong password, it understands what this means and interprets it as an attempted security breach."

But the computer does not actually "see" anything. Even with a camera mounted on top, it does not see. It may detect something, just like a lawn system uses a sensor to detect when the lawn is dry. Further, it does not understand anything. It may compute something, but it has no understanding.

We use language that treats computers as alive colloquially, but we should keep in mind it is not really true. It is important now to make the distinction, because with AGI we are talking about machines going from computing something to understanding something.

Joseph Weizenbaum, an early thinker about AI, built a simple computer program in 1966, ELIZA, which was a natural language program that roughly mirrored what a psychologist might say. You make a statement like "I am sad" and ELIZA would ask, "What do you think made you sad?" Then you might say, "I am sad because no one seems to like me." ELIZA might respond "Why do you think that no one seems to like you?" And so on. This approach will be familiar to anyone who has spent much time with a four-year-old who continually and recursively asks why, why, why to every statement.

When Weizenbaum saw that people were actually pouring out their hearts to ELIZA, even though they knew it was a computer program, he turned against it. He said that in effect, when the computer says "I understand," it tells a lie. There is no "I" and there is no understanding.

His conclusion is not simply linguistic hairsplitting. The entire question of AGI hinges on this point of understanding something. To get at the heart of this argument, consider the thought experiment offered up in 1980 by the American philosopher John Searle. It is called the Chinese room argument. Here it is in broad form:

There is a giant room, sealed off, with one person in it. Let's call him the Librarian. The Librarian doesn't know any Chinese. However, the room is filled with thousands of books that allow him to look up any question in Chinese and produce an answer in Chinese.

Someone outside the room, a Chinese speaker, writes a question in Chinese and slides it under the door. The Librarian picks up the piece of paper and retrieves a volume we will call book 1. He finds the first symbol in book 1, and written next to that symbol is the instruction "Look up the next symbol in book 1138." He looks up the next symbol in book 1138. Next to that symbol he is given the instruction to retrieve book 24,601, and look up the next symbol. This goes on and on. When he finally makes it to a final symbol on the piece of paper, the final book directs him to copy a series of symbols down. He copies the cryptic symbols and passes them under the door. The Chinese speaker outside picks up the paper and reads the answer to his question. He finds the answer to be clever, witty, profound, and insightful. In fact, it is positively brilliant.

We use language that treats computers as alive colloquially, but we should keep in mind it is not really true.

Again, the Librarian does not speak any Chinese. He has no idea what the question was or what the answer said. He simply went from book to book as the books directed and copied what they directed him to copy.

Now, here is the question: Does the Librarian understand Chinese?

Searle uses this analogy to show that no matter how complex a computer program is, it is doing nothing more than going from book to book. There is no understanding of any kind. And it is quite hard to imagine how there can be true intelligence without any understanding whatsoever. He states plainly, "In the literal sense, the programmed computer understands what the car and the adding machine understand, namely, exactly nothing."

Some try to get around the argument by saying that the entire system understands Chinese. While this seems plausible at first, it doesn't get us very far. Say the Librarian memorized the contents of every book, and further could come up with the response from these books so quickly that as soon as you could write a question down, he could write the answer. But still, the Librarian has no idea what the characters he is writing mean. He doesn't know if he is writing about dishwater or doorbells. So again, does the Librarian understand Chinese?

That is the basic argument against the possibility of AGI. First, computers simply manipulate ones and zeros in memory. No matter how fast you do that, that doesn't somehow conjure up intelligence. Second, the computer just follows a program that was written for it, just like the Chinese Room. So no matter how impressive it looks, it doesn't really understand anything. It is just a party trick.

It should be noted that many people in the AI field would most likely scratch their heads at the reasoning of the case against AGI and find it all quite frustrating. They would say that of course the brain is a machine—what else could it be? Sure, computers can only manipulate abstract

symbols, but the brain is just a bunch of neurons that send electrical and chemical signals to each other. Who would have guessed that would have given us intelligence? It is true that brains and computers are made of different stuff, but there is no reason to assume they can't do the same exact things. The only reason, they would say, that we think brains are not machines is because we are uncomfortable thinking we are only machines.

They would also be quick to offer rebuttals of the Chinese room argument. There are several, but the one most pertinent to our purposes is what I call the "quacks like a duck" argument. If it walks like a duck, swims like a duck, and quacks like a duck, I am going to assume it is a duck. It doesn't really matter if in your opinion there is no understanding, for if you can ask it questions in Chinese and it responds with good answers in Chinese, then it understands Chinese. If the room can act like it understands, then it understands. End of story. This was in fact Turing's central thesis in his 1950 paper on the question of whether computers can think. He states, "May not machines carry out something which ought to be described as thinking but which is very different from what a human does?" Turing would have seen no problem at all in saying the Chinese room can think. Of course it can. It is obvious. The idea that it can answer questions in Chinese but doesn't understand Chinese is self-contradictory.

Is there some spark that makes human intelligence fundamentally different than machine intelligence?

Where does all of that leave us?

Now that we have explored the viewpoints of both camps, let's take a step back and see what we can conclude out of all of this.

Is there some spark that makes human intelligence fundamentally different than machine intelligence? Do we each have some *élan vital* that animates our reasoning that machines simply do not have? Is there some X Factor we aren't even aware of that is the source of human creativity? The answer is not obvious. Consider how Rodney Brooks, the renowned Australian roboticist, views a similar question. He thinks there is something in biology about living systems that we simply don't understand. Something really big. He termed this missing something as "the juice" and described it by talking about the difference between a robot trapped in a box who methodically goes through a series of steps to escape, versus an animal that very desperately wants to free itself. Robots, in his view, lack passion for anything, and that passion (the juice) is vitally important and meaningful. (Brooks, by the way, is convinced it is purely mechanistic and categorically rejects the notion that "the juice" as some attribute beyond normal physics.) What do you think "the juice" is?

To try to get some resolution to the question of the possibility of an AGI, I invite you to answer three yes/no questions. Keep track of how many times you answer yes.

- Does the Chinese Room think?

- Does the Chinese Room or the Librarian *understand* Chinese?

- Whatever you think "the juice" is, could a machine get it? (If you don't think it exists at all, count that as a yes answer).
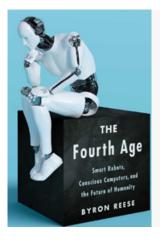
The more times you answered 'yes,' the more likely it is that we will build an AGI. Most of the guests on my "Voices in AI" podcast would answer all those questions with a yes.

There is no middle ground here. AGI is either possible or it isn't. The chasm that divides the two viewpoints couldn't be wider because it has to do with our core beliefs about the nature of reality, the identity of the self, and the essence of being human. There is no real way to bridge the gap on the question of AGI between those with different views on these questions.

But we can at least understand why the views are so different. It turns out that the reason why smart, knowledgeable people come to vastly different conclusions is not because one party has some special knowledge, but because people believe different things.

People don't disagree as much about technology as they do about reality.

Info

Ready to dig deeper into this idea?
Buy a copy of *The Fourth Age*.

Want copies for your organization or for an event?
We can help: customerservice@800ceoread.com 800-236-7323

About the author

Byron Reese is the CEO and publisher of the technology research company Gigaom, and the founder of several high-tech companies. He has spent the better part of his life exploring the interplay of technology with human history. Reese has obtained or has pending patents in disciplines as varied as crowdsourcing, content creation, and psychographics. The websites he has launched, which cover the intersection of technology, business, science, and history, have together received over a billion visitors. He is the author of the acclaimed book, *Infinite Progress: How Technology and the Internet Will End Ignorance, Disease, Hunger, Poverty, and War*. Byron's newest book, *The Fourth Age: Smart Robots, Conscious Computers, and the Future of Humanity*, is described as "entertaining and engaging" by *The New York Times*, was named this year on JP Morgan's Summer Reading List and is a Gary's Book Club pick for CES 2019. He lives in Austin, Texas.

Share this

Pass along a copy of this manifesto to others.

→

Subscribe

Sign up for e-news to learn when our latest manifestos are available.

→