Riccardo Manzotti

# Can Machines Think?

## The question

Can machines think? It's a classic question. The straightforward answer is yes—they already do. At most, we listen to someone's words, but we have never seen a thought.

What about our own thoughts? We say: "I think," "I'm worried," "Last night, I couldn't sleep because I had so many thoughts—about my mortgage, about my car that was parked where it shouldn't have been," "The thought of the car kept me awake."

But is that true? Was it really the thought of the car, or was it the car itself that kept one awake?

When we say that we think of something, what is really taking place? Can we describe what a thought is without any reference to what the thought is about? What about a naked thought, an empty thought? Is it something, or is it nothing?

It is widely accepted that thoughts are always thoughts of something. We cannot think of just anything. If we think of nothing, we do not think. So whenever there is a thought, there is also an object of the thought.

Thoughts by themselves—naked, empty, contentless—simply do not exist. There is no way we can conceive of a thought without its reference, without its object. And that is something intriguing.

Because it suggests that maybe this very notion that we think—like The Thinker by Rodin—might be a form of superstition. In the past, there were animistic superstitions, and people attributed all kinds of magical properties to objects that were invisible. Today, it might be that the very superstition we need to get rid of is ourselves.

# The inner world

The idea that we have an inner world that nobody has ever seen is something to think about. How is it possible that we all believe in something no one has ever seen? And when people say, "Look inside yourself," how do we know in which direction to look?

This is suspicious because, in the past, there were many anthropological and philosophical studies that explored metaphorical directions. For example, the concept of heaven has been embodied in our language: "We feel uplifted," "We feel up," "We feel high." That direction is important in our language. But nobody believes that all those metaphors really mean that there is something upward apart from Elon Musk's satellites.

In Italy, there's a beautiful painting by Tiziano with the Assumption of the Virgin Mary. According to the Catholic religion, the Virgin Mary never died. She had been carried up in the heavens before. In this beautiful painting of the 1550s, angels are physically lifting the body of the Virgin Mary in the sky. At that time, people took that direction, the idea that heaven is upward. Today, even people of a religious background no longer believe that heaven is physically upward and that hell is physically downward. But we still have one metaphorical direction that we all still believe in. This metaphorical direction is inner – the inside.

Where do we get that idea? There is an exact date when we started to think about the inner world. It happened in Milan, Italy. At that time, the Roman Empire wasn't in particularly good health. A famous saint, Saint Augustine, wrote the first book in first person. The book was called “Confessions”. In the book Saint Augustine plays a philosophical and literary trick. Christianity, until that point, didn't have the immaterial soul. They believe in the resurrection of the physical body. Saint Augustine had the problem of putting the soul somewhere. And he had the problem because in the physical world, there was nothing like a soul. In the physical world, everything is filled with physical stuff. If we open a body, we find physical stuff: organs, cells, blood. There is no space for a soul.

Saint Augustine was very smart, and felt that the word inside didn't work for what he wanted to find because he needed to find an empty space that was outside of the physical world. What did Saint Augustine do? He took the Latin word for inside “internus” and made the comparative, the greater, which is basically “more inside”.

Then he made a noun of the word that in Latin means “more inside”. This word became the word that today we use in many languages: in French it is “interiorite”, in Italian – “interiorità”, in English – “interiority”. But it is most commonly used with the expression “inner world”.

Saint Augustine invented the idea that we have an inner world. “Inner” means that it is more inside than everything else which is already inside. It is the idea that we have a kind of niche, or shelter, or bubble that is detached from the external world. And the idea had a big success because people liked to have a place where they could be safe from all the harsh realities of life.

Same idea had been developed even more by Rene Descartes during the Thirty Years war. He said that we are something that thinks. We are not sure about the world, but we are sure that we think. And how can we be sure of something that cannot be seen or demonstrated in any way?

## Why shouldn't a machine think?

That's why, many years later, Alan Turing asked the very question we started with: How is it that we think, and why shouldn't a machine do the same? It is surprising, that Alan Turing in this paper was also addressing theological issues. When he wrote that very famous paper, his dean complained about Alan Turing and told him not to write anything like that again – “This is not serious science”. The dean was Charles Darwin III – grandnephew of the great Charles Darwin.

Today there is a machine that has reached, in many ways, human capacities. The last version of GPT has actually a little bit more – 1,8 trillions of parameters. The estimated amount of connections we have in the temporal area of the brain that deal with the language is between 1 and 10 trillion connections. Other parameters are quite interesting too. The 2000k conceptual space of these large scale language models. Why is it worth mentioning? Because the number of conceptual spaces in the abstract mathematical space of language models is just like the number of words we have in human languages.

In a very big dictionary we have more or less that number of words, and that is not by chance. The reason is that these parameters are beginning to reach exactly the same kind of complexity from a quantitative perspective that our brain has when

dealing with language. AI today is what Isaac Newton would have described as a philosophical experiment. Actually, Isaac Newton described his own work as experimental philosophy, a way to test experimentally our core intuition about ourselves and reality. Today, AI is doing exactly that.

When we ask the question, can machines think, it is the time to ask the mirror question, to use AI as a mirror of a mirror. Because we are the first mirror. We reflect. We speculate. We are a mirror of the world, and AI is a mirror of ourselves.

We may start to wonder whether the fact that AI is able to do the same things that we do with language tells us something very deep about ourselves. We have always believed that we have an inner world, which is the modern version of the soul. Do we really have an inner world? Do we really or AI allows us to take a sharper turn and to get back to philosophy?

## Getting back to philosophy

The goal of philosophy is always to get back to the origin, and then to change our insights about reality. It's not to go forward, but to get back to the origin of everything. What is the origin of everything in this case? It is a famous dialogue done in 390 BC by the only philosopher that Plato didn't mention by name – the Stranger. It's quite intriguing that there is someone that Plato didn't want to mention explicitly. Why? Why was this guy so dangerous for Plato? Because this guy was against the notion of an inner world, was against the notion of ideas, was against the notion of thoughts in a way, thoughts as abstract or formal entities that are not in the world. This guy, the Stranger, is usually referred to as the Eleatic philosopher.

They were philosophers coming from Elia, who were against the idea of abstract or mental ideas. They said: everything takes place in the world, everything is an object. We do not need those objects. We have the objects. What is our body then? Our body becomes a way in which we empower the world. Our body becomes a way through worlds that are real: not inner worlds, but actual worlds. They are my car, not the thoughts of my car. They are able to become causally effective. The principle that this philosopher claimed against Plato was that something is real only if it produces an effect.

There is a male only if there's a female. An object is a key only if it unlocks a lock. Anything is what it is only if it is able to produce a real effect. This view Plato put aside. Because he wanted to find a kind of shelter for the mind, a cave, a refuge, a place where we could be safe from decay, change and death. And Plato invented the idea of the soul. And the idea of the soul afterwards becomes the idea of the mind, the inner world. But is there a better reason to believe in an inner world than there is to believe in the soul? Actually, there is none.

After all, the paradigm shift brought about by this experimental philosophy, called artificial intelligence, is pushing us to question our beliefs about what we are—and that may actually be helpful. It may allow us to improve, to mature. It may compel us to take a step forward, no longer needing the idea that we are sheltered inside an invisible and unproven inner world.

The two most famous sentences written by human beings are: 'To be or not to be,' from Hamlet by Shakespeare, and 'I think, therefore I am,' by René Descartes.

“I think therefore I am” – is it true? Do we know this? We have no idea. Have we ever seen AI? We haven't. Actually, one of the most convincing models of the eye is that it is the center of narrative gravity of our own existence, a concept that we develop during our life just to explain what we do. We see what we do, and we need to attribute what we do to a kind of center of narrative gravity. Luigi Pirandello, an Italian writer who won the Nobel Prize in the last century, suggested that there is no such thing as a hidden identity. Instead, he argued that we develop who we are by acting in the external world—by engaging with it—and that we are, in fact, one with the world.

## Goodbye, inner world

With the help of AI, we can take into consideration the possibility of writing a love letter to our inner world. And we need to say goodbye to our inner world.

> *Dear loved one,*
>
> *My dear interiority, my dear, my dearest inner world, I did everything for you.*

*My love for you was unmatched for so long. You were the love of my life. I know why I stayed together for so long.*

*Our relationship was familiar, comfortable, and all that I wanted to know. I clanged for so long to the hope that you would become the thing I needed you to be. I felt so good with my inner world, which shielded me from the harshness and dangers of the world. Moreover, thinking of myself as a thinking subject filled me with pride and narcissism. Thanks to you, I was the center of the universe, the only thing of value in my world, and I could imagine always having a safe haven to retreat to you, conscious mind, inner world, my beloved one.*

*I was afraid to lose you. I feared there wouldn't be anything better out there for me, but I realized that I was hurting myself by wasting my time. I am so sad and disappointed by our relationship ending. But the time has come to let you go, and in doing so, to enter the world, or rather to become the world, I'm no longer a who, no longer a thought, no longer an inner world, no more "I think therefore I am", but rather just I am. In fact, I am the thing that I am.*

*I no longer have a barrier between myself and the world, no longer the possibility to step back from reality. I am thrown into the world. In fact, I am the world. The card "I think, therefore I am" becomes Heidegger's being in the world, or rather, it becomes being the world or even being a world. Farewell, safe harbor of my inner world.*

*Farewell, ontological privilege. Farewell, adolescence, ontological adolescence that led me to believe I was special. Starting tomorrow, it will be a thing among things, and I will have to confront the harsh reality. Leaving you, my mind, or thought, my inner world is painful, but it is necessary to mature and enter the world.*

*Goodbye, my adolescent love.*

I wrote it this way because I believe we are mostly attached to the notion of an inner world for sentimental reasons. After all, there is not a shred of evidence for the existence of an inner world or for the existence of thoughts. Now what could this tell us? First, I will take advantage of the theatrical version of Galileo. In Life of Galileo by Bertolt Brecht. In that play, Galileo makes a very important statement. He asks: "What is the biggest mistake of science?" Mirroring the work by Thomas Kuhn in the structure of scientific revolution. "What is the biggest mistake that we make in

science? It is to believe in something we don't really know. That's the biggest mistake."

We once believed that the Earth was at the center of the universe — but it isn't. We believed that human beings were a special species among animals — alas, that wasn't true. We believed that space and time were absolute — alas, that wasn't true either, and so forth.

What is something we believe today that might not be true?

It is easy to look back and judge people from the seventeenth century — how foolish they were! But there were many reasons for them to believe what they did.

Today, we believe that in order to think, we must have an inner world. And because of this, we assume that for machines to think, they must also have one.

This thing is the idea that we are separate from reality, the idea that we are not the world. Let's get back to thinking. What is a thought? We don't have a positive way to either describe or to point to thoughts. We don't have a way to say that a thought is a thing which is made in a certain way, which can be seen using certain tools, which has a certain shape, which consumes a certain amount of energy. We cannot do anything like that.

There is neither a scientific way, nor a phenomenological way to refer to a thought. What is the thought of X, the thought of an object, the thought of a car? It is not an object. It is not a car. It is not an X.

And likewise, what is the self? What are we? We are not a thing. But to not be a thing means that we are nothing. We are not a thing. We are nothing. The subject, historically speaking, has not been defined on top of some positive evidence, experience, fact. It has been defined and built on top of the negation of reality because we didn't want to be the world. Because to be in the world is too painful. There are too many bad things in the world. The world decays. There is death, illness, injustice, all kinds of things that make us suffer. And so we want to have a barrier. We want to have a shelter. We deny the reality that we don't like, and we historically develop the notion of the inner world.

## Where lies the difference?

Let's get back to Galileo. Galileo, in 1590, wrote a famous poem to explain how he was going to do his scientific revolution before doing it. He told us how to proceed when we have to deal with something unexpected. At that time, there was the Copernican revolution. Today, there is the AI revolution.

What was suggested by Galileo? In his opinion, anyone who wishes to discover something must deploy imagination, and play with invention and guesswork. And if you can't go straight ahead, a thousand other paths might help you along. Nature seems to teach us this: when one cannot follow the usual path, they take a different road. The style of invention is very diverse. But as I have tested in seeking the good, one must proceed contrary wise. And that's what I am suggesting to you. I am suggesting to you not to look for whether machines think, but to ask yourself, what do we do?

Thinking is merely mirroring the causal structure of the world. This is done primarily through language, because language allows us to reflect the structure of reality. That's why it is a revolution. Don't believe the claim that AI doesn't think — because what exactly do people do that AI does not? When we say that inside AI, there are just statistical, parameters, bias and probabilities mirroring the causal structure of language, exactly what do our neurons do that AI doesn't? Because one neuron is not much smarter than a conditional probability of one parameter inside AI. It is true that at this point, the causal structure of AI may not be exactly like ours. It is also true that at this very moment, the biggest limitation of AI is human beings because AI so far has had no access to the world. It has to go through our description of the world. But how much smarter would a human being be if he or she had no access to reality, but only to other human beings' description of reality.

Therefore, this compels us to take seriously Galileo's suggestion that we must consider entirely new perspectives. We must proceed in the opposite direction because the world has changed dramatically.

If AI is doing what we do—namely, using language to mirror the causal structure of the world—then, as the English writer said: "How could I know what I think if I don't say it?" Likewise, how could we know what we are if we don't act?

If AI is mirroring the causal structure of the world and using language just as we have for the last twenty thousand years—if it has 'hacked the operating system of

humankind'—then is there anything we have that AI does not? It is one thing that has two names. They are apparently different, but they mean the same thing.

We have freedom and values. A practical example is if we have to go back home with the car, what is the first thing that we would do? We would take your navigator, Google Maps, and we would ask Google Maps what is the best route to follow from point A to point B. Is that a value? Is that a choice? No. Because the true value comes before. When you have to choose whether you want to do the shortest, the briefest, the cheapest, the most ecological route. And that is the value on the basis of which a causal structure like that of AI is able to compute the best route.

That is knowledge. Knowledge is the capacity to use the causal structure of reality to reach a goal. We have a different capacity so far. We have the capacity to freely choose a goal. We have the capacity to choose what's worth pursuing. And that's what we have that at this very moment, machines do not have. We have the capacity to create value. How? Choosing it by acting with our own freedom. And this is a lesson that AI is teaching us. It's telling us that life is not just having knowledge. It's not just having the capacity to compute from a cognitive, efficient perspective the best route. It's not only the capacity to use language. It is the capacity to know why we use language.

Like Jack London famously said: “I don't want to waste my days trying to prolong them”. There has to be something that we choose for which it is worth using all the knowledge that we have. This is what AI today does not have.

We are not a negative entity. We are physical worlds that, by means of our body, are able to choose what they want. And knowledge is a tool, the most powerful tool, but it is a means. It is not the end.