Translation of “No veo impedimento para que la inteligencia artificial haga lomismo que la humana” published in *La Vanguardia*, June 25, 2022. (Click here for original article.)

"I see no impediment for artificial intelligence to do the same as human intelligence"

Computer engineer and philosopher

Judaea Pearl

Judea Pearl (Tel Aviv, 1936) laughs when he is introduced as the father of artificial intelligence (AI) and rejects the label. However, he compares the development of machines to raising children and believes that computers should be seen as a new generation of children to be trained and educated in the hope that they will conform to the value system of their parents. , the humans.

And it is in this line that the contribution of this engineer and philosopher stands out: he has translated probabilistic reasoning and cause-effect relationships into mathematical language so that robots learn to think like people do and understand the reality of their environment instead of just accumulating data as is the case with AI based on deep learning.

His conceptual and mathematical contributions have earned him numerous outstanding awards, including the BBVA Foundation Frontiers of Knowledge Award in Information and Communication Technologies.

Can you give me a practical example, an application of your developments?

An extraordinarily promising area is personalized medicine and personalized decision making. We know how to take data from different sources and offer information about a specific individual; not about the population or an average, but about whether you, for example, are going to benefit from a drug or if that drug is going to fail you. And it's already being used to test theories, to quickly find out whether or not a theory is supported by the data.

The layman is amazed at what the AI can already do and fears being overtaken and displaced by it. What do you think?

I would not say that we should be afraid of artificial intelligence, but we do have to observe it, be vigilant. It's like raising children: there is a chance that a child of ours will become a Putin, but also that he will absorb the value system that we share.

Who should establish that value system for the machines?

Philosophers have some theories of value and psychologists too. But in artificial intelligence we have a better one because we know how to apply a system of values to machines, so finally it will be the computer scientists who transmit to the robots a code of values adapted to their condition.

Will each scientist apply his values or will they agree?

I think there should be conferences and committees created to define the value system and how to communicate it, with the participation of AI scientists but also social scientists and philosophers. But for that they have to learn our language, because it is a waste of time to talk about Aristotle, he didn't have a computer.

You have allowed machines to establish cause-effect relationships. Will they one day also be able to reason and imagine?

Today they already do. Establishing the cause-effect relationship is reasoning, and the counterfactual level (thinking about alternatives that could have happened and didn't happen) is your imagination.

You seem to assume that whatever we humans can do, machines will be able to do. And feel and understand emotions?

Yes, and I'm not the first person to say that. Marvin Minsky explained it in The Emotion Machine (Debate). Emotions are easy to program, what is not so easy is to control them.

Does that mean that robots will be able to replace people? In what term?

I don't know how long it will take, but the answer is yes, in many areas. I don't see any kind of theoretical impediment to this happening. I don't know when it's going to happen, I'm not a futurist, I'm an incremental scientist, I only advance what I can demonstrate.

And shouldn't we set limits on their development before they overtake us?

I don't think we can regulate anything now because right now we don't know what we have to regulate, it's premature. Those who want to regulate first have to understand the systems.

The point is that today algorithms are already used to make decisions that people do not understand or know how to explain. Doesn't that entail a lot of risk? Who will be held accountable for those decisions?

There are two questions here. The first is to make a machine explainable, and that is where my cause-effect theory comes into play. You cannot have a system that explains its recommendations unless the machine has a causal model of the world, that the machine understands what the implications of the decisions are in reality, not just at the data level. And today what we have and are using are big data machines, which do not understand the world, only the data, and that is the biggest obstacle we have today. Once the machines are equipped to understand reality we will have an explainable system.

And regarding the responsibility of the decisions of the AI?

It is a combination of responsibilities. If you put the wrong model of reality in a machine and, for example, make it work based on supernatural forces so that its recommendation to get a loan is 'go pray', the person responsible is the programmer who has assumed the risk

of describing reality in terms of supernatural forces. But in addition to the model, the correct logic must be applied to connect it to the data, so both the person who provides the model of reality and those who know the logic that connects it to the data are responsible.

Would you allow yourself to be cared for by a robot or would you leave decisions about your work or your health in the hands of an AI?

If the robot reads my book, yes.

?

Because causal artificial intelligence is much more transparent and understands the implications of its decisions.

And what can citizens do to reduce the risks that the expansion of artificial intelligence systems entails?

Learn some technology. All people concerned with the social implications of AI should learn philosophy of science, epistemology, how we acquire knowledge, what makes us believe that what we know is true, how we can communicate knowledge so that we can speak coherently.