Synopsis

ARTIFICIAL INTELLIGENCE:

DREAMS, RISKS, AND REALITY

March, 7th and 8th, 2017

COSMOCAIXA BARCELONA. ISAAC NEWTON, 26. BARCELONA

www.bdebate.org

CO-ORGANIZED BY:

iBE CSIC upf. | upf. Universitat Pompeu Fabra Barcelona | CSIC Consejo Superior de Investigaciones Científicas | ICREA

Artificial intelligence: dreams, risks and reality

There are many definitions of artificial intelligence out there, which hints at its slippery nature. One approximation would be: a computational system capable of emulating the functions of the human brain. Expressed in popular imagery as robots, in reality artificial intelligence goes much further and can already be found in many areas of our lives. Although there are some limitations, AI is in the selection of news articles we read, conditions or at least influences election results, is behind many of the ads we see and we are on the brink of co-existing with some of its new applications, like driverless cars.

But behind every opportunity there is also risk. To debate both sides, some of the top international experts in the field met for a B·Debate session, an initiative of Biocat and the "la Caixa" Foundation to promote scientific debate. This gathering led to the Barcelona Declaration for proper development and usage of artificial intelligence in Europe.

CONCLUSIONS

- It is difficult to predict how our relationship with robots will evolve. This interaction opens up a new field known as robot ethics.
- Artificial intelligence applied to the massive amounts of data available today is creating echo chambers and filter bubbles on social media, threatening to radicalize ideologies.
- The algorithms artificial intelligence uses are often obscure. They must be translated into our language so that ethical, socially just decisions can be made.
- In light of the many opportunities and dangers, participants drafted the Barcelona Declaration for proper development and usage of artificial intelligence in Europe.

1. ARTIFICIAL INTELLIGENCE: FROM ROBOTS TO DATA

On robots

"Artificial intelligence could wipe out humanity when it gets too clever, as humans will be like ants." Superimposed on top of a photo of a giant robot in The Independent, this quote from physicist Stephen Hawking made its way around the world. The experts, however, didn't give it too much importance. For Héctor Geffner, ICREA professor at Pompeu Fabra University, robots are still clumsy and "**human-level general intelligence in machines only exists in films for now**. They can react but aren't very flexible and don't have any mental life."

Carme Torras, a researcher at the CSIC Institut de Robòtica i Informàtica Industrial, believes that "**today humans are more dangerous than robots**", but defends their current development. Her group is working on care robots that can help dependent people dress themselves, and explained that there are already robots being used in logistics or as receptionists, robots that screen what can go into the trash or the dishwasher, and even pet robots, developed to show and provoke emotion. The latter opens up a whole field of ethics, known as "**robot ethics**", which Torras says is motivated in part by the humanization of robots, the perception that there is *a human inside*." For Torras, the big question is "**How will human nature evolve as a result of increased interaction between humans and robots?** And, by extension, can this evolution be predicted?"

One of the problems in studying this evolution is the limited language available for describing the future. In Heidegger's words, **"It is through technique that we perceive the sea as navigable."** This is why many of the dilemmas are found in **science fiction**, one of the roles of which, according to Torras, is to "anticipate possible scenarios." These can be seen in books like those by **Asimov, Philip K. Dick and Ray Bradbury**, in films like "**Eva**" or series like "**Black Mirror**". Torras herself has written a science fiction novel entitled "**La mutación sentimental**", through which she is developing an educational project.

On data

The increased computing power of computers and, above all, the advent of the Internet and the exponential increase in data available have led **artificial intelligence to permeate many aspects of our lives**. Algorithms are behind much of the news we

read and the ads we see, but it can also be behind our credit rating or loan approval, the first screening of our resumes or studies of some of our health records. And it isn't outlandish to think that it will also be guiding our vehicles, which in the future will surely be totally driverless.

One of the most widely studied areas is social media and access to information, which has noteworthy implications and dangers. For example, in **effecting elections**, as it has been suggested happened by using artificial intelligence to personalize the messages for Donald Trump's campaign.

Apart from one-off campaigns, according to Cornelius Puschmann –a researcher at the Hans Bredow Institute for Media Research in Hamburg, algorithms "are a confusing way people sometimes speak about artificial intelligence," and are responsible for "choosing" –based on our profile and prior preferences– which news articles tend to appear on our social media walls. There are two trends that algorithms and networks have brought to the forefront, neither of which is exempt from danger. The first, as noted Camilo Cristancho, a researcher at the Autonomous University of Barcelona, is the creation of **"echo chambers"** in which people with similar ideas tend to stick together. The other is "**filter bubbles**": ideological bubbles created as a result of the contents each person is shown being customized based on their past preferences, which threatens to over-promote established beliefs, minimizing nuance and plurality.

Should we be worried about filter bubbles? Not too much, says Puschmann. "From a journalistic point of view, this is nothing new. We've been studying it for 50 years." For Walter Quattrociocchi, head of the Laboratory of Computational Social Science at IMT Lucca in Italy, "It is still up for debate whether these bubbles are created by the algorithm or are just human nature."

According to Puschmann, what seems clear is that **fear comes from the feeling of losing control,** the fact that something non-human has decision-making abilities. In an experiment with Facebook users, more than half weren't aware that there is an algorithm behind what they see on their walls, and their first reaction when they found out was shock and anger. However, several months later, after being shown how they work, their satisfaction levels were similar to those seen before "discovering" this fact. And they used Facebook even more than before.

Quattrociochi mainly researches how misinformation spreads online, which he ties to a new word coined in 2016, **post-truth**. This concept "denotes circumstances in which objective facts are less influential in shaping public opinion than appeals to emotion and personal belief." As well as what is known as the confirmation bias: "the tendency to search for information that confirms pre-existing beliefs." Even the World Economic

Forum has pointed to mass digital misinformation as "one of the main risks for our society."

Quattrociochi's research confirms that echo chambers clearly exist, in which mutual-reinforcement dynamics are similar regardless of the shared motivation. Perhaps the most concerning is that the most active and "committed" users in these chambers tend to be **those defending conspiracy theories** (like those who believe in UFOs or a link between vaccines and autism). The following paradox is even more concerning: publishing rigorous information seeking to discredit the theories in these echo chambers not only doesn't work, but is counterproductive: it only strengthens their initial position.

2. LIMITATIONS AND ETHICS OF ARTIFICIAL INTELLIGENCE

Despite the achievements and promises of artificial intelligence, it still faces serious, important limitations. Basically, programs can work very well for specific applications, but are **far from being able to do so in the unpredictable real world**. For Francisco Martín, president of BigML, techniques have hardly changed in recent years. They've merely increased the data they can work with. According to Martín, "a lot of people are reinventing the wheel at the same time," without making any real advances. In fact, the processes "still require a lot of human experience and are done in a very manual way."

One of the limitations has to do with the obscurity of the process programs use to reach their conclusions. Many of them are based on what is known as **neural networks**. These networks are mathematical constructions modeled on the human brain, with information moving between levels and becoming consolidated in a diffuse manner. They are very powerful models, but necessarily opaque, and have led people to speak of "**the black box of artificial intelligence**". Some experts, like Pierre Baldi, don't give any importance to this obscurity. In the end, "You use your brain all the time. You trust your brain all the time. And you have no idea how your brain works." Others, like Marcello Pelillo, professor of Computer Science at the University of Venice, recognize that there are situations in which "it is **important to be able to explain decisions**." For example, if it affects the decision of a judge or social institution. And, in general, to ensure respect for autonomy and personal dignity.

Because the use of artificial intelligence implies numerous ethical challenges and dilemmas. These are some proposed by Francesca Rossi, researcher at IBM and professor of Computer Science at the University of Padua:

- Would artificial intelligence replace human work?

- How would our interaction with other humans, society and education change?

- In the future, what would happen with the possibility of developing autonomous weapons?

- How will artificial intelligence develop its own ethics?

For Rossi, **systems must be developed to discriminate results ethically**. This would obviously have to be the case for autonomous systems, but should apply to others, too. For example, tools that advise doctors based on data analysis. "If these suggestions don't follow a code of ethics, doctors won't be able to trust the system." The key is how to develop that code, as the program's decision-making processes, as we've seen, are obscured by their own nature in many cases.

It isn't simple. On one hand, due to flexibility. What may seem like **common sense** to us, isn't necessarily so for a machine. "A food processor shouldn't cook the cat if there isn't anything in the fridge, even though it may think that would be an acceptable meal for us," explains Rossi.

One solution would be to program according to professional codes. "But these have many gaps that we fill with common sense. This is difficult to instill in a machine."

Can it be done, though? According to Rossi, yes. "There are several ways that we're studying now." Furthermore, the process has hidden advantages. "Doing so could help us be more aware and behave more ethically ourselves."

3. THE BARCELONA DECLARATION

Prompted by the leaders of the B·Debate Luc Steels, ICREA professor at the Institute of Evolutionary Biology in Barcelona, and Ramón López de Mantaras, director of the Artificial Intelligence Research Institute of Barcelona, those at the event came up with and agreed on the "Barcelona Declaration for proper development and usage of artificial intelligence in Europe", **a manifesto open to new signatories and comments** that aims to lay down guidelines for the progress of artificial intelligence. The manifesto reflects the great importance this field will have in the future of the economy and society, as well as concerns about misuse, premature or malicious use of new technology. Based on this foundation, the signatories propose a code of conduct in six points:

- **Prudence:** Despite the development of artificial intelligence and its possibilities, many fundamental problems are not yet solved and will require radical breakthroughs and perhaps some issues will never be solved.

- **Reliability**: New methods should be verified before they are made available. The new agency for robotics and artificial intelligence created by the European Parliament could potentially take up this task.

- **Accountability:** When an artificial intelligence system makes a decision, humans affected by these decisions should be able to get an explanation why the decision has been made. Nevertheless, many techniques that "explain" the internal processes of these systems are still in their infancy.

- **Responsibility:** There is growing worry about AI chat-bots and other kinds of automatic messaging systems operating online that are capable of manipulating public opinion. It should be obligatory to state when messages come from an artificial intelligence system.

- **Constrained autonomy:** Artificial intelligence systems, when embedded in physical systems like self-driving cars, have the potential to act upon their decisions in the real world. It is necessary to have clear rules constraining their autonomous behavior, as well as clarifying who is responsible for failures.

- **Human role**: Enthusiasm for artificial intelligence sometimes gives the impression that human intelligence is no longer needed. This is a very serious mistake. All AI systems depend on human intelligence, so human expertise should continue to be taught, developed and exercised.

Finally, given the growing importance of artificial intelligence in maintaining a competitive economy, the signatories call on European funding agencies and companies to invest in the development of AI at a scale adequate to the challenge, creating the conditions to promote entrepreneurship. Plus, they **highlight the importance of developing open resources**, available as a common infrastructure on which specific applications are built.