Artificial Intelligence and the Evolution of Truth and Language "Semiotics and AI" Roundtable

Dario Compagno

Artificial Intelligence and the Evolution of Truth and Language

"Semiotics and AI" Roundtable – 09 June 2023

Dario Compagno, University Paris Nanterre (dario.compagno@parisnanterre.fr)

My intervention is about truth, language and their relationship with artificial intelligence and especially with programs such as GPT. I will generically call "chatbots" the kind of software capable of producing human language. I'd like to make one main point today: we should be careful when evaluating chatbots through the prism of our current concepts of language and truth, as we might be about to discover that our concepts of language and truth may evolve thanks to chatbots.

Much everyday talk about chatbots today is based on a mistake: the idea that chatbots are imperfect because they do not always tell the truth, that is, that they make factual mistakes, such as attributing false statements to people, or making reference to books that do not exist–as I recently happened to find myself in a student's work. But chatbots are trained to talk, and particularly to entertain in conversation, not simply to tell what is true. This means that chatbots are much more powerful than truth-telling machines, which have existed since a quite long time.

In order to talk, you need to know things, and more importantly you need to understand things. To understand something means that you can elaborate on it. If a chatbot attributes to Einstein something he never said, but that he could have said, this means that somehow it actually understood Einstein's thought. Any mediocre student can learn to repeat what Thomas Aquinas said, but only a smart student can reason *ad mentem divi Thomae*, as they said in the Middle-Ages, which means to think as Thomas did.

We need to decide: do we want machines who always tell the truth or machines that are smart? Because we have been having the truthful ones for a while now and never found much of a collective use for them, while today that they lie instead we are increasingly afraid that they will steal all our best jobs. That's why chatbots are so interesting: as soon as we have a machine that is capable of speaking reasonably well, it immediately starts lying. It starts elaborating on its acquired knowledge, moving beyond what is and towards what could be but it's not.

Umberto Eco famously said that semiotics is the discipline of lying. This sentence has been used in all kinds of contexts, often taken as a joke or as proof that semiotics is bullshit. But what Eco meant is much deeper. In order to talk, there is a logical need for opening the door to falsehood. It would be senseless to say that something is true if it could not also possibly be false. Aristotle *ipse* famously defined truth and falsity at the same time, as faces of the same coin. So somebody bound to truth only would just see half of the coin, and it's not clear how he, she or it would grasp the idea of truth itself. It would be something like asking a fish what water is.

Chatbots access reality exclusively through discourse. Instead of approaching the world directly through perception (with cameras and microphones) and physical action (with mechanical hands), they only rely on what was already told before. This is why chatbots could be called 'deleuzian' or 'barthesian'. In fact, a chatbot is like a person who has seen all the movies about love, who has read all the books about love, who knows the lovers' discourse by heart, but never loved. And still this works: chatbots are fully capable of interpreting love, that is to produce new signs which refer to precedent ones in a way comparable to what a human could also do. In this sense, they understand love.

For years we tried to build a machine capable of passing a Turing test, fooling us into believing it a human, and now that we have it, some brilliant minds say that it's not fair, because it makes stuff up, it makes mistakes. This is an illusion: chatbots do not make any mistakes, because there is nothing wrong in inventing and lying. Mistakes only exist in narrow frameworks in which there are right and wrong answers. And chatbots are built and trained in order to read and write as humans, not in order to make addictions and subtractions. It is true that sometimes their answers do not make sense, but this has nothing to do with truth. And of course what's really astonishing is that most of the time their replies do make sense. Chatbots are proof that semiotic theory is not entirely bullshit, because machines without a mind and a soul can very well become part of human society.

This is how chatbots work. They are built out of attested contexts, in order to reason in new contexts. The representation they have of a word depends entirely on other words, and nothing else. First and foremost, they look for words that co-occur together, that is, which are often found together in the same sentences, paragraphs or texts, such as for example the words 'good' and 'doggy'. These are the so-called syntagmatic relationships of Ferdinand de Saussure. Then there are those words which are not found together, but that tend to appear in the same kinds of contexts, each as a potential substitute of the other, such as 'pasta' and 'rice'. Saussure would talk about paradigmatic relationships here.
And in fact for chatbots absence is key. They are purely structural and differential, even more than any human being. We humans fill up the sense of words with mental representations. The meaning of the word 'house' for us depends on the real houses in which we have lived, and that's why the concept of 'house' changes a little from one person to the other. Therefore none of us has the pure concept of 'house' in his or her mind, what Edmund Husserl called the meaning or *Bedeutung*. As Jacques Derrida stated very well, such pure meanings are never reached by humans, and we use language to approximate a collective understanding of words. This means, in the end, that there is no unique concept of 'house', independent of the historical sentences in which this word has been used. And that's what chatbots do: they reach consensual meanings even if they do not have mental representation for them. Pure semiotics, no psychology.
This is the reason why another critique often made to chatbots is preposterous. Sometimes I read that chatbots aren't good enough because they only take averages from past linguistic productions, from sentences that have already been said. Now, I do not want to scare you, but if that's a problem, well we as a species are pretty much affected by it too. I'm sorry if I am not inventing new words today, or if what I am saying has already been said, in one form or the other, before me. As Eco loved to tell: nothing of what I say is new, just the order in which I say it, is original. And we are certain that chatbots put words in new orders.

I have defended currently existing chatbots from a couple of attacks. But I believe that we should be prepared to go much further, in the next few years, and see human society itself evolve thanks to its newest members. Given the learning curve we are observing, it may be that we will see for language what we already saw for the game of Go. I mean that a problem that was first considered much beyond the capacities of machines suddenly finds a solution. And once this solution is found, algorithms can outperform humans.
Today chatbots talk, this is not a matter of opinion anymore. And all experts agree that they will learn to talk better and better. Honestly, this is just an engineering problem, and not such an interesting one. The really interesting question is whether chatbots can talk *better than humans*. This is what should gather our attention today, and the answer to this question will

shape the centuries to come. The idea that human language can be improved and evolve is so far from our current understanding that it needs some explanation. Let me make an example by analogy.

AlphaGo is a computer program trained to play Go. Go is a complex game, much more complex that chess. Chess was already ‘solved’ by machines incapable of lying, while Go on the other hand needed something more: AlphaGo needed to learn from human data, from real matches. And eventually, first it learned to play well, and then it started playing *more than well*. It started inventing moves, it started playing creatively. It started making moves that looked silly to human Go masters but were actually incredibly smart and profound. Today nobody, no human being, can beat AlphaGo nor can even understand what AlphaGo does when it plays: its moves are totally opaque. It always wins and we don’t see it coming. Its moves look random, but they are not. Imagine a kid watching a master playing Chess: it may look like the master is pushing wood. The master is not pushing wood. He’s not pushing wood, not because of something in his brain, hidden and private, that we access through his ears, but because of something happening on the chessboard, where everyone can see.
Can we imagine chatbots doing the same thing? Are we able to imagine that after having swallowed the entirety of human knowledge in its written form, they start speaking *better than us*? Understanding what this last sentence means could become the future of semiotic research, and possibly also that of applied semiotics. For sure, and of this I am one hundred percent sure, speaking better than us does not mean that chatbots will stop lying. On the contrary they may start lying in new ways. They may make the concept of truth itself *evolve*.

Let’s try to glimpse at what the evolution of language may turn out to be. Imagine that you ask chatbots to write you an ad that will persuade the audience to do something, for example to buy a bottle. Imagine that the ad works: that people start buying bottles after reading the ad. Now keep imagining that you do not understand the text in the ad. Literally: you do not understand what is the meaning of what you are reading, you just go and buy a bottle. This example may help us to see what it is to ‘solve’ the “speaking problem”.
Please pay attention to this: I am not saying that to speak is simply to persuade, otherwise a whip could also be considered an evolution of language, and the NRA would be the greatest eloquence association in the world. What I am saying is that a machine may learn to do what we do with words, but in different ways that may prove more effective. We humans do not exactly know why we speak: whether it is for collaborating on serious tasks or instead just for social grooming. The fact that we do not know why we talk does not mean, does not mean at all, that we do not talk for a reason. Therefore, a machine can become able to understand why we speak, and then to do it better.
Let us think of the language games invented by Ludwig Wittgenstein. Language works not *only* to make you buy bottles but *also* to make you buy bottles. Language works not *only* to make others fall in love but *also* to make others fall in love. Now, it is more than possible that chatbots will pretty soon be able to make us fall in love like apples falling down from a tree. People already say that chatbots are great psychoanalysts, and everyone knows how easy it is to fall in love with your shrink.
Language is a natural phenomenon and like all others it may evolve, there is no reason why it could not. Language that was born within our species and shaped human society may very well migrate onto another species and keep progressing beyond our cognitive limitations. The game Go has evolved to a point in which we do not understand it anymore, but it still remains the same game. Something similar may happen to language.

If this seems already too crazy for you, you may want to consider that AlphaGo has already been surpassed by another program: AlphaGo Zero. Instead of learning how to play Go by studying human games, with all their defaults and biases, AlphaGo Zero just plays against itself. It started as the worst player in the world, not even knowing the rules of the game, and ended up being the best player in the world in only a few weeks of just playing by itself. With a dirty metaphor, it's like if someone became the best lover ever, just through a lot of masturbation.
Can language also be learnt by masturbation? Can a "GPT Zero" be able to learn a language just by talking to itself? I'm not referring to English or Chinese, with all their historical sedimentations of quirks and inefficiencies, but about a language comparable to English or Chinese in terms of its reasons and uses. Of course, a hypothetical "GPT Zero" would need to access the world naively, without any human mediation: it would need eyes and hands to experience reality by itself. So most probably some perception and psychology will have to be added somehow. The day when this happens will be splendid, but I am afraid that we humans will be incapable of fully appreciating that splendor.

If you are interested in the arguments I raised today, I suggest the incredibly good short text by Umberto Eco called "CSP: Charles Sanders Personal", in which Eco argues that according to Peirce's theory of signs there is absolutely no reason why a computer can't think like a human. I also immodestly suggest my own paper called "The Cost of Truth", in which I try to refine the pragmatist concept of truth.