filosofia unisinos | Unisinos Journal of Philosophy

ISSN 1984-8234

Filosofia Unisinos
Unisinos Journal of Philosophy
25(1): 1-13, 2024 | e25115

Unisinos – doi: 10.4013/fsu.2024.251.15

Article

Reflections on the future of artificial intelligence: an interview with Luciano Floridi[1]

Reflexões sobre o futuro da inteligência artificial: uma entrevista com Luciano Floridi

Murilo Mariano Vilaça[2]

https://orcid.org/0000-0001-9720-5552

Fundação Oswaldo Cruz, Escola Nacional de Saúde Pública, Programa de Pós-Graduação em Bioética, Ética Aplicada e Saúde Coletiva (PPGBios), Programa de Pós-Graduação em Saúde Pública (PPG-SP), Rio de Janeiro, RJ, Brasil. Email: murilo.vilaca@fiocruz.br

Murilo Karasinski

https://orcid.org/0000-0002-6099-6968

Pontifícia Universidade Católica do Paraná - PUCPR, Educação Continuada da Escola de Educação e Humanidades, Curitiba, PR, Brasil. Email: k.murilo@pucpr.br

Kleber Bez Birolo Candiotto

https://orcid.org/0000-0002-2000-4776

Pontifícia Universidade Católica do Paraná - PUCPR, Programa de Pós-Graduação em Filosofia, Curitiba, PR, Brasil. Email: kleber.candiotto@pucpr.br

[1] Agradecemos às agências FAPERJ (JCNE) e CNPq (PRÓ-HUMANIDADES).

[2] Bolsista da Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ - JCNE). GN: E-26/201.377/2021. Bolsista APQ do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq): PROGRAMA PRÓ-HUMANIDADES, GN: 421523/2022-0. Bolsista de Produtividade em Pesquisa (PQ) do CNPq, GN: 315804/2023-8. Agradeço à FAPERJ e ao CNPq pelo apoio.

ABSTRACT

In this interview, we delve into the reflections of Luciano Floridi, a prominent figure in contemporary philosophy and a pioneer in the philosophy of information. Recognized globally, Floridi plays a crucial role in addressing the ethical implications of digital technologies. The dialog covers crucial themes in artificial intelligence, exploring "Singularitarianism" and questioning whether perspectives have evolved with the rise of generative AI. The discussion extends to large language models, highlighting the lack of real reasoning and the disconnect between agency and intelligence. The interview highlights the importance of developing intelligence and critical thinking in the face of the challenges presented by ChatGPT, addressing concerns about possible cognitive decline. As for the future of artificial intelligence, Floridi outlines key issues in GELSI, mapping implications for governance, ethics, legality and society.

Key-Words: Luciano Floridi, artificial antelligence, future of humanity.

RESUMO

Nesta entrevista, mergulhamos nas reflexões de Luciano Floridi, uma figura proeminente na filosofia contemporânea e pioneiro na filosofia da informação. Reconhecido globalmente, Floridi desempenha um papel crucial na abordagem das implicações éticas das tecnologias digitais. O diálogo abrange temas cruciais da inteligência artificial, explorando o "Singularitarismo" e questionando se as perspectivas evoluíram com a ascensão da IA generativa. A discussão se estende aos modelos de linguagem de grande porte, destacando a falta de raciocínio real e a desconexão entre agência e inteligência. A entrevista destaca a importância do desenvolvimento da inteligência e do pensamento crítico diante dos desafios apresentados pelo ChatGPT, abordando preocupações sobre possíveis declínios cognitivos. Quanto ao futuro da inteligência artificial, Floridi delineia questões fundamentais no GELSI, mapeando implicações para governança, ética, legalidade e sociedade.

Palavras-Chave: Luciano Floridi, inteligência artificial, futuro da humanidade.

1 Introduction

Recognized worldwide as one of the most authoritative voices in contemporary philosophy, Luciano Floridi is considered the founding father of the philosophy of information and one of the leading interpreters of the digital revolution. He is deeply committed to policy initiatives on the value and socio-ethical implications of digital technologies and their applications and collaborates closely on these topics with many governments and companies around the world.

Floridi is Professor of Philosophy and Information Ethics at the University of Oxford, and Director of the Digital Ethics Lab at the Oxford Internet Institute. His academic career includes stints at several prestigious institutions, including the University of Cambridge and the University of Hertfordshire. He is currently the founding Director of the Center for Digital Ethics and Professor in the Cognitive Science Practice Program at Yale in the United States.

Among his most influential works is *The Philosophy of Information* (2011), a seminal work that lays the foundations for a philosophy of information. In this book, Floridi explores the nature of information, its relationship with reality and the ethical implications of living in an information society. Another relevant work is *The Fourth Revolution: How the Infosphere is Reshaping Human Reality* (2014), in which he examines the fundamental transformations in society resulting from the information revolution. His most

recent books are *The Ethics of Artificial Intelligence - Principles, Challenges, and Opportunities* (2023) and *The Green and The Blue - Naive Ideas to Improve Politics in the Digital Age* (2023).

In this interview, we explore crucial issues in the artificial intelligence landscape, investigating "Singularitarianism" and its dogmas and asking whether perspectives have been adjusted with the rise of generative AI. We also address large language models, highlighting the absence of real reasoning and the dissociation between agency and intelligence in LLMs. We question the relevance of the term "intelligence" to describe these models, while investigating the proposal to "shepherd AI systems". We highlight the importance of developing intelligence and critical thinking in the face of the challenges of ChatGPT, pondering fears of cognitive decline. Regarding the future of artificial intelligence, we explore how Luciano Floridi is mapping fundamental issues in GELSI, considering implications for governance, ethics, legality and society, from Yale University's Center for Digital Ethics.

We hope that the dialog that has begun can soon be broadened and deepened with the visit of Luciano Floridi to Brazil. This is the fifth in a series of interviews with renowned international researchers conducted by Grupo de Pesquisa Filosófica sobre Transumanismo e Bioaperfeiçoamento Humano - GIFTH+ (Fiocruz/CNPq), which has already featured Anders Sandberg, Allen Buchanan, John Danaher, James Hughes and will soon also include Nicholas Agar.

2 Interview

Murilo Vilaça (MV): Professor Luciano Floridi, before we begin our interview, we would like to thank you for accepting our invitation. It is certainly not our intention to make you uncomfortable with an unnecessarily laudatory speech, but we must acknowledge your invaluable importance for the constitution and development of some fields of research that are so crucial to our time. You are a pioneer in the field of the Philosophy and Ethics of Information and Computer, also called Digital Ethics, and your theoretical contributions and practice are priceless. Your intense academic work as a professor in several internationally renowned institutions, as the author of several articles and books, as the editor-in-chief of an extremely important journal like *Philosophy and Technology*, as well as your outstanding work in the field of the development of policies related to your research themes alongside organizations, associations, and internationally relevant boards are extremely significant for people in general and especially for us, scholars, so thank you so much for kindly and quickly making room in your very busy schedule for us.

In this interview, we are going to focus only on artificial intelligence, which many times seems to be a new hype, given the superficial and sensationalist way in which it is portrayed. Precision and ponderation are in short supply in the amount of newspaper reports, op-ed articles, and even articles published in journals. In a discussion that is filled with opinions that do not always have a firm ground, but have strong advertising, a good understanding of artificial intelligence and its issues and implications is essential.

Murilo Karasinski (MK): Professor Luciano, it's truly a pleasure to be here. My name is Murilo as well. We have a few questions for you so that this debate can grow and flourish in Brazil, and I am sure that it's going to be a very fruitful evening for us here in Brazil, so thank you for your kindness and generosity to talk to us. Thank you so much. I will pass the floor to professor Kleber.

Kleber Candiotto (KC): Me too. It's a great honor, professor, to be able to talk directly with you. I have read so much of your work since I was a Doctoral student in Brazil, and so have my students and many researchers who are now graduating after reading your research and publications, so it's a great honor and satisfaction. I'm sure it's going to be a very productive moment for us. Thank you very much.

So, professor, in your 2015 article about Singularitarianism, you criticize the approach of Singularity advocates, which you call "Singularitarianism", for being based on three dogmas. The first is that the creation of some form of artificial superintelligence – a so-called technological singularity – will happen in the foreseeable future. The second dogma is that humanity is in great danger of being dominated

by this superintelligence. The third tenet is that the primary responsibility of the current generation is to ensure that the Singularity does not happen or, in the case that it does, that it is benign and benefits humanity. Since the advent of generative AI, which is after your article, has anything changed in your view? Could there be any updates regarding your criticism?

Luciano Floridi (LF): Well, first of all, *muito obrigado a você* [thank you very much]. Unfortunately, my Portuguese doesn't go very far, despite the fact that my wife is Brazilian, so I should speak much better Portuguese. She is from Rio. *Ela é carioca* [she is a carioca]. But I think that there is… and not just for this interview. I've used this phrase before, in the past. I can start replying to your question by saying that – and it cannot be translated in any other language – *o buraco é mais embaixo* [the hole is much deeper]. And that is the problem – that the approach that we have these days to AI is very superficial, and people don't look deep enough, and don't understand that the hole is much deeper, in a bad English translation. And the Singularitarian position, for example, which is as old as Alan Turing… It was actually one of the students, colleagues of Alan Turing who started this superintelligence story in the middle of last century. It has gone, if anything, worse in the past few years. It has gone worse because we have seen machine learning. ChatGPT is so famous that it doesn't need to be introduced to anyone… achieving things that we thought were really almost impossible, or at least some people thought it was impossible, and therefore people started thinking "today this, tomorrow something else, the singularity is coming", this point of no return, when AI would take over, would dominate, would make us slaves. And we have had plenty of businesspeople and some AI experts claiming that this is a real threat, that it is a real problem. In fact, some people, like Elon Musk, have actually said that this is the biggest problem humanity is facing, that it is an existential risk. It is science fiction. If you have any exposure to real AI in a real basis and you're not invested in producing, selling, researching, developing AI, but you are outside the AI community, you know perfectly well that it is not really very good at all. Let me give you an example for anyone who is listening. You play chess with an AI system. It will win against anybody. Oh, so that is superintelligence! Not really, because if it starts being a risky environment… imagine the fire alarm goes off, you stop playing chess and get out of the building. The AI system keeps playing chess until it is burnt to the ground! Because it plays chess better than everybody else but it has not any other skill or understanding or context. There is nothing more than an extraordinary, fantastic, amazing technology that that thing which is even more amazing, more fantastic, more extraordinary, which is called a human mind, human intelligence, has developed. So bottom-line: the updated article is: I'm afraid the problem has not gone away, it has become worse, and the hype today hides two points, and I will keep it short here. One is that it has become completely detached from real science as we know it. If we look at the claims, they come with no control, no experiments, no margin of error, for example, and no timeline. It's always about "it could happen one day". Well, of course, the problem would be enormous if one day we had zombies going around. Oh, that would be a big problem! The solution is that luckily, there are no zombies, so you don't have to be worried about that. Likewise with AI, if that kind of AI were to be developed, it would be a big problem. But luckily, if you like, there is no such thing coming at any time in the future. The second point is that, contrary to the past Singularitarians who were, if you like, believers, the current people who are pushing for this narrative, they have a vested interest. They make money out of it, or they run big research labs. Again, power, visibility, fame, money, networking… So should I listen when someone who makes money out of it is going to say "look, it's going to be so powerful you wouldn't believe it"? Of course not. They are, in other words – and that is my point – hypocritical. Their hypocrisy is obvious when, for example, Elon Musk signed that letter saying "the singularity might be coming, we are all at risk, block everything, regulate it" etc., and then in a number of days, he started buying all the technology required to launch his own AI company! Now. Anyone in his right mind, wouldn't you stop sort of something that is so risky if you really believed that it is risky? Or is it just hype, advertisement? And, above all, a certain kind of culture, which is, I'm afraid, very Californian-oriented, and a certain kind of world where the focus is not on people who everyday, and I'm talking about 800 million people in the world who don't

have clean water. Everyday. Eight hundred million people. Now that is an existential risk. AI taking over the world? That is science fiction, a distraction, and is not funny anymore.

KC: Very good. In the wake of this question, I have a second question, professor, which is a continuation of it, but with a philosophical provocation of sorts. In your most recent article, in which you even approach these large language models as we call them, like ChatGPT, you discuss these large language models (LLMs) that are causing enormous optimism about the powers of the so-called generative AI, especially on account of the most famous ones like ChatGPT from OpenAI-Microsoft and Bard1 from Google. According to your analysis, these, like any other AI, do not reason or understand anything without any relation with the cognitive processes present in the animal world and, above all, in the human brain and mind, to successfully manage semantic contents. Even in view of this significant advance in generative AI, do you think that the well-known "Chinese room argument" by John Searle is still valid to demonstrate the unfeasibility of a Strong AI?

LF: I think so. In fact, actually, there is a video on YouTube where John and I discuss his argument. It's not very interesting because we agree, so there's no real debate. I think John Searle got it right when he said: "these are syntactic engines", meaning that they apply rules. So if, for example, an odd number comes in, you provide an even number, or for every number that comes in, add plus 1. But it doesn't mean that it understands anything, or it's talking about truth or false, meaning, context. Now. What we normally would treat as intelligence is not there at all. Now. What the world normally calls "intelligence", of course "intelligence" means many things for many people. My normal joke is that there are as many forms of intelligence as there are of stupidity. So it is not intelligent to close the door with your fingers in the door. That is stupid. You would have been intelligent to remove the hand. It is not intelligent to go to the airport and forget the passport at home. It is not intelligent to tell someone something when it is not the right moment, maybe not, maybe tomorrow, maybe the person is not feeling well. So there is a good time for everything. There is also mathematical intelligence, musical intelligence, dancing is a form of intelligence, I mean, there are a gazillion forms of intelligence. It means being human. And of course, in that context, we don't find any AI. So John was and is right. John is still holding on to that argument. Of course, he has been criticized in various ways, but I think the criticism missed the essential point, which is there and stands: syntactic engines don't process semantics. Semantics means understanding what is the content of the message. Let me give you a very simple example, so for people who might not be acquainted with syntax and semantics, it might be a little bit confusing. So imagine you have a gigantic puzzle, say, 5,000 pieces. So it's huge. Say, since we are here, it is, I don't know, a puzzle that describes Copabacana. So it's ~~a~~ fantastic and beautiful... How do you start the jigsaw, the puzzle, putting together the pieces? As a human, I start looking at the colors, the features, the pictures… so this is a piece of blue so it must be either the sea or the sky… there is something there, it could be a face so it goes with anything... Imagine now AI doing the same. AI is looking at the other side of the puzzle – white. It puts all the pieces because of the form, the pattern, the structure of each piece. Now if you look at AI putting together the puzzle, you think "oh my goodness, it's understanding"! Why? Because that's the only way we would be able to do it! By not picking up every single piece and looking at the shape and saying: "Ok, this piece goes with that", but rather by mounting all the right pieces in the right place, etc. So the bottom-line is: you can do syntactically what we would be doing semantically if you are a machine. It doesn't mean that, therefore, there is an understanding behind it. One more analogy before I stop here. Imagine you come to my house, and there are clean dishes on the table. I'll ask you to please guess who or what cleaned the dishes – the dishwasher or me, by hand? You can't tell. What is the implication? None. It doesn't mean that, therefore, the dishwasher and I are the same. It doesn't mean that we did it in the same way. It just means that the outcome cannot tell us who did what and how. That is crucial! It means that, for example, you read a little poem written by Chat GPT or me, and I ask you: who wrote that poem? And you look and say: "oh, I can't tell". Therefore, the machine is intelligent. No, of course not! Because it would be like saying: "therefore, the dishwasher is like Luciano". It did the dishes exactly like Luciano. It didn't. By hand in one way, the other one is a completely different mechanism. So

what we are understanding, also in line with John's argument, is that syntax and semantics can generate the same content – maybe a movie, maybe a picture, maybe a text, and lead to the same sort of successes, playing chess, parking a car, finding the right ticket for a concert online. It doesn't mean: 1) that the process is the same; and 2) that the source is the same. The only thing it means is that the outcome is either equal or even better! But then it would be like saying that a motorbike runs faster than me. Of course. I think we are anticipating some of the questions that we want to discuss later because the real point is: who drives the motorbike? Who controls the dishwasher? But that's maybe for later.

KC: No, that's perfect. And just to conclude, like the washing machine, the dishwasher also needs human beings to remove the dishes and to put them back, so our interaction with AI also needs a human command and a human filter for something significant to be produced. The metaphor was very good, so thank you for that.

LF: Absolutely. Let me comment just briefly on this DALL-E that produces all those images. Well, it needs scripts! And the scripts are very complex, if you want to get something decent. I mean, I've tried several times and my outcomes are ridiculous, the video is very poor, because … a robot with some flowers, boom, that is rubbish. But I've seen experts, real experts doing that, and the outcome is amazing. So the real point becomes: how are we going to use this technology? Who is in control of what, according to which rules etc., but more on this during our conversation.

MV: Professor Luciano, before I ask my first question, I just have to make a brief comment. My wife is a linguist, and she is a linguist of the Chomskyan theory. Recently, in an interview, professor Noam Chomsky challenged the possibility that systems like these large language models systems could reproduce human language, because in his theory, that is a capability that is unique to human beings. Your comments on syntax and semantics led me to discussions that I have with her because we are starting to write a text together and we have been testing… I don't know if I can call google translator a linguistic artificial intelligence, but testing google translator, it cannot recognize idiomatic expressions like the one that you mentioned in the beginning – *o buraco é mais embaixo*. I am not a linguist, so if I'm saying nonsense here, I'll ask my wife to correct me later. But human creativity and the resource of language would not be emulatable, so they could not be copied by any artificial intelligence system. If I'm not mistaken, in the beginning of this text that Kleber mentioned, you make a lot of linguistic claims, saying that this is nothing new, that this is something old, and then you go on to make an analysis of ChatGPT. This is just a comment, I don't know if you'd like to comment on the comment before I ask my first question….

LF: Well, first of all, I think that Chomsky was absolutely right. The human mind is, to a large extent, and I mean a really large extent, mostly a mystery. And I'm not just a philosopher saying so, because speaking of wives, my wife, she is the chair, or was until last week the chair of Neuroscience at Oxford. And she just moved to Yale, and she's now directing the… She's professor of Neuroscience here, chair of Neuroscience here at Yale, and she directs the Yale Center for Human Neuroscience. And speaking to her, as you do with your wife about linguistics and Chomsky, you learn that we don't know almost anything about how the brain works, or even what the relationship really is between mind and brain. Now, there is not necessarily anything magic, mysterious, but we need to acknowledge that we don't know. And it's not like: "oh no, we don't know"! We just don't know. It's an immense continent that we have just started exploring. My usual analogy here is that we landed on the beach and it is Australia in front of us. Now, what do we know… or maybe, going back to our cause – it's Brazil! I mean, what do you know about what… you are going to find or what is it like? And then, by landing on the beach, we already claim that we are creating technology that is like a brain? You don't know what you are talking about! So how do you say that you are generating something that you have no idea how it really works? So back to us, of course, you know, Neuroscience has made an enormous amount of gigantic steps forward, but, as I said, it's really at the beach level of this immense continent. So back to us, the point becomes more like: Chomsky is right, it doesn't have to be something necessarily mysterious as in "oh, there is a ghost in the

machine", or something that is necessarily religious. It may or may not be. I'm happy to leave that door open to believers and non-believers. But there are many mysteries around us. They have nothing to do with religion. All they have to do is with our lack of understanding. There is a second point here, which I'd like to comment on your comment. Sometimes you hear people say: "oh, but nature has developed brains". Forget about humans, say, even the brain of a dog. So, "if nature has done it, we must be able to do it!" Well, this is simply ridiculous! Imagine you look at the solar system and say: "oh, nature has created this solar system. If nature has created, then we must be able to do it!" No. We might be able to understand it, but building it, impossible! Like, beyond our wildest dreams! Unless, of course, you're watching Star Wars, in which case now we can create planets… but then you call Elon Musk and… or running an AI company. So the fallacy here is: "nature has done it and, therefore, we must be able to do it." Not really. Even think about cold fusion. I mean, maybe one day, but even then, we are not quite there. But creating a whole solar system, that is beyond our means and capacity and power. And the brain is a bit like the solar system. Just because nature did it, it doesn't mean that we will be able to do it again. Not necessarily. We might be able, one day, to understand it, yes, but it might be just that in the same way in which we understand the solar system, but our understanding cannot be leading to an engineering of the solar system. But back to you, I know you had another question so I don't want to take too much of your time.

MV: No. Please. Thank you for your comment, which is at the same time enlightening and comforting in view of so many adamant and sensationalist allegations about artificial intelligence. But I'd like to go back to a point that was mentioned by my colleague before. In the same article that he quoted, published in 2022, you say that the LLMs represent a dissociation between agency and intelligence. Given these statements, would there be a better term than "intelligence" to conceptualize the large language models? What kind of strange and never seen before agency is this? Considering that there is a difference between agency and intelligence. So, what kind of agent are we dealing with? What is exactly new about this "agent without intelligence", if it can be defined that way? How will we "shepherd" it, considering that the "algorithmic language" is "the home" of this new "being" that *agere sine intelligere*? In other words, what do you have in mind when you say that we will be "shepherds of AI systems"?

LF: Thank you. That's another very good point. These are all wonderful questions, thank you. Let's start from the initial point: is there a better term? Well, "artificial intelligence", as everybody knows, is Wikipedia level. It was John McCarthy and a bunch of kids at Dartmouth… He was the one who came up with the expression. I had the pleasure to interact… and the honor to interact with John when we were part of one project of technology of information etc. And so there was time during those meetings to ask him a couple of questions, so "John, how did you come up with this?" "Was that a good idea?" And he was astonished by the success, of course, and he said: "we had to find an expression and we found this one". It has been essentially a keyword ever since. It is an unfortunate expression, because it generates a lot of confusion. Now people understand artificial intelligence, they think human intelligence, biological intelligence, engineered intelligence, but here is where I'm very happy, so to speak, to be told that I was completely wrong, now fast forward 10 years from now, and someone looks back at our conversation and says: "look how silly that professor was"! But here is the point I want to make and I hope to be right. I'm the only one as far as I know that will say this but no, someone has to say it: we're not creating a new form of intelligence. This is not a marriage between biology and engineering. So we're not engineering some kind of intelligence that would be even remotely resembling the intelligence of a mouse. What we are doing, which is equally extraordinary, as extraordinary as creating a new form of intelligence, is detaching intelligence from the ability to do something – perform a task, solve a problem… with success, in view of a goal, learning from the output without exercising any intelligence whatsoever. Now this is extraordinary, and leads to your second question, but let me add just one comment on the expression before we move on. So "artificial intelligence" is an unfortunate oxymoron, meaning a contradiction in terms: if someone is intelligent, it's not artificial; if something is artificial, it is

not intelligent. So it's like talking about a happily married bachelor: it doesn't exist. Anyway, it's too late. So the first answer to your question is: we won't have another expression. By now we have some couple of 50 years or so of "artificial intelligence". It's part of our vocabulary. What I think will happen and again, I'm saying this with you today but let's see what happens in 10 years, is that in the same way as we don't look for horses when we talk about horsepower – hp – you look at the engine and say: "oh, the 'hp' of the engine is 'x'". No one in his right mind will ever look for horses inside the engine! Why did it happen? Because when we developed cars and engines for cars, all we had were carriages and horses, and so the comparison was: "look, we don't have the vocabulary, we need to speak about this new technology, here is a horse, this is like one horse, two, four eight sixteen thou… gazillions of horses". Now with "hp", no one looks for horse power anywhere… one will not look for horse nor power or artificial intelligence inside a computer, for example. One day, I hope so. So we will get used to the expression. The second example that I usually have in mind is, I think in any language but you will correct me if it's not true in Brazilian Portuguese, certainly in English and Italian, we say that "the sun rises". If we think about it, the sun is not going anywhere. It's the Earth that is moving. But we still say "oh, the sun rises" and etc. because there's a millennia of culture, but no one who actually tomorrow is going to tell you "oh, the sun rises at 6 o'clock", and you say "oh, you need to study more your astronomy, the sun doesn't go anywhere". It's a manner of speaking which we have inherited from our culture. So there are no horses in horsepower, the sun doesn't go anywhere, artificial intelligence is not artificial. Sorry, if it's artificial, it's not intelligent. We will get used to it. So I'm afraid the first answer to your question is: no. We will keep the terminology, trust me, but hopefully we will get used to it and just say: "oh yeah, I know what you mean". It's just like a dishwasher. It's just a very smart dishwasher, if you like, or does things, you know, better etc. Now. More complicated is this hypothesis that I have developed for some time of this detachment between agency and intelligence. Because normally, if we do something without intelligence, it is a disaster. Try to do the dishes without intelligence and you break everything. Play chess without intelligence, you lose immediately. Park the car without intelligence, you will make a mess, it's a crash. So how come that this engineered artifact is successful without any intelligence? Well, because what happens is that, first of all, we have amazing algorithm, statistics, amount of data, sensors, but above all, we have organized a world around the limits of the machine. So when you have, say, a driverless car… I just bought one and it's impressive. I mean, actually the wheel moves under your hands. I'm scared, so I'm constantly paying attention, precisely because of this conversation, so I don't believe there is any intelligence, so this thing is dangerous. But what happens is that there are satellites, there are… roads are made so that cars can actually go on those roads. If you take that car in the middle of nowhere, say, a simple open path in some corner of Brazil, that car will not go anywhere, because it needs to check the lines, the white lines, and so on of the road. So we are building the environment within which that form of agency is successful as zero intelligence. There is a big risk here, which is, by building the environment around the technology, we might be at risk of forgetting that we live in that environment. And there is, you know, the kind of language that they speak. If we make sure that the whole world speaks "algorithm", we who are at this point "foreigners", who do not speak their language, do not act like machines, we will be… the entities left out. Now if this sounds a little bit too strange, imagine once again, say… Rio de Janeiro, by the way, is one of the most advanced cities in the world in terms of digital technology to manage the city. People may or may not know, but it is Chicago, Barcelona, etc., it's over there. Imagine we are in Rio de Janeiro, and someone in Rio decides that they need to build traffic controls to make sure that driverless cars can be successful. You rebuild all Rio de Janeiro around driverless cars. But then we live there. And so you with your car, me with my dog, the kid playing etc… we will be sort of at risk of being used or neglected or marginalized because the real problem needs to be solved so that driverless cars can go, say, driverless buses can go through Rio successfully. Obviously, this is not the scenario we want to see, and this is the risk we are running, among many others. So the points are very simple: either you believe that we are generating a new form of intelligence – I think that's science

fiction – or you believe that we are creating a new form of agency as zero intelligence. How can it be successful? Building the environment around it. What's the risk? That we forget the humans, who are the receivers of that policy. We need to make sure that humans are always the end, not the means for the development of a technology. This is very Kantian by the way, as you all know. So the final point in your very interesting and complex question is: so what kind of agency are we talking about? Well, the kind of agency that we are talking about does the following thing: it's anything that 1) changes the world; 2) so it interacts, because if it doesn't interact, it's not agency, or he's not an agent, so if he's an agent he interacts with the world, he interacts with the world in a way that he learns from his interaction about the world; and 3) can change its own rules. So by learning from the interaction with the world, the data, feedback, change the rules, improve the performance. There are 3 elements: interaction, successful with the world; feedback, so that learning procedure; and autonomy meaning it can regulate itself. So interaction, regulate itself, and learning. These three things make this agency successful given the environment built around it. What is lacking is any mental… any intelligence, any understanding, any semantics, any meaning, any relevance, any context – what we would normally call "human intelligence". That's fine. Now, really, two final comments. One, there is a difference between a river and artificial intelligence. A river modifies the valley by going to the sea but it cannot learn from its own experience and cannot change its own rules. A volcano, an earthquake, they are agents in the very uninteresting sense of making a difference in the world, but they don't learn from the difference that they make. The waves of the sea, of Copacabana, they don't learn from what they are doing. They never go back to their own feedback. And, therefore, they cannot change their behavior. So between, say, the sea, the waves, an earthquake, a volcano, on one hand, and human agency on the other, you have something in the middle, something that is a bit more than just an earthquake, or a river, or a wave, or a volcano, and much less than human agency. It's that middle ground that we have never seen - and that's the last comment – in our history. Until yesterday, if you wanted to be successful, you had to exercise some biological intelligence. Maybe a dog, like a Shepherd, or maybe, say, a mouse catching… sorry, a cat catching a mouse. But you had to have some level of intelligence to be successful. We now have this middle ground, which is not as brutal as a simple cause and effect of the river, but it's not as intelligent, not at all, or mindful, as a human. The middle ground is the agency which we need to study. So the final, final comment to your question is: we need to do much more work in understanding what we are developing, as opposed to science fiction and thinking that we are somehow creating a sort of human intelligence or even superintelligence.

MV: Thank you very much, professor. Murilo, you have the floor now.

MK: Perfect. Well, professor Luciano, I have two questions but I think that because of the time, we don't want to keep you from anything, I believe you have many commitments, so I'm going to summarize my two questions into one, and with that I think we're able to have an interesting discussion because both questions are similar in a way, in the sense that my questions have to do more with a perspective on human beings. So the debate that I'd like to foster here begins with an analysis of an article of yours that is entitled "GPT 3: Its Nature, Scope, Limits, and Consequences", in which you say that in view of the challenges of ChatGPT, it is crucial that mankind develops greater intelligence and critical thinking. You say that we should seek this complementarity between human skills and artificial capacity, and promote successful interactions between people and computers. To do this, it would be fundamental to create a more robust digital culture, making current and future citizens, users and consumers aware of the new reality of the infosphere in which they live and work. Here in Brazil, we have a researcher who is a very famous neuroscientist here but I believe he is also famous abroad, he's called Miguel Nicolelis, and his understanding, which I believe is similar, is that algorithms could contribute to a cognitive decline of human beings, as artificial intelligence systems in a certain way, they could prevent humans from learning how to reason, how to think, and also how to do basic tasks that humans from older generations could do. And making a parenthesis here, we are university professors, and we notice nowadays how many students today have difficulty to write texts without using AI systems to make the texts better, or even

have difficulty to write texts that are not texts… or rather, texts that are handwritten. If you remove the computer from them, they are not even able to write an essay, so we wonder if the world overnight… if the laptops were gone and these algorithms vanished, how would people produce texts? Because those are things that for the older generations, they were the easiest and most obvious things in the world. So the first question is: how do you assess this fear that artificial intelligence can lead to a world that doesn't think, with this diminished intelligence and maybe impaired critical thinking for human beings? And how could we make a more robust digital culture? So that would be the first question.

And the second is similar to the discussion, and it is based on an article of yours, also from 2019, called **"**What the Near Future of Artificial Intelligence Could Be", and if you could, I'd just like you to comment a bit on matters that were known as ELSI (Ethical, Legal and Social Issues), which are crucial issues for us to deal with in a legal, ethical and social manner about artificial intelligence. A few years after this article of yours, I would like to know if you see that there was an advance or a regression in terms of these ethical, legal and social impact issues of artificial intelligence. So basically that would be a discussion that has to do with human beings.

LF: Thank you, that is a very important question. As it happens, because of Kia, my wife, I met Nicolelis a couple times in Brazil, in fact in Natal, if I remember correctly, in the north. So… very very smart, very famous scientist. I think he's wrong on this particular point, but then, philosophers… quite arrogant! Whenever we have deep transformations like the one we're talking about, transformations that are not only just technological, but they touch the nerves that… the spinal cord of our culture, of how we even consider ourselves to be human, there is often that sort of reaction, the negative reaction, to say "oh, people are losing their fundamental skills, young people no longer know how to do x, y and z, when the past generation is gone, who will be able to have critical thinking, or writing skills, reading habits, etc". Well, considering that Plato, as we all know, was already complaining when writing was invented: "oh, that's the end of the world, no one will ever remember anything, the memory and the oral culture, that's the end of…" He was right, in one sense! It was the end of a… Writing was the end of an oral culture as they knew it. In fact, by the time Plato was writing… Yeah, so commenting on writing, writing had been around for a long long time, for centuries! You know the Egyptians had been writing for centuries and centuries. It will be like someone complaining about computers in 2000 years, ok? So, writing was invented roughly six, six and a half thousand years ago in that area of Europe in the middle east. It was invented four times, as far as we know, around the world, and Plato was writing a few centuries before Christ, so he was late, but he was still complaining. Now let me compare, to be kind… Nicolelis to Plato, I know… The complaint "oh no, kids these days"… Well, honestly? It will be like saying: "oh, kids these days they don't know how to record a song on a tape"! Exactly. Who needs to be able to record a song on a tape? Which I did when I was 18 years old! Now you had the tape, and you had to listen to the music, and as soon as the guy stopped talking, click! So you get the song, and you were hoping that the guy would not talk while you were playing the music on the radio! That skill is gone. Do I feel that, unfortunately…? No! It was a recording system, it was different... Today you don't even dream about it. Of course, there are things that we lose and things that we gain. The important thing is to understand the balance here. No one should complain about the fact that we do not… I'm sure that no one in this conversation knows how to shoot a horse. I have no idea. Do I miss it? No. This was a great skill centuries ago, of course, I mean, there were only horses. Imagine if someone were to say: "look, one day no one will be able to type anymore". Ok, why? Well because the vocal recognition or even some implant… a little chip here will simply transform your thoughts into, no, your brain waves into text. Am I going to miss the ability to type on a keyboard? I mean, who cares? I'm sure that year someone will say: "oh, good old days when we were able to use the keyboard", etc. So there is something like that, which we need to avoid. There is something more, which is critical thinking, the ability to understand… or everything will be delegated to AI and we'll be just passive... I hope I haven't lost you. Can you still hear me? Hello?

MV: Yes, we can hear you.

LF: OK. Let me try to go back to that point. These technologies, like all the other technologies in the past, they polarize. They don't put humanity on one side. They just polarize humanity between those who will be subject and victim of these technologies, and those who will be empowered by these technologies. You can do so much more, for example, using ChatGPT if you are smart. Like, for example, I've already seen and I know and even myself, you can use it to do a lot of bureaucracy that is a waste of time. You need to fill forms? Well, ChatGPT can do that for you. You need to say: "compile". And I did that only a few weeks ago: "please compile the last five publications with a summary in a 100 words of each publication". Who has time? But the bureaucracy wants that. So, "ChatGPT, list the five latest publications by professor Floridi with a summary of 100 words", and it was almost perfect – 90%. It took me 10 minutes to just touch and go, boom, send, everybody happy. So imagine if someone were to say: "look, the invention of the car made us fat, obese, because you don't walk anymore, you drive everywhere etc." Well, it's also the car that takes you to the gym! So if you really want to complain about something, it's what you do with it. So that is… I find that really unconvincing, and is kind of old-fashioned. It's the talk of some people who are old, and therefore, they don't adapt, they don't want to see the novelty, and they think that the past is better than the future. The moment under the shower, one day, you start thinking "oh, good old days", mark my words: you have become old! And I hope that will never happen to me! I don't want to go down that road of… in my, you know, say, eighties thinking "oh, when I was young it was so much better"... Because it wasn't! It was different and sometimes it was worse. The second point that was in that question is: so what do we do? This technology is so immensely powerful, it is one of the most powerful technologies we have ever designed and engineered. We need to be in charge, in control, that's why the metaphor of the shepherds of AI, it was a bit of a joke on Heidegger. But that means educating especially the young people, the new generation, the people who will follow us, about how to be the people who control the technology, not the people being controlled by the technology. Now, this is doable and is perfectly reasonable and is what education does all the time. Unfortunately, we also have to be realistic here and maybe I'll share some thoughts with Nicolelis, which are: we need to offer to everybody the opportunity to use these technologies in a way that is empowering and enabling. However, not everybody will do that. Some people will just not make it to the other side. For those people, we need to have rules to protect them. Otherwise, not only they will not be empowered, enabled, but they will also be means to an end. And therefore, together with education, we need a lot of good legislation.

MV: Professor Floridi, I'd like to ask you if we have time for one more question. Would that be possible?

LF: Of course. With pleasure. Yes.

MV: My colleague referred to the acronym ELSI, which has gained a lot of attention in many debates on various topics related to science and technology in general. Recently, you and other colleagues are focusing on something new. You are mapping the fundamental problems in the field of GELSI, that is, the Governance, Ethical, Legal and Social Implications of digital technologies. You are revising the GELSI framework. From the research findings so far, what would you highlight as most interesting in the field of GELSI? If you want to emphasize more on governance proposals, feel free. In that sense, would the Digital Ethics Center at Yale University Center for Digital Ethics, of which you are the founding director, play a strategic role? Please tell us a bit more about it.

LF: With pleasure. Yes, GELSI is just using the famous expression ELSI by adding the "Governance" to it, which I find, these days, absolutely crucial. As we've said before, it's not so much about digital innovation the real problem, or the real challenge. The real challenge is not digital innovation, it is the governance of the digital. It's not what you are going to develop, but what you do with what you have developed that makes the difference. So imagine someone discovering a gold mine. Well, of course that's amazing, but it's what you do with it that makes a difference. To give you a more concrete example, Norway has a lot of petrol, and they did a fantastic job, a foundation, etc... everything is there for

the people and for the future. They manage their richness in a very intelligent way. So are we going to develop the governance and the policies that are able to cope with the challenges of this amazing technology? Or do we have the risk of just, say, follow the hype, the fashionable statements, letting business do everything, letting the market decide…? Well, I don't think so. Coming to the second half of your question, I hope that this center, this new center that we are literally establishing, that I'm directing here, the Digital Ethics Center, the DEC, will play, I hope, a significant role in developing the right governance, the right ethics, the right legal and social analysis of the impact and implications of this technology. We need this enormously, desperately, and so ultimately, if I were to ask for only one thing or one wishful sort of moment, I think… You have only one wishful thing for that center, what would that be? Find the right people to be members of that center, human minds that can actually help us solve all these challenges that we discussed together today and many more, and give us some help in shaping the technology and shaping the future, because ultimately, and I don't mean to extend more of your time, but ultimately this technology could be an enormous force for good. It could really help us to do two things that we need to do desperately today: save the planet, the environment, and save our society on this planet. If we could make sure that this enormous power is sent in the direction of supporting, say, climate change, the fight against climate change, the UN SDGs etc. and eliminate, or at least ameliorate, minimize the injustice, the level of inequality in our society, and many other forms that are linked to that, migration and so on, that would be… the plan for the 21st century… It is what I call the "green" and "blue", this particular marriage between… We spoke about the divorce, we can end with a marriage. The marriage between the "blue" of all these technologies, amazing, including AI, if we use them properly, otherwise they'll be a tool for evil in the wrong hands… wrong hands, not by themselves. Wrong people behind these technologies will make a mess. And the "green" of all our environments – social, family environment, working environment, political and, of course, biological. If we could put these two things together, then we have a recipe for the human project of the 21st century. That's amazing. That is our man on the moon kind of moment, the Normandy landing. It's a challenge the new generation should embrace. So I hope that the center with the help of I hope as many people as possible, will find the right people to develop the projects that will support a GELSI philosophy that will make a difference in the 21st century. If anything – and I'll close here – we want to be remembered in the future by future generations with a sense of gratefulness. Imagine two or three generations from now, looking back at us, and saying: "thank you, you did the right thing". That is the ultimate goal. If we are not there, if we miss this opportunity, if we make a mess of the environment and our society and this technology, which could help us to solve these problems, we will be remembered as the most wasteful and immoral generation for a long time. Now, the past is even worst sometimes, but surely we have a great opportunity. And as they say even in football: the game is ours to lose. We can win this game, but we need to put the "G" of "Governance", and therefore the "P" of "Politics" on the right track, and that's an effort for everybody, it's not just the center, so I hope there will be many more joining.

MV: Professor Luciano, it's a pity to say goodbye to kind, intelligent, and well humored people, but this is the moment to thank you so much for the generous and rich interview. You clarified a lot of things to us, and I believe we have a lot of material here that is so rich, and it's hard to thank you enough, but again thank you so much for the interview, and I will pass the floor to my colleagues so they can say goodbye to you as well.

LF: Thank you.

KC: Thank you, professor.

LF: Thank you! *Obrigado*!

MK: Yeah, I am very grateful and I fully agree with Murilo.

MV: Thank you very much, professor Luciano. I hope we keep in touch in any way, because since the first time I dared to contact you by sending an email, I was surprised by a very kind response, which was so fast, so I will keep in touch with you. You are very important in the field, and this contact can

enlighten this debate, which seems still very polluted, you know, with science fiction and speculations that are unreal, so your participation in this discussion is essential. Thank you so much! A great hug to you. Ciao, professor.

LF: Thank you! Muito obrigado. Ciao!

Ciao!

Grazie mile!

Ciao!

References

FLORIDI, L. 2011. *The Philosophy of Information*. Oxford, UK: Oxford University Press.

FLORIDI, L. 2014. *The fourth revolution: How the infosphere is reshaping human reality*. OUP Oxford.

FLORIDI, L. 2015. Singularitarians, aitheists, and why the problem with artificial intelligence is HAL (humanity at large), not HAL. *Philosophy and computers*. **14**(2): p. 8-11.

FLORIDI, L. 2019. What the Near Future of Artificial Intelligence Could Be. *Philos. Technol.* **32**: p. 1-15. https://doi.org/10.1007/s13347-019-00345-y

FLORIDI, L. 2023. AI as *Agency Without Intelligence*: On ChatGPT, Large Language Models, and Other Generative Models. *Philosophy & Technology*, **36**(15). https://doi.org/10.1007/s13347-023-00621-y

FLORIDI, L. 2023. *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*. Oxford, UK: Oxford University Press.

FLORIDI, L. 2024. *The green and the blue: a new political ontology for a mature information society*. Available at SSRN 3831094.

FLORIDI, L.; CHIRIATTI, M. 2020. GPT-3: Its Nature, Scope, Limits, and Consequences. *Minds & Machines* **30**: p. 681–694. https://doi.org/10.1007/s11023-020-09548-1

GHIONI, R.; TADDEO, M.; FLORIDI, F. 2022. Open Source Intelligence and AI: A Systematic Review of the GELSI Literature. Research Paper Series, Available at SSRN: https://ssrn.com/abstract=4272245 or http://dx.doi.org/10.2139/ssrn.4272245

SEARLE, J. R. 1980. Minds, brains, and programs. Behavioral and brain sciences. **3**(3): p. 417-424.

Submetido em 30 de outubro de 2023.

Aceito em 11 de janeiro de 2024.