© Lo Sguardo - rivista di filosofia
N. 34, 2022 (I) - Algoritmo

DOI: 10.5281/zenodo.7768566

Interviste/1

On Algorithms: Ethical and Epistemological Questions
Interview with Igor Pelgreffi

Luciano Floridi 0000-0002-5444-2280

Luciano Floridi tackles the topic of the algorithm and the ethical and theoretical challenges it entails, starting from a survey of some of his ideas, including the concepts of *Fourth revolution*, of *Infosphere* and of *On-life* (the latter indicating an extended condition of close interconnection between our life and the being online situation). Firstly, Floridi analyses the concept of algorithmic dependency on the complex data/software, with related responsibilities and risks. Then he moves on to the question of Machine Learning, recalling both the 'historical' theme of the so-called 'machine to machine interaction' and items from game theory frame. Interaction means not only sharing data but also modifying data: Floridi underlines therefore the role of human subject and of ethical responsibility in data managing, instead of the 'machines hyper-autonomy' perspective. Subsequently, Floridi interprets the meaning of the various problems raised here, regarding the algorithm and its forms, by retracing his main works: *The philosophy of information* (2011), *The ethics of information* (2013) and *The logic of information* (2019), thus also discussing the overall meaning of his research path. He also reflects on the problem of writing about the politics of information and, in perspective, of elaborate a reflection on 'meaning', that is on what it means to be human in a digital age. A philosophical thought around the issues of algorithmic era, if conceived in such a broad way, can therefore imply a profound and gradual change in our self-understanding, considering ourselves as 'informational organisms', like a slow evolution of the Aristotelian *zoon politikon*. Following Floridi, the ways in which the opportunities related to the digital society are spent will allow or not the informational organisms to produce the ultimate good, that is to say the creation of a good society.

***

Igor Pelgreffi – *Thank you for accepting this conversation, that would be on the themes of the algorithm and the problems it poses to us. We can also, if you believe, connect these issues to 'your' concepts, which are also very well known: I am thinking of the* fourth revolution*, of the* infosphere*, or even of* on-life*, the latter indicating an interconnection more and more extended between our historical condition of being-on-line and the vital or existential element of our life. So I come to my first question: can the algorithm be considered a* condition of possibility*, or even a* condition of existence*, of that all, that is of the infosphere, fourth revolution, on-life dimension? Of course, the algorithm is not the only condition of existence*

(there is also infrastructure, internet, mass mobile devices, etc.), but it is certainly one of the most relevant today. Here: how do you see the hypothesis according to which the algorithm is an element that supports *all this? An element that, at the same time, makes possible all this and also represents its* sustainability? *A sustainability, therefore, linked precisely to the properties, logical but also ontological, of the algorithm?*

Luciano Floridi – I will take a step back to approach your interesting question. The step back is the following. When we talk about Computer Science, AI, ICT, Digital Technologies, and the new wave of innovation, we may have two items in mind: data or software. And we know that we need both. It's a bit like cooking, you know: you need ingredients, and you need the recipes. Sometimes we emphasize the data rather than the software, almost as if the software were in the background doing all the job invisibly. But we should not forget that we need data all the way. We hear the metaphor 'data is the new oil', which I do not like. It is a metaphor particularly misleading, but it is very common. I prefer to think that data is the new gold, data are the new material, the new stuff that we manipulate in a digital society, almost as if software were there like a tool, something that you can not ignore as you work with these new resources.

Well, I think it is quite important, maybe as you were suggesting in your question, to realize that not only do we need both (because the one without the other makes no sense. Using the previous metaphor, it is like to say that with the engine without petrol or with the petrol without the engine: you do not go anyway). But also to realize that software has a support role that increasingly makes our society work.

The dependency that we are building with respect to software is growing daily. And it's a dependency that we do not notice: it works well, or even better, when you do not see it. If everything works smoothly, that's because the software, therefore the algorithm 'AI-oriented', is doing its job properly.

This reliance raises a couple of issues, if we have potential philosophical questions in mind. One is: how risky is this reliance on algorithms, as society becomes increasingly complex, and I mean *complex* in the technical sense, not just as complicated. Really complex: its elements are always intertwined with each other, and small local changes can reverberate and produce big differences in the whole system. I mean, and it is well known, that a system is complex since you cannot really imagine or calculate what can it happen to some corner of the system itself, when something really tiny occurs elsewhere. Now: as society becomes complex in that sense, increasingly, we need algorithms to run the system. But how risky is that? How much of our individual well-being and social welfare depends on the algorithm question?

To be honest, society has always moved towards interactions with complexity using more complexity. For instance, as we moved through urbanisation, we created more urbanization, as well as the car-based society, with more complexity in schedules, traffic, traffic control, traffic lights, if you like, etc. ... And again: more cars, more ways of handling cars, therefore more complexity and even

more complexity. Now, as complexity grows, we would like to see control, accountability, and transparency; all those items grow with complexity.

Thus, essentially we have two sides of the same point in front of us: as society becomes increasingly *algorithmic dependent* – data-dependent or if you like more generally data/software dependent, and therefore dependent on the digital – we should have a much better comprehension about what policies are in place to reduce the risks that all this implies, and also about the policies for better control, accountability, and transparency. Well, I think that in this condition, there should be some understanding of some risks concerning dependency on algorithms, specifically about what an algorithm or software could do when it goes wrong. Let us notice that it represents a completely different sense of risk. The risks we are talking about are not the risks when things go wrong, but when things break down. And it is a little bit like the risks you run when electricity is down, when there is a blackout: I'm not talking about being electrocuted or analogous damage. I'm talking about not having electricity in the house at all. That is the risk of dependency.

About that, I don't think there is enough sense of awareness, even if we have nowadays much more awareness of our dependency on energy sources. But for technologies is something different: depending more and more on them, we should have all the policies in place (resilience, for example, and so on). The same happens in terms of responsibility. We have a sense of responsibility that is almost based on the frame like 'when things go wrong', that is, when something is misused, overused or underused, for the wrong reasons.

In your questions, you said that there is an increasing dependence on algorithms, that possibly they support or sustain our society, so that the word sustainability can be paradoxically used here, but I think we have left the analysis and understanding of *good*, not evil risks. And we usually think the analysis and understanding of the policies for the governments of all this would be charged for the future generations, so we do not have to care about that, because one can think there are more pressing issues. On the contrary, this should be done today. I think that the risk management and the governments of the algorithms is an issue that we have to consider now, because it is now that we are building this dependency for the future.

Igor Pelgreffi – *Among other things, a final question of this conversation would have been about the future scenarios that you imagine. Particularly interesting, here, is the concept of 'functioning well', or 'too much well', which always accompanies that of 'functioning unwell': a duality that we could consider structurally linked to a certain logic of the algorithm, if you want.*

I take you now to the great subject of automatic learning *by an algorithm, that is, Machine Learning. Today the algorithm learns, and learns some kind of 'behaviour', but it can also work... badly: it is a remarkable boundary, even regaining from the discussion just made about risks. Could you tell me something about this*

connection between algorithm and learning? Does the machine learn? Can it make mistakes? Can it self-correct?

And, in connection to this, one more question: according to you, and if it is not just science fiction, is there a social dimension of machines? I mean: machines or algorithms that can 'communicate' with each other, that is, therefore algorithms that somehow learn but no longer individually but within a dimension social, even if the social term is probably too anthropocentric? What I am asking could seem extrapolated by a scenario from a science fiction film or novel, but in actuality we already have algorithms that feed themselves, 'nourishing' on data with increasing speed and acceleration, so much so that in some cases, it is no longer possible to have the technical control of the output.

Luciano Floridi – This is an important question. I think that what you called the 'social' aspect, I put into brackets the term 'social', I think that this is… anywhere… and it is not eccentric at all. I should bring a couple of examples, but as of now, I can remember that the problem of 'machine-to-machine interaction' has been with us since we started to use the first computers. Now, to anyone who has or uses a computer, probably the most elementary, almost trivial phenomenon or example of interaction, is the one you have between your computer and your printer: they need to talk with each other. In the past, we used to talk about so-called 'protocols of interaction' between two computers and their modem. That was called *handshaking*, speaking of metaphors: the systems were handshaking, finding a way of communicating through a protocol and sharing data between A and B, and B and A: between the two machines. It is just to say that all this has been with us since we started building the first computers. In fact, at the very beginning of the history of Computer Science – we are talking about the Fifties – the idea was that there would have been only a few immense computers and many tiny terminals having access to these mainframes. Therefore communication was everything, not least because maybe I needed to interact with another computer that was trying to access *the same* resource *at the same time*, in terms of scheduling access to the same resource, and that means having some sort of parallel, a-synchronic or synchronic access, and so on.

From a historical point of view, this was an old, big branch of Computer Science. Then, said briefly, things changed: we began living in a sort of 'do it yourself' dimension: my computer, your computer, my PC, your PC. A personal computer became *my own* machine, *your own* machine. Reaching today, in the actual stage of technical development, we are going back almost metaphorically to the mainframe idea of that sort of enormous computational power that is elsewhere. And, moreover, we were subjected to that computational power… So all this has been with us for some time.

The 'social aspects' of computation, and I put the 'social' into quotation marks, that is, metaphorically speaking, have received – for example, like in the communication I'm sharing with you now – a strong input from the well-known

aspect of artificial intelligence. Think of the so-called GAN, or Generative Adversarial Networks. It is a very common technology, no science fiction. It is used successfully in developing security systems, for example, the security of a website. These networks are 'challengers net': they actually play one against each other, literally. So: one tries to crash or to hijack the website, the other one is trying to defend it; one has no data, while the other is trying to use the inside data to defend: there is an attack, a defence, another attack, and so on, through a learning process. And there is a lot of what I would like to describe as *game theory* here, again: theory in a scientific sense.

So – going back to the previous question – we *support* a society increasingly through data and software, with software also being increasingly smart algorithms. It is also a world in which all of this is linked by significant interactions. Interactions are more than just sharing data, sending data, and receiving data. Interactions also mean modifying the data, generating new data and learning from the data and the interactions themselves, and therefore updating and upgrading the behaviour of the algorithms based on the interactions with other algorithms. And that is what happens in adversarial networks, for example. This is the future in the most ordinary sense, and it has nothing to do with Science Fiction, nor with the image of a few scientists that are dreaming in their office. Yet again: in a world where algorithms interact with other algorithms to learn from and improve their behaviour, the question about responsibility is: *who designs what* and *who enables what*. I mean: contrary to what happens in the animal world, algorithms do not interact with each other spontaneously, like cats and dogs fighting each other. No: you need to put them together, you need to make them communicate, you need to create the conditions for that particular interaction, and so on.

Now, it is this *design issue* that seems to me fundamentally human and, therefore, fundamentally ethical. For example: is it really a good idea to train an algorithm by interacting with another algorithm to make sure that it builds the best possible deepfakes? Or to enable an algorithm to learn how to exploit weaknesses in a credit card system? Once again: algorithms are playing with each other, interacting, improving, and then it becomes possible to hijack or crack the security codes for a credit card system or get my personal data. But the point is that the design of all this is in our hands: this is all about us. In other words, we entirely design the 'social' nature of algorithms: it is made possible by us. We are engineering these social interactions. And therefore, the responsibility for what kind of engineering is allowed, for what, whether we like it, etc., is 100% human. So that is the big difference about using 'social' with quotation marks. This is not the biological world, where animals or people socialise, but it is an artefact that we can build, that can or cannot, literally may or may not be enabled and allowed to do this or that.

Igor Pelgreffi – *Very clear. Thank you. Including the clear final ethical position you have just expressed...*

Luciano Floridi – Sure. Because it seems clear to me that, otherwise, we begin to think that technology really interacts only with technology, which I don't think. And then, at that point, we can also ask ourselves something like: 'what happens when you leave them alone'?

Igor Pelgreffi – *Yes, but they are still huge issues. I agree with you on this type of ethical requirement; however, it should still be remembered that the problem is* in re*: is the current reality. Just think of that branch of Computer Science that has recently emerged – for example, the so-called XAI (eXplainable AI) – which deals with algorithms that are 'intelligible', 'interpretable', or, more generally, 'explainable'. This need arises probably because algorithm programmers often report certain outcomes, actions or 'choices' of their algorithm (generated or written by them) that are not rationally explainable. But let's move on.*

It is interesting to briefly review your parable of studies on these issues, also in order to better understand your actual position. Your first books from the Nineties were more focused on logic and epistemology. I should remember: Scepticism and the Foundation of Epistemology. A Study in the Metalogical Fallacies *(Brill 1996);* Internet. An Epistemological Essay *(Il Saggiatore 1997);* Philosophy and Computing. An Introduction *(Routledge 1999). Gradually then the ethical aspect, perhaps present since then, seems to emerge more clearly, with consistency or, perhaps, urgency. I think of* Infosfera. Filosofia e Etica dell'informazione *(Giappichelli 2009) and many more books, including* The Ethics of information *(Oxford University Press 2011). In short: the ethical aspect has perhaps become paramount. I wouldn't say it denied them, but that it completed those early searches. And, in fact, it is also felt in this conversation...*

Luciano Floridi – Well… it is a very long project that I started developing a long time ago: decades ago now. When I embarked on this particular journey, initially, I thought I would publish three volumes. One on *the philosophy of information*, which would have been essentially logic, epistemology, metaphysic, etc.; one on *the ethics of information*; and the last one on *the politics of information*. One, two, three: ten years each, more or less, it means… thirty ears: that's my life! Unfortunately, things became a little bit more complicated. I realized that I needed much more space, also for publishing reasons: that book could not be a book, for Oxford University Press, of a thousand pages! So it ended up being volume one, *The philosophy of information* (Oxford University Press 2011), which is very much as you said, epistemology, so to speak, but also metaphysics. After that, I published volume two, *The ethics of information* (Oxford University Press 2013). But then, I realised that there was much more that needed to be done in terms of *The logic of information* (Oxford University Press 2019), which could become volume three [*laughing*]; but at that point, volume four, unfortunately... split into half. The reason for that splitting into two is because initially, I thought that *The politics of information*, which was supposed to be only a single book,

was going to discuss two forms of agency from the perspective of what can be called the digital revolution, that is the fourth revolution: agency as in socio-political agency (in other words the impact of the digital revolution on politics and society) and agency understood as 'artificial agency' (that is, understanding philosophically the ethics of artificial intelligence). Therefore, when I gave the *Ryle Lectures* at that time, I tried to keep agency at the centre of the socio-political and the artificial. But then I realised that it was just too much. I mean: the architecture on the volume was getting too complicated. And so, also speaking with the philosophy editor at OUP, I followed his recommendation to split the text into two parts. So this is just to say that volume four is now 4.A and 4.B [*laughing*], and volume 4.A is actually *The Ethics of Artificial Intelligence* (Cortina 2022), which I published in Italian and is coming out in English soon, I hope. I could say that now I have started the long way, the long journey towards *The politics of information*, volume 4.B, understood as the philosophy of the digital transformations being prompted by a different way of understanding political actions. It will take some time before I finish that book. Hopefully, there will be, after that, if I get lucky, I really hope to finish with a final book: number five, which is more a reflection on life's meaning, and in particular on what it means to be human in a digital age. At the moment, I'm calling it *The hermeneutics information* because I have no imagination! So we will have the philosophy of information, the ethics of information, etc., with, in the end, the hermeneutics of information that will hopefully also be the end of my intellectual journey [*laughing*]…

Igor Pelgreffi – *Thank you for this explanation around your path, or... your journey. In particular, the reference to hermeneutics seems to me to better clarify, at least in perspective, the ethical* quality *of your work: on the one hand, in fact, the use of an hermeneutics of information may integrate the perspective of the sign, of a linguistic and logical problem, and therefore, if you like, of the concept of interaction we have talked about, present since the beginning of your research path; but on the other hand it welds it more firmly to the theme of the human, of the return to the human, so to speak, within a hermeneutic instance of interpretation and domestication of these complex issues. Almost to limit the risks of them becoming divergent or hyperbolic...*

So let's go to the conclusion with one last question. The theme I am now presenting to you is really massive, so I can only ask you for a short answer that can show what is your basic position. The theme is that of corporeity. We know that algorithms (understood as a sequence of steps to solve a problem in a finite number of operations, etc.) represent an autonomous or abstract systematicity linked to the code, to the 'sign', to an impersonal protocol sequence of symbols, etc.; but we also know that nowadays algorithms tend to integrate themselves more and more deeply into living systems, in the psyche *or in our 'flesh'. Let us think about the theme of hybridization, if you want a reference: where does the algorithm start and where does it end, in my body? Where are our bodies – please note: not so much and not just the*

mind or the Subject, but the bodies – therefore, the attention, the psyche, our entire being onlife, *to take up one of your key-concept, are inside the living sphere, or the infosphere? In short: is the body really detached/detachable from the algorithm? Does it really hybridize, or is the so-called hybridization just... appearance?*

Luciano Floridi – Oh, this is another very interesting question. And I have to keep the answer short… In the past, I considered the body as an interface. I have not changed my mind. Maybe I will change it in the future, but at the moment, I still think that way. So: the body is an interface. This could be a synthetic answer to your very complex question. But what kind of interface? The interface between, on the one hand, the self, or the I, or the mind – depending on the philosophical orientation – the soul, even, for some people, our intelligence etc., something that is on this side; and, on the other hand, the bodies, other bodies, in other words, other interfaces, in the world. So, when we perceive the world, not only with our own interface, which is the body, we also perceive another interface, that is, the object which is on the other side. So imagine a conversation like the one we are having among us. What happens? It is me and my interface interacting with you and your interface. From that, you see that there are four items interacting: there is not only me and you, but there is me and my interface meeting, encountering, exchanging messages with you and, as well, your interface. This becomes obvious when someone has a different interface. I might be blind, and I have a different interface. Or I might be able to see or hear better than you, maybe using different interfaces. Although interfaces are also interfaces in terms of action, of what you can actuate using your body, so what you can interact with in the world: maybe I am strong, perhaps I am weak, maybe I am old, maybe I am young… And interfaces are also full of features: gender, colours (eyes, hair, skin), elements (legs, fingers etc.). Once you get the interface perspective, of course, this interface gets challenged when we shape it with the technologies we developed. But digital technologies not only operate on my body as an interface, but they also operate on my conceptual modelling of my body as an interface.

It is important, therefore, to remember that it is true that we have different abilities to work with the world, but it is also true that we conceptualise all this differently: to put it a little bit more strongly, what the digital does is to *re-ontologise* the body but also to *epistemologise our understanding* of the body. So, the digital does both, and sometimes it is in the asymmetric interaction between the re-ontologisation of the body and the re-epistemologisation of the body – from a digital perspective – that things do not quite work well. One thing I could mention, for example, is the *Quantified Self Movement*, that is – in short – people who are very keen to measure any bodily parameter, health information, and everything in their interaction with the world. So with their interface, wearable device etc., they may measure their heartbeats, breathing, the number of steps they take, calorie intake etc... For instance: 'how many steps I take', or even 'the level of red blood cells', sometimes even with chips under

their skin. Probably it is a small movement, up to now, and we do not have to emphasise it so much, but it offers some evidence of what kind of interactions we can have *ontologically but also epistemologically* with our body. We can speak, about this, in terms of *re-ontologization and re-epistemologisation* of the body. This means that one can see oneself as quantified minimally, for example, for the ten thousand steps you should take daily. The mechanism is that if I do not take the ten thousand steps, maybe I do not feel good with my self and with my body, at the end of the day. It is also self-training, therefore. By the way, you probably know that the question, the number of ten thousand steps, is not based on a scientific study; it merely started as a good round number in the Sixties, when it was advertised by a Japanese company Yamasa, which produced the first, successful step counter, called Manpo-kei, which literally means 10,000 steps in Japanese. Walking is good, but there is no reason to take that number of steps. But back to the main topic… The consequences are clear: these amazing technologies enable us to shape not only the world but also our interaction with the world, and all this happens through, and with, the interfaces we are, the body-interfaces through which we communicate and interact with the world. This may become an extreme thesis, which I do not like, things like virtual reality, new artificial skins, the idea that we can transfer minds through different bodies and so on. This is sci-fi. What is interesting for me is the cultural aspect: the deep, massive transformation in our self-understanding. That is, we are talking about a change in our philosophical anthropology: what does it mean to be human? What do we expect a human should and could be? So you see, here and again, the ethical implications of the whole question of algorithms and of their widespread diffusion, etc. All this will not happen one day to the next: it is a long, profound and gradual change in our self-understanding. In the same way as it took quite a long time since the agricultural revolution – many many years – to reach Aristotle and the social or political animal; and only after that stage the modern urbanization, for example, could take place: you don't have urbanization without a specific culture… But we should never forget that it takes millennia to get to the Aristotelian view, as a philosophical anthropology of the individual as *zoon politikon*, the political animal. Well, about all this, I think it will take quite a while for us to see ourselves as, as I like to call it, *informational organisms*. But that it's where we are going. And, again, we move through the flux of information, we exchange information, we can damage ourselves through the wrong use of information received, we can damage society through wrong interactions... but, apologies, we are running out of time … I better stop here with a final comment.

In everything I said, there is no attempt to develop a metaphysics, as if one could have access to the ultimate essence of the world. I don't know what that means. I'm not that kind of philosopher. I am much more a Kantian philosopher, in that sense. So: what is the world in itself? God knows, literary. I don't. We don't. But the world is conceptualized in a variety of ways, and today we are conceptualizing what we call 'ourselves' as *informational organisms*:

I mean, we are developing technologies that handle the informational side of those organisms. And once again, it becomes an ethical problem to make sure that we don't shape these informational organisms in the wrong way. At the end of the day, both education or *paideia*, and norms or *nomos*, are ways of shaping informational organisms in such a way that we could speak, in the end, of a 'good society'.

So, as philosophers, what can we do in order to prevent absolute and horrendous mistakes? That seems to me the pressing task in front of us. Once again, I think we have not thought enough about this, about the whole question of algorithms and their implications in the sense we said before. We should – as philosophers – focus on understanding more the digital revolution we are undergoing, to be able to design our future better.

Luciano Floridi
Università degli studi di Bologna
✉ luciano.floridi@unibo.it

Igor Pelgreffi
Università degli studi di Verona
✉ igor.pelgreffi@univr.it