

# BIG DATA, SCIENTIFIC RESEARCH AND PHILOSOPHY

Un coup de dés

Jamais

n’habolira le hasard

*Stéphane Mallarmé, 1897*

## THE DELUGE

In a novel by Isaac Asimov, an author that in times of Artificial Intelligence it would be good to rediscover, he imagines a huge computer called Multivac, able to answer any question one may want to ask it.<sup>1</sup>

In 1956 Asimov had thus already imagined the Big Data, the idea of a machine capable to digest enormous quantities of data but also to find more and more complex correlations among more and more heterogeneous data. But what remained clear for Asimov was that the production of truth (or rather of more knowledge) needed a third ingredient, one that Multivac could not have, the capability to ask the correct questions; thus, the role of the Grand Masters, humans with special intuitive capacities able to ask the right questions. The limit of Multivac is then not technical, it has to do with the essence of Truth, which cannot be reached only with quantity.

It is this limit that the most enthusiastic fans of Big Data fans believe can be overcome. In an article on the magazine “Wired” Chris Anderson claimed nothing less that “the deluge of data makes the scientific method obsolete”.<sup>2</sup> His idea is that the dimension of the machine and its databases are so “big”

---

<sup>1</sup> ISAAC ASIMOV, *Jockester*, 1956

<sup>2</sup> CHRIS ANDERSON, *The End of Theory: the Data Deluge makes the scientific method obsolete*, “Wired” June 23, 2008

that they can get to correct answers (working answers) without need to ask questions. The traditional scientific method, where one formulates a theory, then tests it experimentally and looks at the results, makes no more sense. Finding the right correlation between data is enough to say that knowledge has increased.

In a previous paper we have already highlighted the progressive "detaching" of Machine Learning from Artificial Intelligence as "continuation of Philosophy by other means"; the evolution of the theory of Big Data summarized above confirms us in this intuition, given that the algorithm's automatic learning and availability of data in increasing quantities are two sides of the same coin.<sup>3</sup> Instead of simulating or reproducing human intelligence, here a new type of intelligence emerges, an intelligence capable of acquiring new knowledge in a different and better way than humans, who remain analogic and limited in terms of storage capacity. Knowledge thus becomes a mere matter of quantities.

## **A NEW (A LA KUHN) PARADIGM FOR SCIENTIFIC RESEARCH**

This is no scandal. Research should be able to explore all the possibilities and to go in as many directions as possible. And it would not be very serious for a philosophical positioning to suddenly become the defender of the scientific experimental methodology, considering that for any philosophy worthy of the name sense-certainty could never be a "proof" for Truth. Besides, Chris Anderson's article is quite clear in the pragmatical (which means business) interest behind this new science paradigm; the theory about "the end of theory" should allow corporations to impose the market's timing to research, to do away with cautious scientists and their epistemological doubts.

As Thomas Kuhn teaches us, when a new paradigm arises the old paradigm's supporters implement various defence strategies. The excellent work of Sabina Leonelli<sup>4</sup> and her team does exactly that, to try to bring these new instruments of research within the tradition of experimental science, the one and only capable since Galileo to teach us something about the world surrounding us. The arguments here are not toothless, even without excluding Big Data in toto. It is interesting to notice that the conclusions of this defence are very similar to

---

<sup>3</sup> GIOVANNI LANDI, *Intelligenza Artificiale come Filosofia*, Trento, Tangram Edizioni Scientifiche, 2020, p.16

<sup>4</sup> SABINA LEONELLI, *La ricerca scientifica nell'era dei Big Data*, Milano, Meltemi Editore, 2018. Even more interesting in this book is the subtitle it bears "Five ways Big Data threaten science, and how to save it".

those promoted by those worried by the impact of AI on our lives: the need for an “ethical” approach, less pressure on researchers (less hurry in implementation even if it costs a couple points of ROI), more democratic users’ participation in the decision making.

These are all valuable suggestions, suggestions one can agree to, especially considering the courage it takes to make them against the propaganda steamroller the whole world, and not just the data industry, is being subject to. However it is a rear-guard fight, like with the Ptolemaics trying for centuries to adapt their theory to astronomy anomalies. Chris Anderson’s change of paradigm of scientific research has many analogies with Galileo’s arrival on the scene.

The unbeatable anomaly, in this case, is the fact that Big Data works, that Google has become the largest advertising company not through customers’ introspection but only looking at data; it is the fact that we worry about our privacy, but we are ready to give it away for some free apps; and so on. “*A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it*”<sup>5</sup> says Thomas Kuhn, and it is hard to believe that the next two-three generations of data scientists won’t be Chris Anderson’s followers. The “deluge” is coming, and it is coming fast.

## **THE ARK OF PHILOSOPHY**

How can we face the “deluge”? Where can we find an ark strong enough to keep us afloat? What can we answer to the claims that a duly trained algorithm could forecast if and when we will have cancer, by relying on 2-3-4 thousand of our voice’s parameters? Or that we won’t be needing passwords anymore as the algorithm will recognize us thanks to unconscious patterns it can detect in the way we press buttons? After all, if millions and millions of observations between two phenomena show a statistically significant correlation how could we not accept the idea that there is there a relation (maybe even a causal relation)? And how could we not accept the idea that with even more data and more triangular correlations we would finally get to some kind of certainty?

---

<sup>5</sup> THOMAS KUHN, *The theory of scientific revolutions*, Chicago (Usa), University of Chicago Press, 1962.

The problem, however, is that Data, no matter how many we collect, cannot give us an answer because in front of them there is nothing less than the infinite. And the same goes for the triangulations and crossings we can devise. This seemingly endless wealth of data loses all truth when it comes to knowledge:

*« Because of its concrete content, sense-certainty immediately appears as the richest kind of knowledge, indeed a knowledge of infinite wealth for which no bounds can be found, either when we reach out into space and time in which it is dispersed, or when we take a bit of this wealth, and by division enter into it. Moreover, sense-certainty appears to be the truest knowledge; for it has not as yet omitted anything from the object, but has the object before it in its perfect entirety. But, in the event, this very certainty proves itself to be the most abstract and poorest truth”<sup>6</sup>*

Once data are too many to be understood, and all that remains are the correlations a software program can find, it is obvious that any kind of correlation can be found among any kind of data. Of course a supercomputer can find meaningful correlation between the thousands of parameters which constitute our voice and, say, a particular disease; but the same supercomputer could find as meaningful correlation between the same disease and, say, other data like for example the rate of growth of our hair. And the same supercomputer could subdivide our voice in not thousands but tens of thousands of parameters, with maybe different results. Where could we possibly set the limit?

It is the infinity, both in extension and divisibility, that render the parameters epistemologically useless.<sup>7</sup>

This kind of predictive extrapolation, even if they sound scientific, does not even hold common sense analysis. And this is where the storytelling comes into play, with the same strength Plato denounced when talking about the arts of Sophism. Sabina Leonelli rightly underlines the contradiction of our time: on

---

<sup>6</sup> G.W.F.Hegel, *The Phenomenology of Spirit*, Oxford, Oxford University Press, 1977, p. 58 (first edition 1807). We reproduce this passage from the Phenomenology of Spirit only to prove that technical progress does not bring forward new questions, as many claim today; these questions have all already been asked, and technology is not per se the answer.

<sup>7</sup> The Hegelian concept of Bad Infinite (schlechte Unendlichkeit) is relevant here for those who are interested

one side an apparently hyper-objective (because based on pure data) scientific method, which on the other side gives birth to countless interpretations:

*“This ocean of data inevitably transforms itself in a series of contrasting interpretations. We find data according to which drinking wine is bad for our health, but also data which prove that people with a heart condition should drink one glass of wine per day. Data which confirm the negative impact of plastic on the environment, but also data which deny it.....And – this really is the most disconcerting aspect of it – people who come to opposite conclusions using exactly the same data....our judgement on what to believe comes to down to a judgement to whom to believe.”<sup>8</sup>*

But the contradiction is only an illusion. Big Data and Storytelling are two sides of the same Digital Transformation the world is going through. No matter how “Big”, Data cannot give us Truth or even Certainty for that matter, so that what remains is the rhetoric ability to give one’s point of view with the right set of data to support it.

Therefore it is not enough to remind ourselves that *data lakes* can become obsolete, or that data collected with an objective in mind cannot necessarily be used for other scopes: these are all technical precautions, which do not touch the heart of the problem.

Sabina Leonelli rightly questions the epistemological status of Data, and she rightly concludes that they be taken as the objective representation of a fixed Reality. But defining them “relational” does not help a lot, because relations have to face the same infinite which makes everything senseless. It does not matter how many relations we can find or imagine with our computers, there will always be an infinity more to find or imagine.

The question on the epistemological status has to reach Reality itself, something a purely epistemological (or Philosophy of Science) approach cannot do because it takes Reality as granted, existing, external and given. The coming back, or better said the rediscovery, of this way of asking questions is the Ark we will need to face the “deluge”. And this way of questioning is what makes it a truly philosophical method.

---

<sup>8</sup> SABINA LEONELLI, *La ricerca scientifica nell’era dei Big Data*, Milano, Meltemi Editore, 2018, p. 4 (our translation)

## CONCLUSION

It is impossible to know what technology will bring us in the years to come, and it is ok to accept this: what matters is how we position ourselves in front of it, it is the ongoing and tiring effort to understand, without following the advertisers' sirens. It is important, as the work of Sabina Leonelli shows, to deny technology its triumphalism only because "it is bigger" (as if Big by itself was a guarantee of bigger Truth).

But Philosophy is not just demystification, it is understanding. And to begin with, it is understanding that understanding does not mean to find solutions: that is the job of sciences, whatever their methodologies may be in the future. Philosophical comprehension is something else, it has to do with what people, not with what people do or think they are doing. It can seem a useless effort, but it is what remains in history, technologies and even societies come and go, while the Truth remains. And this is why we do not accept the idea that "we live in the era of post-truth"<sup>9</sup>. Truth, if it is Truth, exists forever, regardless of how many followers it may have.

---

<sup>9</sup> SABINA LEONELLI, *La ricerca scientifica nell'era dei Big Data*, Milano, Meltemi Editore, 2018, p.