Giovanni Landi – November 2020 – Independent publication

ETHICS OF ARTIFICIAL INTELLIGENCE VS. ETHICAL ARTIFICIAL INTELLIGENCE

INTRODUCTION

A line of thought arguing that Artificial Intelligence is first and foremost "philosophy by other means ", and not simply a wonderful technology, must inevitably confront the ethical debates that have been going on for years on the responsibilities, moral choices, decisions that the advent of these systems will require. At the same time, this is not an easy choice: the debate, indeed the ethical debates on IA, are so many and varied as to discourage a philosophical approach, precisely because such debates are too "ethical", and this be said without irony.

In almost all these discussions, in fact, "Ethics" means a moral reflection on behaviors, responsibilities, rights and values; but all this, while correct, is at the same meaningless if one does not consider at the same time Ethics not as "a field" of Philosophy but philosophy at large and nothing less. Problem specificities, or fields of study separation, does not work in Philosophy. On the contrary, everything is held and must be held together, otherwise we fall back quickly into edification and contradiction. The specificities of ethical debates cannot be separated from the rest of the Truth, and in this concrete unity lies the difference between a philosophical truth and other types of truth. Forgetfulness of this specific unity produces two opposing yet mirrorly identical results: on the one side the ethical debate is forced on the path of abstraction (as in the famous cart dilemma and other problems similar to game theories[1]); on the opposite side the thirst for concreteness is satisfied with pragmatic solutions that perpetuate the existing while a political, economic and social revolution comparable to the industrial revolution of the eighteenth century is announced.

[1] For example, PATRICK LIN, "*Ethics of autonomous cars*", in The Atlantic, 2013

We will outline both these outcomes when we are meet them in the course of our discussion.

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Most ethical questions on AI, especially at the journalistic level, concern the so-called biases[2] present in machine learning algorithms. These algorithms, which run on neural networks, learn by extrapolating from large amounts of data correlations trends not noticeable at first sight. Once "trained", the algorithms are used to predict possible evolutions on new data homogeneous with those on which the training took place.
The problem, of course, is the difficulty, some argue the impossibility, of making sense of the parameters on which the predictions are based, which is another way to say making sense of what and how the algorithm has learned exactly; reverse engineering simply does not allow us to reconstruct the algorithm's self-changes, so that they are rightly described as "black boxes" without an understandable logic. Tons of ink have been used to find possible solutions to this problem, which concerns the quality of the data used for training but also the complex mathematics that governs self-learning choices of the algorithms.

One of the most striking and well-known cases is that of the algorithm developed by Amazon for the selection of candidates to technical specialized jobs:

"Once implemented, this software had to use sophisticated AI algorithms to learn key traits of the successful candidate resume for a period and look for similar features in curricula submitted for screening."[3]

After a certain period it was noted that the algorithm took into account (negatively) CVs of female candidates, probably because a large number of CVs of male candidates had been used for its training. Amazon, obviously sensitive to the issue, announced the pure and simple abandonment of the program. Considering gender as one of the elements to be taken into account for selection was not considered appropriate, as well as being legally punishable. It should be noted that the simple removal of any such indication in CVs was not deemed sufficient, as the algorithm could still extrapolate it from other seemingly neutral data.

---

[2] Bias it's a word somewhat strange whose etymology is ignored. It seems that the word comes from ancient French and indicated originally the tendency of a bowl not perfectly homogeneous to deviate from a straight-line path. Today it is practically used as a synonymous of “prejudice”.
[3] AKHIL ALFONS KODIYAN, *“An overview of ethical issues in using AI systems in hiring with a case study of Amazon's AI based hiring tool”*, 2019, p. 1

The short history of Machine Learning is packed with episodes such as this. And once suspicion is raised, how can one consider entrusting a ML system with decisions on bank credit assignment, or prisoners' parole, or arbitration in civil disputes, and so on?
To put a remedy on this loss of confidence, which among other things is a big threat to a very promising business, proposals have not been lacking. It is the so-called Ethics of Artificial Intelligence:

"The response to unregulated artificial agents tends to be of three general types: avoid algorithms altogether, make underlying algorithms transparent, or control algorithm output. Avoiding algorithms is probably impossible; few other options are available to make sense of the current deluge of data. Algorithmic transparency requires a more educated audience that understands algorithms. But recent advances in learning technology suggest that even if we could deconstruct the procedure of an algorithm, it may still be too complex to give a useful meaning to that intuition. Christian Sandvig's recent work argues that the last option, the algorithm audit, should be the way forward (Sandvig et al., 2014). Some types of audits ignore the internal functioning of artificial agents and judge them on the basis of the fairness of their results. This is similar to how we often judge human agents: from the consequences of their outputs (decisions and actions) and not from the content or ingenuity of their code (thoughts). This option makes more sense for policymakers and sets the standard for an ethics of accountability for artificial agents. Regulation is much easier in this context."[4]

To begin with, we eliminate even the possibility of doing without these systems, thus eliminating social and political human freedom to decide one's own destiny. Then we eliminate the possibility of understanding how these systems come to their conclusions, thus declaring them endowed with freedom (the algorithms, not the humans). So in the end humans are left just with the option of judging the decisions according to their sense of fairness.
In our opinion, this is a striking example of the pragmatic approach we mentioned in the introduction. Let's ask ourselves for example, who should be the judge of the 'fairness' of an algorithm decision on a denied bank loan. It could certainly not be the person who has been refused the loan credit, and of course it could not be the bank that refused it; we should then entrust the judgment to an external arbitration panel (composed of course of humans, because if another ML system is involved the issue would go on forever). At this point it is not clear anymore what the advantages would be from using ML in the first place.

---

[4] OSONDE OSOBA, WILLIAM WELSER IV, "*An intelligence in our image*", RAND Corporation, Santa Monica (CA), 2017, p. 25

And we can ask the same question in the case of Amazon's algorithm. It discriminated against women, it is argued, but it is not clear who should decide and by what parameters whether a different decision would instead be considered "fair". What is the right proportion of female candidates that should be hired in order not to raise any warning flag or not to trigger an investigation into the algorithm's decision-making process?

Having recourse to the law does not solve the problem either, the situation is equally uncertain, despite the use of strong words to convey the intention of being inflexible on this point:

"In essence, who do we hold responsible if an algorithm throws an innocent person in jail or diagnoses an incorrect cure: with the user, with the mathematician or with the manufacturer? According to Floridi, it is necessary to move from the concept of objective responsibility, "which provides that in the event of a serious malfunction it is the manufacturer who must prove his innocence." [5]

We can find out if the person condemned by the algorithm is actually innocent only by making a new trial, and the same goes for the diagnosed cure, whose incorrectness can only be established by a doctor. If this human double-check cannot always occur, who decides which cases to review? And if we decide instead to always double check always then let's save time and not use algorithms at all – otherwise algorithms' decisions will be considered as long as criminal justice and its three degrees of judgment.

The inadequacy of this pragmatic approach, laudable in its intentions, must also be measured by the trumpeted promise of Artificial Intelligence as a "revolution" equal to the industrial revolution of the eighteenth century. Revolution is a heavy word, it should be used with caution except by professional marketing people. The industrial revolution happened together with the French Revolution, a reversal of all the epoch's principles; it is risky to announce such a change is the technology behind it is not up to the task.

BIAS AND ETHICAL CONTRADICTION

So what do we have left but a sense of pessimism or pure and simple rejection? What remains is the effort to do ethics as is it done in philosophy, not a specialized discipline but a discussion where ethical issues do not go separate from epistemological, logical and ontological ones. And where the ethical

[5] LUCIANO FLORIDI, interview in "*Flying machines*", October 2017. Luciano Floridi is one of the most cited experts in artificial intelligence ethical issues in existing literature.

questioning does not align with the logic of the object of study, in this case machine learning.

We should start with acknowledging that things are not simple not because of technical issues or because of bad faith on the part of some of the interested actors, but simply because bias (prejudice) is an integral part of any ML algorithm. In philosophical terms, prejudice ***is the very essence of Machine Learning, not its flaw.*** To be trained to make decisions an algorithm must be fed data, and data are prejudice since they discriminate, as they are based on difference and not on unity. For there to be a judgment there must be prejudice (pre-judgement), otherwise there is literally no decision to be made. The choice of data to be omitted (e.g. the gender of candidates) would also be a prejudice, justified perhaps socially but without any value of truth. Why omit gender and not age?
To reduce the bias is to reduce the algorithm; to delete the biai would simply mean to eliminate it. As we have seen that this is precisely what happens in the approach which calls itself pragmatic, where the last control retained by humans would result in greater waste of time than originally saved or in other forms of prejudice as to what results to double check or not.
Incidentally, let’s note that any "corrective" solution taken in this perspective would perpetuate the existing order, whatever it may be. It is a kind of risk particularly felt by the law community, who knows from experience how principles evolve in its domain. Justice is one of the most conservative areas of human activity, when in doubt the judge always finds comfort in past decisions, but Philosophy of Right is very clear in always leaving the door open to new interpretations, even radical ones, which may be ahead of their time but will be rediscovered and become in common use when the right time comes. This ability to listen, to review the context in addition to data, is one of the best qualities that can be found in a judge, it is jointly responsible for all the progress of the Law. And the context can never be reduced to a given.

So what is the ethically philosophical problem of Machine Learning, the fundamental one, what remains when we let go of the accidental?
Let's take the case of Amazon's algorithm again, forgetting for a moment the discrimination scandal:

The company's experimental hiring tool used artificial intelligence to give job candidates scores ranging from one to five stars - much like shoppers rate products on Amazon." [6]

[6] JEFFREY DASTIN, "*Amazon scraps secret AI recruiting tool that showed bias against women*". In: https://www.reuters.com/article/ us-amazoncom-jobs-automation-insight/ amazon-scraps-secret-ai-recruiting-tool-thatshowed property -bias-against-women-idUSKCN1MK08G (2018)

Male (and female!) candidates were therefore assimilated to products one can buy on Amazon. Obviously the idea behind such a system was that a candidate is a simple piece to be inserted into a mechanical gear, and in fact the parameter used to form the algorithm were CVs previously received of candidates hired and whose recruitment had been judged a success. Human beings learn not only from their successes but also from their failures, and the same should be possible to say about companies if they are places of human community and not just mechanisms. The choice to teach the algorithm only successful cases again shows a vision where an individual is only the sum of his acts and not also the ability to improve himself, in short, a vision of the human being as a mechanism.
The point here is not to criticize the specific policy of Amazon's Human Resources Department, but to understand that this vision of the human being is exactly the one behind Machine Learning, as it clearly emerges when one reads the theoreticals of this discipline. Human beings, single or in society, are seen as essentially repetitive, predictable in their actions, be analyzed as any other object of nature because subject to laws (internal, psychological, social, etc.) of a scientific nature. In short, the human being is conceived without freedom, a freedom that is instead attributed precisely to the algorithm in charge of studying it. The potential to change itself (learn) over time without being capable of identifying the causes of such self-changes ("the black box") means that these self-changes are not the result of a deterministic causal process, that they are the result of a free choice. In previous works we have already indicated this peculiar ideology of Machine Learning,[7] which is also reflected in the replacement of "Artificial" with "Machine".[8]

Let’s take the European Commission's "Ethical Guidelines for Reliable IA," certainly one of the most impressive institutional efforts to find acceptable solutions. They say:

"The foundation that unites these rights can be understood as rooted in respect for human dignity, thus reflecting what we call an "anthropocentric approach" in which the human being enjoys a unique and inalienable moral status of primacy in the civil, political, economic and social field."[9]

The indisputable value of these concepts strides, precisely because of their altitude, with the vision of humans advocated by Machine Learning. One cannot uphold such principles while at the same time conceive that human beings are thinking “machines”, that phenomenal consciousness does not exist

---

[7] GIOVANNI LANDI, *Artificial Intelligence as Philosophy*, Trento, Tangram Scientifich Editionse, 2020
[8] GIOVANNI LANDI, *"Machine and Artificial, Artificial Consciousness, New Year”* In www.intelligenzaartificialecomefilosofia.com, independent publication, 2020.
[9] Curated by the High Level Expert Group on Artificial Intelligence, p. 11, April 2019

or is just an illusion, that neurons are the cause of human emotions and feelings.[10] There is a contradiction between *that unique and inalienable moral status of primacy in the civil, political, economic and social field* that one wants to preserve and a technology (Machine Learning) that denies this primacy, that reduces learning to a statistical calculation, that presents itself as merely another "mechanical limb" to support man, with the difference that the limb to be replaced is not a leg or arm or even the heart, but our brain.

To reveal the contradiction, to bring it out in its pure form, this is the (only) task of a truly philosophical ethic. It's a time-consuming task, which requires from philosophers to patience to study Artificial Intelligence down to the detailed technological aspects, and requires AI researchers to understand what an ethically authentic philosophical questioning is. Plato said that only when kings are philosophers and philosophers are kings will there be perfect government. Similarly, we say that ***ethical AI will only be possible when AI researchers are philosophers and philosophers are AI researchers.***

CONCLUSION

It is easy to understand that no proposal, solution, answer or suggestion that refers to questions on the ethics of Artificial Intelligence will be found in this text. Actually, why are these questions called or defined "ethical"? After all, these are social, legal, economical and in the end political issues, to be publicly discussed as rules of democracy dictate. Everyone's opinion should have the same weight, a specialized "ethical" competence does not exist, we are not in a scientific field (such as biochemistry or meteorology). To do otherwise would take us down a dangerous slope. Let's see why:

"Moreover, Floridi argues, "a second analogy even goes back to Roman law and concerns the relationship between master and slave in ancient Rome: although the slave is obviously smarter than any robot we can ever build, when he committed a crime the legal and economic responsibility fell on the owner". Floridi concludes: "The Romans knew very well that if they laid all the blame on slaves, they would be completely disresponsibly. In this way, however, it

[10] Let's clearly state that we do not intend to arouse suspicions about the opinions or the intentions of the mentioned experts; we only intend to faithfully report what the theory of Artificial Intelligence (be it Philosophy of mind or more generally Cognitive Sciences) explicitly declare.
We refer to our previous works - www.intelligenzaartificialecomefilosofia.com – for a detailed analysis of these declarations.

was made sure that the master was careful and kept the situation undercontrol. Which obviously didn't stop the slave from finishing crucified..."[11]

It surely sounds like an appropriate solution, pragmatic to the right point, that's why Floridi uses the analogy. Besides it comes guaranteed by the Romans, a notoriously practical and non-philosophical people, great innovators in subjects of law. But the obvious contradiction in the reasoning, namely that if the slave is not responsible he should not be crucified, that is something the pragmatic solution does not say, precisely because it is Ethics completely detached from Philosophy. The goal of preserving the existing (slavery) makes the ethical solution itself philosophically unethical!
Similarly, it is philosophically unethical to support the social acceptance of Machine Learning on the basis of the "missed opportunity costs" that we would otherwise have to pay. The “missed opportunity costs” theory says that we should calculate how much individual and social freedom we are prepared to give away by having in return the advantages that Machine Learning would give us.
The problem is not that "missed opportunity costs" are assessed differently by different people with different roles and places in society; and it is not whether there is a formula for calculating the common benefit beyond the individuals. The philosophical (and ethical) problem is that even if the advantages could be calculated, freedom cannot, precisely because humans bear the primacy mentioned above. To quantify freedom is to align one’s reasoning with the same ideology of Machine Learning that we have above underlined, it is adopting ethical but philosophically non-ethical solutions.
But this also means that the opposite reasoning is equally unethical: a society that renounced Machine Learning and Artificial Intelligence all in the name of freedom would quantify the latter in exactly the same way.

A society that is capable of understanding this is a society that can find the way for an "Ethical Artificial Intelligence"[12], a free society who dominates its productions and technologies instead of being dominated by it.

[11] LUCIANO FLORIDI, interview in "*Flying machines*", October 2017. Luciano Floridi is one of the most cited experts in artificial intelligence ethical issues in existing literature.
[12] As opposed to the various Etics of Artificial Intelligence.