

Which ethics for the new (digital) world?

**Guido Boella, Tiziana Catarci, Amedeo Cesta, Ines Crispini,
Piergiorgio Donatelli, Maurizio Mori, Giovanni Sartor**
Società Italiana per l'Etica dell'Intelligenza Artificiale
(www.sipeia.it, info@sipeia.it)

Abstract

This paper describes SIPEIA, the new Italian Society for the Ethics of AI, established by the authors of this paper to fuel the scientific debate about Ethics and AI and to disseminate the discussion among citizens and institutions.

1 Introduction: the SIPEIA Manifesto

SIPEIA, the new Italian Society for the Ethics of Artificial Intelligence, founded by the authors of this paper at the end of 2020, is deeply aware that the introduction of technologies based on Artificial Intelligence will profoundly change the way of life that today is known and familiar to us. The use of machines capable of performing actions hitherto carried out by humans will involve a profound modification of the current social arrangements.

Without considering the status of intelligent machines - a non-negligible but future problem - we note how the scenario that is looming involves a sort of redesign of social life, since intelligent machines will significantly modify the conditions of our existence. The task ahead is exciting, but it is not without its dangers. The new machines open new horizons, but there are also negative aspects, which could also be amplified. Therefore, great attention is required to be able to govern the process. For this reason, ethical and legal analysis becomes central.

We cannot here examine all the various problems that arise, but it seems decisive to us to establish some basic points crucial to be placed at the centre. We are against the so-called *inevitabilism*, that is the perspective according to which the development of machines will continue according to its own logic, regardless of human needs. We are well aware that the interests at stake are great, and that the rhetoric of inevitabilism is rather appealing when intelligent machines are involved: we believe that processes can and should be governed, and for this reason we identify three areas in which intervention is urgent and dutiful.

The first concerns the world of work. We know that work as we have known it will be rapidly transformed, because robots will replace human operators. Think how even a complex job like that of the pizza maker can now be done by a robot. This means that the advances of AI will lead to the disappearance of many traditional jobs and the birth of new

professions. This has already happened, since, for example, at the beginning of the last century the farrier's profession was flourishing, which gave way to that of the tire dealer. However, there is a need to reflect on the speed of implementation of the change process and above all on its final goal, which must be to ensure a richer quality of life for everyone. In this sense, we welcome the progresses of AI, but we believe three points should be considered as indispensable:

1. We must seriously address the question whether highly technological work allows human self-realization.
2. We must seriously think about decreasing working hours to guarantee humans the opportunity to have free time for self-realization.
3. We must seriously think about how to fairly redistribute the wealth produced by the intelligent machines, the quantity of which is destined to an enormous increase.

The second area on which it is urgent to intervene concerns the world of economics, which today is based on the primacy of the market and on the liberal assumption that centralized management of a complex production system is not possible due to the cognitive limitations of humans who are unable to adequately collect, merge, and process all localized information. As machines vastly exceed human capacities in sensing, storing, processing and transmitting data, these limitations do not apply to them. Consequently, new ways of managing production become possible, as it is shown by the unlimited capacity of scaling-up of leading internet companies. This assumption also needs to be examined with great attention: we do not have pre-established solutions, and rather believe that the exercise of critical thinking is fundamental, especially in this foundational phase of new social arrangements.

The third area of intervention concerns the problem of monopolies connected with the information industry [Wu, 2010]. The trend in this regard seems to be increasing since the data needed for machine learning is available only to some companies capable of processing it. This aspect touches a point of enormous importance, because monopolies can constitute a threat to democracy itself. SIPEIA is in favour of an expansion of democratic prerogatives and a decentralization of powers. There are many other issues to be addressed, but

those mentioned seem to be in a priority position. SIpEIA intends to open academic discussions to see how it is possible to control the process of rapid transformations taking place.

2 Research topics

SIpEIA focuses not only on the ethics of AI, which addresses the ethical issues raised by AI. It also addresses ethics and AI: how AI will affect Ethics itself. Moreover, it adopts a multidisciplinary perspective considering also legal issues raised by AI.

What is morality? Ethics or morality is a peculiar social institution. An institution is a social construct. An example of institution is law. Law is a solid institution as there are defined places in which laws are created and applied, according to pre-established procedures. There are other institutions that are not as solid as law. For instance, there are semi-solid institutions like politics. Politics is in part solid like law (there are parliaments, parties, political acts, etc.) but it is also fluid: when you talk to a friend in a bar, your talk could be a political act. Elections are a solid part of politics, but when you suggest to your neighbour to vote someone this is very fluid. Having distinguished two different sorts of institutions, we can introduce another great institution: language. This is a prototype of fluid institution. The meaning of words and the structure of language is not invented by anyone of us, but rather given. Yet certainly it is not solid (there is no such a thing as the “house of language”). Language is spread among people. Language does not depend on anyone and yet it exists. Morality is a kind of institution like language. It is spread in society and it is similar to language in this respect. No one owns morality. Morality is like the wind: you can feel it, but you cannot capture it. It has an influence on you, but it is difficult to grasp as it is pervasive and has no definite location. The analogy between language and morality is classical and it appears to be adequate, at least so far, i.e., up to our time or a few decades ago. For this reason, our view is that morality is secular. Morality is a social institution that we follow because it has norms that are rationally justified. The goal of morality is to achieve the adequate level of social coordination needed to allow personal self-realization. When the circumstances change, social coordination may modify as well and therefore some norms require to be qualified and changed. To have a rough idea, think of dressing: in the winter, when it is cold, we wear heavy coats and hats in order to be warm, but in the summer, when it is hot, we dismiss them and we wear linen instead. In our time the historical circumstances changed radically, and in fact moralities change because people try to adequate norms and values to the new conditions.

Thus, morality as a social institution can change along time. It is not the first time that technology brings changes in morality. In the '70s there has been a revolution within Ethics as a consequence of the advances on bio-technologies. It is possible that AI will force further innovation in Ethics.

There are already several signs that AI will change morality. For instance, the possibility of recording everything and to

keep any sort of data is going to modify our sense of history. For thousands of years humans lived without scripture and recording the facts. That was pre-history. Then 5,000 years ago humans invented writing in order to record what is happening. In this sense, we all have birth certificates, pictures of graduations, of weddings, etc. This is history: History is a time in which we have some written documents recording of some events. “Pre-history”, a time in which there were no written recordings, while history is characterized by the recording of the most important events: actions, treatises, etc. Now the situation is quickly changing. Let's take Venice as an example: tourists used to take few pictures of Venice in the past. Now, tourists in Venice record everything. This is a dramatic change. We are entering what we propose to call “post-history”. This empowerment will change deeply the historical condition and for this reason our morality will be modified. This is the main point of our thesis. In post-history we have recordings of everything. An important aspect is that all this information can be examined and analyzed. They can be used for good purposes, like to detect in advance medical problems. Yet also for bad purposes. This will deeply change our lives. We are entering in post-history in which everything is recorded, duplicated, and it will be hard to get what is relevant. For this reason, Big Data, data mining are aspects of the future.

We are in a stage in which we start something new because history as we know it will cease to exist. Of course, this is a remark concerning our description or representation of the matter. In reality, the flow of human events will continue as always, but will include technological events. And this is a real problem because in essence technology enhances humans. We are already enhanced humans. We are no longer mere humans. We are supermen/wonderwomen! This is a problem for ethics because is not to be enhanced normal or is it a lack of something? Everyone has a mobile phone. Is not having a mobile phone like having something less than the others? Is it a kind of deficiency or defect? This is a permanent change in our condition and therefore it will influence morality. This is a great problem. Algorithms will change our real life because life will be strongly conditioned by technology, peculiarly AI. This new technology will be pervasive and it will induce a moral change. It is too early to predict the lines of the change, but it will occur eventually.

Moreover, we now think mostly of AI as a tool, as other technologies, but this is a reductive perspective, since AI contributes to shape our social reality, in ways that we have not planned for. The historian Yuval Noah Harari [Harari, 2014] reminds us that the success of Sapiens with respect to the other human species is not due to the invention of tools. The real explosion of Sapiens happened with the cognitive revolution 70,000 years ago, a change that allowed us to talk not only about the reality, but also about fictional entities such as myths and gods. 6,000 years ago, in the middle east, artificial fictions such as temples, gods, and then laws, allowed the Sumer to coordinate their society. Temples were employing people, owning lands, collecting taxes, etc., like they were real persons. Harari notices that these fictions are not so far from the modern concept of corporations, legal

entities who own properties and employ people: they have a legal personhood (now detached from religion). Thanks to the possibility of coordinating people using such fictions, in a couple of millennia empires of over one million people became possible.

The American philosopher John Searle [Searle, 1995] explains that we can coordinate in such a way our society thanks to the fact that we all share norms and costumes that create a socially constructed reality in which we are embedded in a seamless way: money, citizenship, laws are all fictional entities that work as long as we all agree upon them. Technologies up to now had impact on our lives and only indirectly affected the socially constructed reality. Probably AI will have an impact on this dimension. So, the greatest impact of AI may not be autonomous weapons but how it will change these structures surrounding all of us. Consider, as an example, the notion of “Public sphere” by the German philosopher Jürgen Habermas [Habermas, 1989], the realm of social life where public opinion is formed by public discussion: the social sites or arenas where meanings are articulated, distributed, and negotiated, as well as the collective body constituted by, and in this process, “the public”. These mechanisms are increasingly substituted by personalization algorithms of social networks, deciding which different content to provide to each of us. How this change in the functioning of “Public sphere” will change our socially constructed reality?

Finally, up to now, many scholars, such as [Floridi and Cowls, 2019], proposed to adopt for AI the same principles developed by bio-medical Ethics: beneficence, non-maleficence, autonomy, and justice. Take for example the beneficence principle: “*Beneficence: promoting well-being, preserving dignity, and sustaining the planet*”. At some point some AI system will be used to provide information and judgements for the decision maker that must ethically evaluate another AI system according to these principles. So, we will use AI to evaluate AI itself from the ethical point of view. We have here a circular way of applying current ethical principles to AI, which would invalidate the use of these principles if their wider context of use is gone missing. For this reason, it is not possible to simply import them from the field of bio-medical Ethics in the AI field without further considerations.

Moreover, we should not take for granted that a list of principles is the right approach to AI Ethics. As claimed for example by the advocacy group Algorithm Watch¹, the generality of AI ethical guidelines makes them ineffective: paper tigers. But also, as the authors of the *Principles of Bio-medical Ethics*, [Beauchamp and Childress, 1979], write: *Principlism, then, is not a mere list and analysis of four abstract principles. It is a theory about how principles link to and guide practice. We will be showing how these principles are connected to an array of transactions, practices, understandings, and forms of respect in health care settings, research institutions...*

Beauchamp and Childress ground the principles in common morality: “*norms about right and wrong human conduct that are so widely shared that they form a stable social compact*”. The identification of very broad principles is only part of the contribution of Beauchamp and Childress: the other part is to use specification, balancing and (deductive) application to create a bridge between the moral situation and the relevant principles. In the Ethics of AI field, we are still lacking the array of transactions, practices, understandings, and forms of respect, that make concrete these principles. Hence, importing principles, as many authors do, may not be the right way to proceed. We should learn from the history of bio-medical Ethics in creating a new Ethics of AI tailored to the new risks created by technology.

Coming back to the ethical problems created by AI, the list is endless and here we can only remind some of them:

1. Bias in machine learning algorithms: As Ruha Benjamin [Benjamin, 2019] said, “*feeding AI system on the world’s beauty, ugliness and cruelty but expecting it to reflect only the beauty is a fantasy*”. Bias is not simply a type of technical error: it concerns human beliefs, stereotypes, or forms of discrimination. Limiting discussion to the statistical meaning of the term makes us forget that the term has to do with equality and justice. Who is in charge of deciding what are the equality and justice principles that should correct the bias of an AI system? As suggested by Geburu and Mitchell [Bender et al., 2021], this is a political question and not a technical one.
2. Kate Crawford in *Atlas of AI* [Crawford, 2021] argues that we cannot save ourselves from the bias problem by improving algorithms only, because the AI systems we are talking about essentially make classification. Classification is an act of power: classifications are created to justify the order, the status quo of who is in power. The people who classify can decide which differences make a difference. They are political interventions hidden to us and which are taken for granted, but they still play an active role and shape the social and material world. To contrast them we must actively resist.
3. AI in image recognition is going to create a so-called “digital epidermalization”. The only data the machine can have, our images, do not contain the information machines tend to take out of images: gender, cultural and political views. These categories are not biological traits that emerge from our face images, but are politically, culturally, and socially constructed relational concepts. Leaving image recognition, the use of AI systems for inferring information may mean alienating our right to define our true identity: they do not consider what are our subjective claims.
4. Explainability: A large part of the machine learning community is trying to create systems that are not entirely black boxes, but that are able to justify the classifications they make. This concept is so widespread in the community that it even ended up in the General Data Protection Regulation of EU (GDPR) of 2016. In recital 71, it is said that people have the right to obtain an

¹ <https://inventory.algorithmwatch.org/>

explanation of the decision reached after assessment made by AI and to challenge the decision. Before rushing to say that achieving explainability of ML decisions would save AI, we need to understand what an explanation is. Do we really know? Can humans give explanations? Different disciplines have different points of view on the matter. Law is often advocated as an example, since judges motivate their judgements. However, Amedeo Santosuosso [Santosuosso, 2021] reminds us that in law, motivations of judgements are just a recent social construct. In some legal systems, writing the motivations can be even delegated to assistants. Moreover, as the economist Michael Polanyi [Polanyi, 1966] discovered, explicit knowledge (the part that can be expressed linguistically and can be represented in a logical way) is only a part of our knowledge. There is a bigger part of knowledge which is the “tacit knowledge”, concerning how to do things, for example how to ride a bicycle or how to be able to speak or even to reason. This knowledge can hardly be codified. Then there are technical problems. How to explore the dependencies among tens of trillions of nodes constituting a deep neural network to reconstruct its reasonings? Finally, as discussed in “*Everyday Chaos*” by David Weinberger [Weinberger, 2019], deep neural network systems could be literally unexplainable since, given their computational power, they are able to reason without resorting to the traditional conceptual models made by humans to decide and at the same time explain their decisions. A new venue for research is broached here which challenges the traditional picture of knowledge.

3 Projects

SIpEIA was established also within some projects such as CANP² and CH4I³ on AI and healthcare. SIpEIA participates to the project *AI Aware* funded by the public engagement call of University of Torino. *AI Aware*⁴ is a review of actions that will take place in 2022: events, serious videogames, podcasts and a graphic novel.

In its first year of activity SIpEIA organized several events for the larger public, an online PhD course and a tutorial at ACAI2021:

1. Foundational event “*Per una nuova etica della IA (Intelligenza Artificiale): le tecnologie rinnovano e cambiano l’umano*” 3/12/2020.
2. Webegno “*Medici: Primi attori o comprimari? Intelligenza Artificiale e Etica Medica*” 13/04/2021.
3. Webegno “*Intelligenza Artificiale ed Educazione*” 07/06/2021.

4 References

- Website of SIpEIA with the recordings of our events: <http://www.sipeia.it>
 Online PhD course on Ethics and AI in English organized by SIpEIA: <http://www.shorturl.at/zOSY0>
 Facebook page of SIpEIA: <https://www.facebook.com/Sipeia-100974895423092>
- [Beauchamp and Childress, 1979] Tom L. Beauchamp, and James F. Childress. *Principles of Biomedical Ethics* (5th ed.). New York: Oxford University Press, 1979.
- [Bender *et al.*, 2021] Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, Shmargaret Shmitchell. *On the Dangers of Stochastic Parrots. Can Language Models be too Big?* FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. pp. 610–623, 2021.
- [Benkamin, 2019] Ruha Benjamin. *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity Press, Cambridge, UK, 2019.
- [Crawford, 2021] Kate Crawford: *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*, Yale University Press 2021.
- [Floridi and Cowsls] Luciano Floridi and Josh Cowsls. *A Unified Framework of Five Principles for AI in Society*, Harvard Data Science Review, 1.1, 2019.
- [Habermas, 1989] Habermas, Jürgen, *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*, Thomas Burger, Cambridge Massachusetts: The MIT Press, p. 30, 1989. Translation from the original German, published 1962.
- [Harari, 2014] Yuval Noah Harari, *Sapiens, a brief history of humankind*. Random House, 2014.
- [Polanyi, 1966] Michael Polanyi. *The tacit dimension*. Garden City, N.Y., Doubleday, 1966.
- [Santosuosso, 2021] Amedeo Santosuosso. *Intelligenza artificiale e diritto. Perché le tecnologie di IA sono una grande opportunità per il diritto*. Mondadori, 2021.
- [Searle, 1995] John R. Searle, *The Construction of Social Reality*. Free Press, New York, 1995.
- [Weinberger, 2019] David Weinberger. *Everyday Chaos*. Harvard Business Review Press, 2019.
- [Wu, 2010] Tim Wu, *The Master Switch: The Rise and Fall of Information Empires*. Alfred A. Knopf, New York, 2010.

² <http://casanelparco-project.it>

³ <https://ch4i.di.unito.it>

⁴ <http://www.ai-aware.eu>