

## **Ethics, Artificial Intelligence and Responsibility: Contemporary Challenges**

Alvino Moser<sup>1</sup>  
Luís Fernando Lopes<sup>2</sup>  
Thereza Cristina de Souza Lima<sup>3</sup>

---

### **Abstract**

*The purpose of this paper is to approach, from a philosophical perspective, the relationship between responsibility and artificial intelligence, emphasizing some contemporary challenges of this relationship. To this end, we started with a brief exposition of the notion of responsibility from a philosophical perspective, highlighting its social aspect. We then analyzed responsibility in relation to objects developed by artificial intelligence (AI), pointing out that intelligent machines are programmed to think or imitate or replace human intelligence, in order to optimize problem solving and replace humans more efficiently in actions that are difficult or require much effort. Finally, we addressed some contemporary challenges concerning the ethics of artificial intelligence and responsibility. From the reflections presented, we consider that the responsibility for the actions of autonomous objects and their consequences lies upon their manufacturers, programmers, sellers, or users, since it is not possible to attribute personality to an automaton.*

**Key Words:** *ethics, responsibility, artificial intelligence*

---

Date of Submission: 21-05-2021

Date of acceptance: 06-06-2021

---

### **I. Introduction**

For some time now, artificial intelligence has become a reality in our daily lives. To perform simple activities, from scheduling and remembering appointments to more complex operations, such as helping to control a vehicle, we count on the increasing collaboration of digital technologies. Cell phones and other digital devices are already built, developed and programmed by AI. We particularly emphasize autonomous objects.

Notwithstanding all the great benefits that come with the use of these contemporary resources, it must be recognized that accidents and other problems are also part of this reality. But can automated objects be held responsible for the consequences of the decisions they are programmed to take?

Considering such point of view, this article aims to approach, from a philosophical perspective, the issue of responsibility in relation to artificial intelligence, emphasizing some contemporary challenges resulting from this relationship. To do so, we started with a brief presentation of the notion of responsibility from a philosophical perspective and highlighted its social aspect. We then analyzed responsibility in relation to objects developed by AI, pointing out that intelligent machines are programmed to think or imitate human intelligence in order to optimize problem solving and replace humans more efficiently in actions that may be too difficult or require too much effort from human beings. Finally, we consider some contemporary challenges concerning the ethics of artificial intelligence and responsibility, pointing out that the responsibility for the actions of autonomous objects and their consequences lies with their manufacturers, programmers, sellers, or users, since it is not possible to attribute personality to an automaton.

We hope that the reflections presented here as well as the results achieved may contribute to the dialogue and the expansion of research in the field of ethical responsibility related to the development and use of digital technologies, with focus on improving the quality of life of human beings and the preservation of the planet and all its biodiversity.

---

<sup>1</sup> PHD in Philosophy from the Catholic University of Louvain. Dean of the International University Center Uninter. Professor of the Master's Program in Teacher Training and New Technologies. Email: [moseral.am@gmail.com](mailto:moseral.am@gmail.com).

<sup>2</sup> Degree in Philosophy. Master and Doctor in Education from the Tuiuti University of Paraná. Professor in the Humanities Area at the UNINTER International University Center. E-mail: [luis.l@uninter.com](mailto:luis.l@uninter.com).

<sup>3</sup> Degree in Letters - Portuguese and English from Santa Úrsula University / RJ. Master and Doctorate in Linguistic Studies from Universidade Estadual Paulista Júlio de Mesquita Filho / UNESP / SP. PhD internship PDEE by the University of Manchester / UK. Director of the School of Languages at the UNINTER International University Center. E-mail: [thereza.l@uninter.com](mailto:thereza.l@uninter.com) and [teslcrisrina@gmail.com](mailto:teslcrisrina@gmail.com).

## **II. Ethics And Responsibility**

The notion of responsibility has gained importance in the ethical field, but it has often been mentioned in the political field as well. Responsibility is understood as the obligation to be accountable for the actions of the subject agents and the consequences that proceed from these conscious and voluntary actions.

Legally, to be responsible means to be accountable for one's actions and to be subject to sanctions as well as to be obliged to repair the damage of the consequences of those actions. One is not obliged to answer for involuntary actions, such as actions caused by reflexes, because according to Boethius Boécio (2005, p. 282): "*personae est definitio: 'naturae rationabilis indiuidua substantia'*" (it is called a person: an individual substance of a rational nature). Note that it is a person with rationality and therefore conscious and free, and as such must answer for his acts, that is, assume the consequences of his actions. Thus, being of an intellectual nature, it is only possible to demand from a person responsibility for conscious, free and voluntary acts and omissions.

In Greek, the word that comes closest to our notion of person is *prósopon*: "one who directs his look forward, who dares" (MOUNIER, 1961, p.470). "The person exposes himself, expresses himself: presents his face, is face" (MOUNIER, 2010, p. 65).

The Greek word *prósopon* has the original meaning of "theatrical mask"; this mask was mandatory for all characters on stage. The word is composed of two ideas: the face and what is in front of the face, the mask itself, for meaning and naming the act or the effect of the actor posing and representing, by the sound (per + sona) of his voice, a character (PAULA, 2010, p. 50).

For Paul Ricoeur (1968), Mounier, like no other, knew how to bring together, in the term person, the multidimensional sense that this word condenses in this concept:

But it seems to me that what attracted us to him is something more secret than a theme of many faces - the rare agreement between two shades of thought and life: the one which he himself called strength, in the wake of the ancient Christian moralists, or even the virtue of putting ourselves face to face - and the generosity or abundance of the heart, which corrects the crispness of the virtue of strength by something gracious and delicate; it is the subtle alliance of a beautiful "ethical" virtue with a beautiful "poetic" virtue that made Emmanuel Mounier this man, at the same time, irreducible and giving (RICOEUR, 1968, p. 165).

A person is morally responsible when he is free and autonomously responsible for what he has done and does, that is, for his actions,

Interpreting actions broadly, as to include both overt actions, such as walking to the store or shooting with a weapon, and mental actions, such as choosing to go to a law college or calculate a move in chess. In addition to being responsible for actions, a person can be responsible for the consequences of their actions. By putting salt in Sean's tea, Meghan is responsible not only for spoiling Sean's tea (what she did), but for his tea being ruined (the consequence of what she did). A politician's opponents may claim that he is responsible for an economic crisis, the result of his decisions and policies. Eventually, a person may be responsible for omissions, sometimes known as abstentions or omissions, and their consequences (ROBB, 2020).

Therefore, it is a duty to keep one's promises, to repair the damage caused to someone by one's actions or by those for whom one is responsible to monitor, or even to bear a penalty.

Another meaning was gradually added to this legal and moral meaning, namely, everyone would be responsible for other persons under his guardianship or his care, because of the vulnerability or dependence of the guardians. The idea of reparation is replaced by the idea of precaution. This irreproachable responsibility, after Levinas (1982), we will call "responsibility for others" (BERNARDI, 2019).

But, how can we talk about responsibility of autonomous objects or of intelligent machines?

## **III. Responsibility Regarding Objects Developed By Artificial Intelligence**

Intelligent machines are programmed to think or imitate human intelligence, with the intention of optimizing the resolution of problems and replacing the human being, in a more efficient way, in actions that are difficult or require a lot of effort. In short, they are developed to make your life easier and provide you with more comfort. Imagine yourself without an airplane or a car or bus, traveling long distances, such as to go from Brazil to Europe or another continent, and commuting in Brazil between cities, whether small or large distances. How much effort and time are diminished (BERNARDI, 2019).

The problem arises when intelligent machines and autonomous objects can decide independently. Although intelligent machines and autonomous objects are developed according to patterns and from given data, they have a certain autonomy when the type of programming involves neural networks, composed of several layers that contain hidden nodes (relationships). Moreover, although we are still responsible for the programming, the machines that operate as machine learning, as well as the neural networks, act through programming structures combined with the relationships and patterns they weave with the databases from which they draw the information to test and operate.

The expression machine learning can be understood as "machines that learn". Such machines are developed by artificial intelligence from the data that is fed into them and through the algorithms that guide their operations. The objects developed in this way, despite being programmed, can solve problems and learn, by themselves, new knowledge not foreseen by the developer. These objects act in a black box by means of relationship between the information embedded in it and can identify and create new patterns. They are self-regulated and self-programmed. This capacity gives them the ability to decide autonomously.

Pedro Domingos (2017, p. 27, our own bold and interpretation) states:

Sometimes, machine learning is confused with artificial intelligence (or AI for short). Technically, they constitute a subfield of AI, but they have grown so large and so successfully that they have overshadowed their proud mother. The goal of AI is to teach computers to do what humans currently do best, and learning is arguably the most important of these tasks; without it, no computer can match a human for long; with it, the rest comes as a consequence.

The programming of machine learning by AI, as well as of other objects so developed, makes the control of these machines more rigorous and enables the programmer greater authority over the results he or she wishes to obtain with the machines (BERNARDI, 2019). The control of AI-developed machines and objects limits, to some extent, human autonomy. However, there are also limitations regarding the autonomy of the machine, so that possible harmful consequences can be prevented.

Yet, notwithstanding these objects ability to learn, it cannot be said that they experience feelings and emotions and understand norms of morality, an exclusive capacity, in our view, of rational, conscious and free beings, different from machines. "What the programmer can do is embed limits based on ethical principles into the action of autonomous objects" (BERNARDI, 2019, p. 56). It seems impossible for autonomous objects to have "autonomous morphogenesis, invariance and teleonomy, characteristic of living beings" (MONOD, J. 1970)

But the question remains: Is it possible to teach morality to the machine, when not even all people can internalize it? Even human individuals do not always internalize it, especially if we consider the factors indicated by Freud: the Id, the unconscious, that is, the impulses from the deep regions of the brain inaccessible to consciousness, the paleocortex; the Superego, namely, the eco-socio-cultural conditioning. Id and Superego limit and condition the Ego, the consciousness.

The traditional description of unconsciousness as conceived by Freud is of historical significance and has not only gained wide acceptance, but has also attracted much criticism. However, it is now known that the fundamental mode of processing brain functions is of an unconscious order. Parts of the brain's symbolic-declarative processing and processing of emotional functions are permanently unconscious. Other parts of these processes are conscious or can be brought into consciousness or, alternatively, can be excluded from consciousness. Highlighting the essential role of unconscious psychic processes, Winograd cites, as an example, the verification that the behavior of patients unable to remember past events, because of lesions in brain structures responsible for memory storage, is clearly influenced by the "forgotten" facts. These data corroborate Freud's theory, according to which the unconscious dominates most of the psychophysiological processes (LIMA, 2010, p. 282)

In recent years, there has been a considerable increase in research and initiatives related to the ethical implications derived from the use of artificial intelligence in our daily lives. Surfing the internet, using social networks, apps, appliances, vehicles, shopping, banking, among others, we notice the increasing presence of AI in our lives.

But, and when we bring ethical issues to the table? What are the implications? In Brazil, for example, we are now putting into practice the application of the LGPD. But to what extent will this law be observed and citizens will have their data preserved, especially considering the arsenal of technological resources available today, capable of capturing data and offering faster and faster responses based on individuals' "digital behavior" history?

When considering automated objects, there are several implications related to the horizon of ethics. Although it is possible to discuss the rights of robots, the major problem concerns the relationship between human beings and intelligent machines, and the use of these intelligent machines and their resources by some human beings and others who run the risk of becoming disposable or even hostages of the control of these machines.

Ethical discussion requires the consideration of responsibility. Thus, transferring the responsibility for the consequences of the use of artificial intelligence to robots does not in itself seem to be an attitude that really considers ethics as a reflection on moral behavior. However, the more the level of complexity of automata, whose internal development is processed in a black box increases, the more complex also become the discussions about the ethical implications of the use of these objects. With the increasing rapprochement between AI and human intelligence, intelligent machines can make decisions that are not always programmed

and predictable. Thus, they seem to act on their own, as they consider the needs and circumstances of each situation beyond what was previously programmed.

Ever since human beings started using technological instruments, and especially machines, accidents have happened. But not for this reason has the responsibility for these accidents been placed on the instruments or machines themselves. Even when natural catastrophes occur, it is necessary to consider the fact that human beings have made and are making choices that can increase or reduce the impact of the consequences. When building a dam, for example, even if all legal and regulatory requirements are met, the risks of a breach and the possible disastrous consequences cannot be ignored. In this sense, the movement of trying to transfer the responsibility for the damage caused by the use of intelligent machines to these very machines seems to want to exempt the initial agent, the creator of the artifact, the one who ordered the construction and made the technical plans for the dam, from responsibility for the product developed and made available to society.

Even if one cannot know how the reasoning of algorithms works, which occurs, as already mentioned, in a black box, one must recognize that the development and use of automata is the responsibility of their creators and those who agree to use them. More and more, automata are approaching human intelligence, to the point that some works of fiction may already have been surpassed by what reality has provided.

Notwithstanding all the benefits provided by the use of AI and automata objects, we need to keep in mind the risks of accidents and their consequences. In this regard, a striking example was the fatal accident that occurred to Joshua Brown in March 2018 in Tempe, Arizona USA. She was hit by an automaton vehicle manufactured by Tesla, owned by UBER. Although there was a person inside, the vehicle was in autonomous mode, without a driver, at the time of the accident. According to Cano (2018), this was the first fatal car accident involving an automated car. However, the first known death of a passenger in a car in autopilot mode occurred in July 2016, with a Model S by Tesla that crashed into a truck in Florida USA. At the time of the accident the driver was watching a movie.

Recently, on April 17, 2021, an accident involving a 2019 Model S caused the death of two people in Texas. Since neither occupant was in the driver's seat, local authorities believe that the car was being driven by Autopilot. Tesla CEO Elon Musk stated, via Twitter, that this is not quite true, as the data recovered so far have indicated that the autopilot was not on. According to Elon Musk, such a vehicle does not drive autonomously on roads that do not have lanes, as was the case on the road where the accident occurred. Among other issues, investigations are seeking to discover whether there was a failure of sensors or even if the autonomous driving system was being used in inappropriate conditions.

Although it is still necessary to wait for the results of the investigations, these accidents cause at least some distrust among users and potential customers<sup>4</sup>. Tesla has already been criticized for using terminology such as Autopilot and Full Self Driving. According to specialists, these expressions may lead users to understand that it is a matter of a fully autonomous operation of the system, when what we have, in fact, are driver assistance mechanisms (ALECRIM, 2021).

As can be seen, these accidents reveal short-term concerns regarding the use of automata and their consequences. Even if the goal is to protect and offer greater safety and convenience to humans, the use of automata can cause accidents, including fatalities. With regard to long-term concerns, Eric Horvitz (2017), director of research at Microsoft, commented on the possibility that one day we will lose control over artificial intelligence systems due to the rise of superintelligence. Once these powerful systems do not act according to human wishes, they could even endanger humanity itself. But Horvitz himself questions whether such dystopian consequences would be possible and what to do about such reality. According to Horvitz, more investment is needed in research in this area, since the tools already available for exploring artificial intelligence, such as, reinforcement learning and other utilitarian functions, are not adequate to address the problem of superintelligence control and offer a robust solution.

Since it is not yet possible to know how algorithmic reasoning, elaborated in a black box and whose effect depends on deep learning, (artificial neural networks) works, the use of automata is not free from criticism and raises several ethical questions.

From this perspective, a topic that has caused debate around the world is the possibility of creating electronic personalities. However, it should be noted that the responsibility of autonomous robots lies with the programmers, the vendors, or with those who command them, the users. The European parliament has even discussed the proposal to treat autonomous objects as electronic personalities and suggested that they should have insurance to cover the consequences of their use.

As it is not possible to treat autonomous objects as human beings, the proposal put forward by Mady Delvaux-Stehres from the European Parliament to endow them with specific rights and duties, intends to consider them "electronic personalities." The basis for this proposal to give robots an electronic personality starts from the mistaken consideration that liability for damage caused would be impossible to prove. Despite

---

<sup>4</sup> It is worth noting that, currently, in addition to this one, the U.S. National Highway Traffic Safety Administration (NHTSA) is investigating 27 other accidents involving Tesla cars (ALECRIM, 2021).

this misconception, we must emphasize the concern of the proponents to offer legal security to companies that operate in this market segment. Likewise, there is a concern to offer security to customers and to all people who, directly or indirectly, are affected by the use of autonomous objects. Added to this concern is the fact of establishing a unified legislation in order not to be held dependent on criteria established by others. Diversified legislation is a major obstacle to the globalized market and can even hinder the very use of automata by their owners.

Meanwhile, the director of NBIC Ethics (Nanotechnology, Bioethics, Computer technology and Cognitive technology), Laetitia Pouliquen, who was one of the mentors of the Open Letter to the European Commission on Robotics and AI, denounced the resolution of the European Parliament Regulation on Robotics Law adopted in February 2017, proposing the creation of a specific legislation for autonomous objects. Laetitia Pouliquen criticized the composition of the AI Ethical Guideline Group for being mostly formed by industry representatives, with almost no presence of philosophers, sociologists, theologians, anthropologists or health professionals. This way, Pouliquen considered that the discussions might be reduced to the economic aspects of cost-benefit, without taking into account the whole body of human knowledge, which would enable the dependence on algorithmic decisions.

If the goal is to protect humans from their own inventions, it is necessary to broaden the field of analysis and consider the knowledge produced in different fields, because somehow all sciences are human and seek to provide results that are beneficial to humans. From this perspective, Artificial Intelligence (AI) is not an end in itself, but a means that needs to be oriented by well-defined goals.

According to experts in the field, the six fundamental ethical principles of AI, are benevolence (do good), non-maleficence (do no harm), human autonomy, justice (i.e., non-discrimination of AI), and applicability to ensure autonomy, informed consent, and data protection."

Caroline Burle and Diogo Cortiz (2020), who performed a multisectoral and non-exhaustive mapping of Artificial Intelligence principles<sup>5</sup>, based on six dimensions: 1) Fairness; 2) Reliability & Safety; 3) Social Impact; 4) Accountability; 5) Privacy & Security; and 6) Transparency, have verified that:

[...] three principles: Equity, Reliability & Security, and Accountability are addressed by the six mapped initiatives. The principles of Social Impact and Transparency are made explicit by five initiatives. And only the Privacy & Security principle is detailed by four of the six mapped initiatives (BURLE AND CORTIZ, 2020, p. 14).

Burle and Cortiz, (2020) also pointed out that, in general, all the mapped initiatives mention the Accountability principle, considering the need to "ensure the accountability of AI systems and their results, before and after their development, implementation, and use". Regarding this principle, the Beijing Academy of Artificial Intelligence additionally states that: "AI researchers and developers need to have sufficient consideration for the possible ethical, legal and social impacts and risks brought about by their products and take concrete actions to reduce and avoid them" (BURLE and CORTIZ, 2020, p. 13).

As can be seen, the concern with the impacts and legal and social risks arising from the development and use of autonomous objects is latent, just as the need to plan actions, so that they can be avoided or at least reduced. Thus, we emphasize the importance of studies and research dedicated to this issue considering multiple viewpoints. In face of this necessity, we believe it is pertinent to make considerations about some ethical perspectives.

#### **IV. Ethics For Artificial Intelligence**

Based on what we have presented so far, it is possible to see that current discussions on the development and use of AI that consider the ethical dimension present a tendency to be guided by utilitarian or consequentialist principles. In short, notwithstanding the problems that can be raised regarding this perspective, utilitarian ethics has, as its guiding principle, the purpose of providing the greatest good to the greatest number of people.

Utilitarianism was founded by the Englishman Jeremy Bentham (1748-1832), whose principle is the greatest happiness for the greatest number of people. In other words, "an action would be moral if it resulted in the greatest benefit for the greatest number of people. In turn, John Stuart Mill (1806-1873) in his work *Utilitarianism*, assuming this moral doctrine, emphasizes its altruistic character. It is a typically Anglophone doctrine, writes Jacqueline Russ (1994).

---

<sup>5</sup> The authors mapped six international initiatives, two from the government sector (European Commission and US Department of Defense), two from the business sector (Google and Microsoft), one international organization (Organization for Economic Cooperation and Development - OECD), and one composed of academia and business sector (Beijing Academy of Artificial Intelligence) (BURLE and CORTIZ, 2020, p. 5).

John Stuart Mill (2000, p. 49) states:

The utility or greatest happiness principle, as the foundation of morality, argues that actions are right insofar as they tend to promote happiness and wrong when they tend to produce the opposite of happiness. Pleasure and absence of pain is understood by happiness, by unhappiness, pain and deprivation of pleasure (MILL, 2000, p.49).

For others, utilitarianism would be equivalent to consequentialism. Thus, for a theory to be considered consequentialist, it must accept that certain normative properties depend exclusively on their consequences. However, philosophers such as McNaughton and Rawling (1991), Howard-Snyder (1994), and Pettit (1997) claim that a moral theory should not be classified as consequentialist unless it is neutral on the part of its agents, which practically makes the practical application of this proposal impossible.

Among the critics of utilitarianism or consequentialism is John Rawls (1997), because individuals, by themselves, seek to satisfy their needs and well-being. Now, how is it possible that the sum of all individuals in society have the same interests and consider happiness as something common to all? How can society achieve maximum satisfaction if each individual, each subject is a single being? What guarantee is there that all members of society will be reached by this calculation of the sum of all isolated subjects? It is a utopia. There is no "universal I", because what exist is "empirical I"<sup>6</sup> with its culture and its determinations. Therefore, consequentialism is controversial. According to various non consequentialist views, morality is concerned with doing one's duty, respecting one's rights, obeying nature, obeying God, obeying one's heart, updating one's potential, being reasonable, respecting all people, or not interfering with others - no matter the consequences. How can a single man determine the greatest good or the greatest satisfaction for all members of society? "Utilitarianism does not take seriously the difference between people" (RAWLS, 2000, p. 30).

Susan Leigh Anderson, in her article: GenEth: a general ethical dilemma analyzer, takes as her perspective utilitarian ethics or consequentialism, which goes against Kant's and Rawls' deontological ethics, intuitionistic ethics, Aristotelian-Thomistic ethics, to name a few. She proposes "embedding in autonomous systems", with the basis "GenEth uses inductive logic programming (ILP)", that is, reasoning from examples, which is problematic, to say the least, for ethicists, because "utilitarianism is an empiricism that starts from facts, [...] considers the tendencies and inclinations of men and then strives to satisfy them (RUSS, 1994, p. 87)".

In his work *A Theory of Justice*, John Rawls (2000) states that legislators should vote enveloped in a veil of ignorance, that is, those who dictate the laws, those who legislate, should put themselves in the position of not considering whether the law will favor their interests or harm them, regarding the good of all and not of majorities, whatever they may be.

Afterwards, Rawls (2000) proposes two principles: the first, that of freedom: he considers that all people have the same demands for basic freedoms. The second, that of equality: he understands that social and economic inequalities should be ordered in such a way that they are at the same time considered to be advantageous to all within the limits of what is reasonable (principle of difference), and linked to positions that are accessible to all (principle of equal opportunity).

In the first principle, we understand the rights to life, to subsistence, to work, to education, to leisure, the civil, social, and political rights. It is the principle of equality. The distribution of rights, duties, and other social goods. They can be applied (at different stages) to the judgment of the political constitution, the ordinary laws and the decisions of the courts. The distribution of these rights and the redistribution of income would be determined with the hypothetical condition pointed out, namely, "the veil of ignorance" (RAWLS, (2000, p. 146-153).

Thus, with justice as equity, Rawls intends to speak of a reasonable notion of justice, which allows mediating political coexistence through the contract (making mutual agreements between people in equal conditions). In justice as equity, the concept of right comes before the concept of good. However, Rawls' theory is still a utopian attempt based on liberalism.

Liberalism does not reign in all nations, and even if one were to accept Rawls' principles that aim to "perfect" consequentialism, the ethical challenge remains: What will happen to current jobs and employees, since the unemployed and under-employed exceed millions worldwide (MOSER, 2019).

Justice, according to Aristotle (1994) in the *Nicomachean Ethics*, is to give to each one what he deserves, to give to each one what he is entitled to, but there are cases in which the people to whom justice is applied are not in the same condition. For example, a single worker and a worker who needs to support his wife and two children. Justice is blind to these differences, and equity does not deny justice, but seeks to correct it. Also, according to Aristotle (*Nicomachean Ethics*, book V, 1994, p. 49-236), the measure of equity is similar to the ruler of Lesbos, a famous instrument used by architects to achieve better results, because, being flexible, it adapts better to the imperfections and hardness of the rock. The measure of the law, on the other hand, is similar

---

<sup>6</sup> There is no plural of I. The I is unique and irreplaceable. We is you and me.

to the inflexible linear ruler. Thus, for Aristotle (1994,) a correction of the law must be made when it is deficient because of its universality. Some of these principles are even taken from the Hippocratic oath.

Laetitia Pouliquen (2019, s/p, our translation) questions:

But then, without autonomy, would the user have no rights? Likewise, will the transparency of algorithms, the non-discrimination, data protection be sufficient principles to guarantee the respect of our freedoms? Probably not.

Well, considering man as a consumer and user, while trying not to harm him, is depriving him of his autonomy and freedom. Autonomy, according to philosophy, is the ability to determine oneself, the ability to choose and to impose rules on oneself, that is, to legislate for oneself and for the principles one adopts for one's conduct of life. In this sense, the President of the European Commission concludes that it is inappropriate to apply the term "autonomy" to mere artifacts, even if they are sophisticated, indeed "intelligent," advanced systems.

Our aim is not to deny the need for an AI Ethical Guideline (Ethical Guide to Artificial Intelligence). We dare, however, to suggest that the Guideline be established following the communication ethics proposed, among other theorists, by Habermas (1988), Rorty (1995), and Karl Otto Apel (2000).

This is an ethics of discussion or argumentation. It presupposes Universal Pragmatics, that is, the use of language that is common to all in the same universe. Thus, all those who engage in a practice of argumentation need to presuppose pragmatically that, in principle, all those possibly affected can participate, as free and equal, in a cooperative search for truth, in which the only coercion admitted is that of the best argument.

In the same way, it is required that everyone who participates in the discussion, in conditions of equality, be committed to accepting the conditions of the agreement. In this perspective, Gilbert Hottois (1990), quoting Richard Rorty (1995), affirms that there is nothing, neither metaphysical, nor rational, nor any other instance<sup>7</sup> above the consensus among those who discuss.

The ethics of communication aims to achieve agreement and consensus, without any kind of coercion, be it moral, psychological or, above all, physical coercion. It is based on the assumption, already stated, of the freedom and autonomy of the subjects involved. It cannot be forgotten that this also applies to apps, Uber, 99, Airbnb, Rippa, among others, in which few are the ones who keep the profits, while most of them enter with the work and must bear the costs and risks.

## V. Concluding Remarks

This paper has sought to analyze the relationship between responsibility and artificial intelligence, emphasizing some contemporary challenges of this relationship. To do so, we started with a brief exposition of the notion of responsibility, from a philosophical perspective, highlighting its social aspect.

We, then, analyzed responsibility in relation to objects developed by artificial intelligence (AI), pointing out that intelligent machines are programmed to think or imitate human intelligence, in order to optimize problem solving and replace humans more efficiently in actions that are difficult or require much effort.

Finally, we addressed some contemporary challenges concerning the ethics of artificial intelligence and responsibility. Based on the analyses carried out, we corroborated the thesis that responsibility for the actions of autonomous objects and their consequences lies upon their manufacturers, programmers, vendors, or users, since it is not possible to attribute personality to an automaton.

Thus, in questions concerning Artificial Intelligence and the resources made possible by it, one should not only listen to those who produce it, or make it available and profit from it, but also to the users, who are of diverse and variable tendencies. The ethics of discussion imposes itself democratically. And these assumptions have incidences and implications for politicians and rulers, who, as representatives should not be limited to experts or their supporters, but need to meet the entire community.

Only by this way will the well-being of living together be achieved. In order to have a better world, it is necessary to "give up one's own certainties and seek objectivity", writes Humberto Maturana in the book "The Tree of Knowledge". And he stresses, in the same work, the need for discussion. Ethics is not only necessary and indispensable, but urgent.

## References

- [1]. APEL, K.-O. **Transformation of Philosophy 2: the a priori of the communication community**. 1. ed. São Paulo: Loyola, 2000.
- [2]. APEL, K.-O. **Discourse and responsibility**. The problem of the transition to post-conventional morality. Frankfurt am Main: Suhrkamp, 1988.

---

<sup>7</sup> To understand what transcendentalism or empiricism is in ethics or morality see: WILSON, Edward Osborne. The unity and of knowledge, Consilience. Is science capable of explaining everything? Rio de Janeiro: Campus, 1999, pp. 229-256.

- [3]. APEL, Karl Otto. **The ethics of discourse as ethics of political responsibility in the current situation of the world.** Magazine of the Faculty of Education. São Paulo, Vol. 18, number: 1, pp. 113-121, Jan. 1992.
- [4]. APEL, Karl Otto. **Modern moral studies.** Petrópolis - RJ: Vozes, 1994. APEL, Karl Otto. **Ethics and Responsibility: The problem of the transition to post-conventional morality.** Lisbon: Piaget, 2007.
- [5]. ARISTOTLE. **Nicomachean Ethics.** In: ARISTOTLE. *Metaphysics, Nicomachean Ethics, Poetics.* São Paulo: Editora Abril, 1994 (Thinkers Collection).
- [6]. ALECRIM, E. **Elon Musk reacts to criticism for fatal accident with Tesla without a driver.** Available at: <<https://tecnoblog.net/434056/elon-musk-reage-criticas-acidente-fatal-carro-tesla/>> Accessed on: April 2, 2021.
- [7]. BERNARDI, M. R. **The indiscriminate use of artificial intelligence: an ethical discussion / Mariana Rocha Bernardi.** - 2019. Dissertation (Master's Degree) - University of Caxias do Sul, Graduate Program in Philosophy, 2019. <<https://repositorio.ucs.br/11338/4856>> Accessed on: April 24, 2021.
- [8]. BOÉCIO. **Writings (Opuscula Sacra).** Translation, introduction, introductory studies, and notes Juvenal Savian Filho. Foreword by Marilena Chauí. São Paulo: Martins Fontes, 2005.
- [9]. BURLE, C.; CORTIZ, D. **Mapping of artificial intelligence principles.** [authors' translation]. São Paulo: Information and Coordination Center of Ponto BR, 2020. Available at: <[https://cweb.br/media/docs/publicacoes/17/20200721143359/digital\\_mapeamento\\_principios\\_IA\\_portugues.pdf](https://cweb.br/media/docs/publicacoes/17/20200721143359/digital_mapeamento_principios_IA_portugues.pdf)> Accessed on: April 24, 2021.
- [10]. CANO, R. G. **Uber driverless car causes first fatal accident.** New York, EL PAÍS, March 14, 2018. Available at: <[https://brasil.elpais.com/brasil/2018/03/19/tecnologia/1521479089\\_032894.html](https://brasil.elpais.com/brasil/2018/03/19/tecnologia/1521479089_032894.html)> Accessed on: March 23, 2021.
- [11]. DOMINGOS, P. **The master algorithm.** São Paulo: Novatec Editora, 2017.
- [12]. HABERMAS, Jürgen. **Theory of Communicative Action.** v. I and II. Madrid: Taurus, 1988.
- [13]. HABERMAS, Jürgen. **From The Ethics of Discussion.** German translation by Mark Hunyadi. Paris: Les Éditions Du CERF, 1992.
- [14]. HOWARD-SNYDER, F. **"The Heart of Consequentialism.** Philosophical Studies, number: 76, pp. 07–29, 1994.
- [15]. HOTTOIS, G. **The bioetiche paradigma.** Brussels: De Buek Wesmael, 1990.
- [16]. LEVINAS, E. **Ethics and infinity.** Paris: Fayard, 1982.
- [17]. LEVINAS, E. **Totality and infinity.** Paris: Pocket Book, 2001.
- [18]. LIMA, A. P. de. **Freud's structural model and the brain: an integration proposal between psychoanalysis and neurophysiology.** Rev. psiquiatr. cl. vol.37 no.6 São Paulo 2010. Available at: <[https://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0101-60832010000600005](https://www.scielo.br/scielo.php?script=sci_arttext&pid=S0101-60832010000600005)>. Accessed on: April 24, 2021.
- [19]. MCNAUGHTON, D., and RAWLING, P. **"Agent-Relativity and the Doing-Happening Distinction"**. Philosophical Studies, number: 63, pp. 85-167, 1991.
- [20]. MILL, J. S. **Utilitarianism.** São Paulo: Illuminations, 2000.
- [21]. MONOD, Jacques. **Chance and Necessity.** Petrópolis, RJ.: Vozes, 1970.
- [22]. MOSER, A. **Ethics and Artificial Intelligence.** Available at: <<http://professorfrontino.com.br/wp/2019/06/25/etica-e-inteligencia-artificial/>> Accessed on: April 24, 2021.
- [23]. MOUNIER, E. **Personalism.** Lisbon: Text & Graphic Editions, 2010.
- [24]. MOUNIER, E. **Works - volume I-IV.** Ed. Paulette Mounier. Paris: Ed. Du Seuil, 1963.
- [25]. PAULA, R. A. de. **Crisis of the person and the crisis of education: a study in the personalist perspective of Emmanuel MOUNIER.** Goiânia: UFG, 2010. [Doctoral thesis].
- [26]. PETTIT, P.; SLOTE, M. **"The Consequentialist Perspective"** in Three Methods of Ethics, by M. Baron. Oxford: Blackwell, 1997.
- [27]. POULIQUEN, L. **Can Artificial Intelligence be ethical?** Availability em: <<https://www.lefigaro.fr/vox/societe/2019/02/21/31003-20190221ARTFIG00141-l-intelligence-artificielle-peut-elle-etre-ethique.php>> Accessed on: January 24, 2021.
- [28]. RAWLS, J. **A Theory of Justice.** 2ed. São Paulo: Martins Fontes, 2000.
- [29]. RICOEUR, P. **History and truth.** Rio de Janeiro: Forensics, 1968.
- [30]. ROOB, D. **Moral Responsibility, and the Principle of Alternative Possibilities.** In: Stanford Encyclopedia of Philosophy, 2020. Available in: <<https://plato.stanford.edu/entries/alternative-possibilities/#WhatMoraResp>> Accessed on: April 29, 2021.
- [31]. RORTY, Richard. **Science and Solidarity.** Paris: L'Éclat, 1990.
- [32]. ROERY, Richard. **Hope instead of knowledge.** Capítitulo III: Ethics without moral obligations. Paris: Albin Michel, 1995.
- [33]. RUSS, J. **Contemporary ethical thought.** Paris: PUF, 1994.
- [34]. RUSSELL, Stuart; HORVITZ, Eric and TEGMARK, Max. **On Science Friday: Is AI Safety a Concern?** April 11, 2015/by Jesse Galef. Available in:<< <https://futureoflife.org/2015/04/11/russell-horvitz-and-tegmark-on-science-friday-is-ai-safety-a-concern/> >> Accessed on: February 05, 2021.
- [35]. TUAL, Morgane. **Beyond fantasies, what are the concrete problems posed by artificial intelligence.** Le Monde, August 04, 2017. Available in:<[lemonde.fr/pixels/article/2017/08/03/au-dela-des-fantasmes-quels-sont-les-problemes-concrets-que-pose-l-intelligence-artificielle\\_5168330\\_4408996.html](https://lemonde.fr/pixels/article/2017/08/03/au-dela-des-fantasmes-quels-sont-les-problemes-concrets-que-pose-l-intelligence-artificielle_5168330_4408996.html)> Accessed on: April 02, 2021.
- [36]. WILSON, Edward Osborne. **Unity and knowledge, Consilience.** Would science be able to explain everything? Rio de Janeiro: Campus, 1999.

Alvino Mose, et. al. "Ethics, Artificial Intelligence and Responsibility: Contemporary Challenges." *IOSR Journal of Research & Method in Education (IOSR-JRME)*, 11(3), (2021): pp. 01-08.