Gregorianum
Roma 2020 - 101/3

LOUIS CARUANA, S.I.

Electronic Persons?

Pontificia Università Gregoriana
Pontificio Istituto Biblico

Gregorianum 101, 3 (2020) 593-614
Louis Caruana, S.I.

Electronic Persons?

On 29 June 2015, a twenty-two year old worker at the Frankfurt branch of a Volkswagen factory was killed by a robot[1]. This is not the only case of a person being seriously injured or killed by a robot. Who is responsible when such things happen? Given that nowadays robots have decision-making capabilities dependent on pattern recognition and prior experience, some people are convinced that in such incidents the one we should hold responsible is the robot. Does this make sense? Is it just a move to facilitate legal procedures? A few months after that Frankfurt incident, and probably partly as a consequence, the Committee on Legal Affairs for the European Parliament produced a Draft Report, dated 31 May 2016, in which we find the proposal that robots be recognized as legal persons. In its exact wording, the proposal was to create «a specific legal status for robots so that at least the most sophisticated autonomous robots could be established as having the status of electronic persons with specific rights and obligations»[2]. On 16 February 2017, the European Parliament adopted this proposal as a Resolution on Civil Law Rules of Robotics. The resolution represents a new extension of the notion of legal personhood and, if approved as law, would probably have very significant consequences for the future. The entire issue deals not only with the level of machine sophistication but also with our own self-understanding. The questions it raises branch out in various directions, involving neuroscience, cognitive science, philosophy of mind, ethics, legal theory, and even theological anthropology.

In recent decades, research in artificial intelligence and robotics has had, in fact, a significant impact on philosophy, social sciences, legal theory, and moral theology. Researchers have studied the implications from various angles. The Vatican as well has contributed to these studies especially through two high profile workshops organized jointly by the Pontifical Academy of

[1] «Roboter tötet Arbeiter bei VW in Baunatal», *Frankfurter Allegemeine*, 01.07.2015.

[2] Committee on Legal Affairs for the European Parliament, Draft Report (Mady Delvaux, Rapporteur), p. 12 (Retrieved from <https://www.europarl.europa.eu/doceo/document/JURI-PR-582443_EN.pdf>.

DOI: 10.32060/Gregorianum.101/3.2020.593-614

Sciences and the Pontifical Academy of Social Sciences. The first one was on the Power and Limits of Artificial Intelligence (30 November–1 December 2016) and the second on Robotics, AI and Humanity, Science, Ethics and Policy (16-17 May 2019)[3]. In all these studies, a number of philosophers and legal scholars have focused on the specific question of attributing legal personhood to sophisticated machines but one area remains somewhat ignored, namely the semantic dimension of the problem. This paper will attempt to make a contribution in this area. The general aim is to examine this complex problem by determining what is happening at the level of meaning and of logical consistency within our understanding. The paper has three parts. The first presents the relevant historical background, with particular reference to philosophy of mind and cognitive neuroscience. The second part reviews the current debate as regards the applicability of legal status to intelligent machines. The third section then pushes the debate a little forward by exploring the neglected area of semantics. Overall, the main question will be the following. What is happening at the level of meaning when we try to attribute legal personhood to an intelligent machine, and what insights can we gain by studying the issue at this semantic level?

At the very start however, it is useful to clarify the distinction between robotics and artificial intelligence. In this paper, robotics will be taken to refer to the area of engineering that deals with the construction of finely controlled structures in view of performing some desired physical task, such as metallic grasping mechanisms, arms, or mobility devices. In this sense, what robotics produces, namely robots, are artificial instruments at the service of humans, devices that sometimes enjoy some autonomous capacity for social interaction among humans[4]. Artificial intelligence will be taken to refer to the branch of cognitive science that deals with the construction of machines that simulate not human physical abilities but mental capacities like calculation, speech-recognition, learning and problem solving. To refer to the collaborative product of these two disciplines, the paper will use the expression intelligent machines.

---

[3] See A.M. Battro – S. Dehaene, ed., *Power and limits of artificial intelligence: The proceedings of the Workshop on* «*Power and limits of artificial intelligence*», Città del Vaticano 2017. For the final statement of the 2019 workshop, see <http://www.pas.va/content/accademia/en/events/2019/robotics.html>.

[4] The social interaction of robots with humans is very important. It has an impact on how we describe them. Apart from taking them to be servants, we can take them to be companions, objects of entertainment, therapeutic instruments for old people or autistic children, and so on. This interaction determines also the extent to which we attribute emotions to robots. For a philosophical assessment of this point, see P. Dumouchel L. Damiano, *Living with Robots*, tr. M. DeBevoise, Cambridge Mass. 2017.

I. Background ideas

The relation between the mind and the body has been the object of philosophical and religious speculation for centuries, probably since the dawn of history. One might conjecture that very early humans started to assume the existence of immaterial souls because of the experience of dreaming about their deceased loved ones. Plato, in his book *Phaedo*, famously defended a clear distinction between soul and body, considering the latter a kind of prison from which the former seeks liberation. The body is corruptible but the soul, since it grasps mathematical and other kinds of necessary truths, is immortal. His student Aristotle adopted a different method and drew inspiration from biology. He saw the relation between soul and body as a special case of the more general relation between form and matter, between mover and moved. He recognized from the start that the question of the soul's location within the body is misguided. Soul, being the form of the natural body or «the principle of animal life»[5] has no specific location within the body. In this sense, form is similar to shape. It makes no sense to ask, «Where is the shape of your face located?» The shape of the face is the face. To have a soul is not to possess something or to be related to something but to be, to exist, in a specific way. Notice therefore how Aristotle's analysis of the soul was primarily conceptual, not empirical. For him, the soul was certainly not some material constituent of the person[6].

A major turning point occurred in the seventeenth century through of the work of René Descartes who proposed that, as regards explanation, we need to assume that non-human animals are machines. Humans on the contrary are made up of two kinds of substance: the extended type, the machine, and the thinking type, the mind. For Descartes, humans are completely conscious of the contents of their minds and have infallible access to this content via introspection. In spite of these clear distinctions, Descartes saw that naïve dualism could not be the right answer as regards humans. Human reality is more complicated. He wrote, for instance: «nature […] teaches me that […] I am not only lodged in my body as a pilot in his vessel, but that, apart from

---

[5] Aristotle, *De Anima*, tr. J.A. Smith, 402a7-8, in R. McKeon, ed., *The Basic Works of Aristotle*, New York 1941, 535.

[6] Although «soul» and «mind» are nowadays often used interchangeably, it is important to recall the Aristotelian-Thomistic important distinctions. These can be summarized briefly as follows. The «soul» refers to the principle of autonomous motion of organisms, whether these organisms are human or not. In other words, «soul» refers to the principle of movement, «body» to what moves. Saying that there is a variety of living things is the same as saying that there are many kinds of soul. One kind is the human soul, called a rational soul. Two main faculties characterize the rational soul: the mind and the will. The mind seeks the truth; the will seeks the good.

this, I am so closely united and intermingled with it that I compose with it one whole»[7]. In spite of this warning however, his distinctions took root in academia and brought forth a form of science-inspired dualism that is still with us today.

Consider for instance the development of neuroscience from the eighteenth century onwards. One of Descartes's conjectures was that the crucial location where the non-material mind interacts with the body was the part of the brain called the pineal gland. Descartes's interest in this question represents a new research project, namely the project of determining those parts of the brain that are responsible for specific intellectual or physical activities. Eventually this project came to be called the theory of Cortical Localization. The first attempt to map the brain in this way could not, of course, resort to human vivisection. It started rather by assuming that the localization of mental function resulted in outward physical manifestations. Researchers like Franz Joseph Gall (1758-1828), the originator of the now obsolete discipline of phrenology, assumed that bumps on the skull correspond to specific mental capacities, some enhanced more than others in line with the character of the individual person. This idea of external manifestation was eventually disproved. Nevertheless the determination of cortical localization remained a rewarding research project. The fine structure of the cerebral cortex was eventually mapped and our knowledge of the function of cortical sites became increasingly specific, arriving even to the identification of those parts of the brain that are responsible for the movement of just one finger[8]. Advancing further in microscopic brain anatomy, neuroscientists discovered the structure and function of the specialized cell of the brain, the neuron. A very significant point here is that, as opposed to the higher-level localization of the brain, we find no localization at the neuron-level. In other words, we find no one-one correspondence between neuron and brain function. And certainly none between neuron and bodily function. We find rather collaboration of many neurons, or cell-assemblies, for any specific function. To explain brain anatomy at the cell-level therefore, we need a holistic approach.

Up to now, this quick historical overview has highlighted the path of inquiry regarding the study of the brain's role in human physical and mental functions. This is however not the only line of inquiry worth mentioning

---

[7] René Descartes, *Meditation* VI : «La nature m'enseigne […] que je ne suis pas seulement logé dans mon corps, ainsi qu'un pilote en son navire, mais outre cela, que je lui sois conjoint tres-étroittement & tellement confondu & meslé, que je compose comme un seul tout avec lui.», in C. Adam – P. Tannery, ed., *Œuvres de Descartes*, IX, Paris 1996, 64; my translation.

[8] For a good overview, see Z. Folzenlogen – D. Ormond, «A brief history of cortical functional localization and its relevance to neurosurgery», *Neurosurgical Focus* 47/3:E2 (2019). DOI: 10.3171/2019.6.FOCUS19326

here. There is another long line of inquiry regarding simulation. The main question in this second line of inquiry has been, «How can we simulate what humans do intellectually?» Simulation starts with the construction of simple instruments, for instance the construction of a spade that, in a sense, simulates and extends our capacity to dig with our hands. As regards thinking, the first steps were taken in ancient times with the construction of instruments that help calculation, like the abacus. Progress continued with the construction of sophisticated mechanical adding machines, like Blaise Pascal's mechanical calculator invented in the early seventeenth century. With the advent of electronics, the second half of the twentieth century saw the rapid development of digital computers. Current machines can simulate many human intellectual abilities. They can respond correctly to human speech, compete successfully in high-level strategic games like chess, operate cars autonomously, and so on.

Where are we today? The question regarding legal personhood attribution to machines emerges forcefully because intelligent machines have now become capable of simulating more and more human intellectual skills. Consider for instance two impressive and significant areas of current development in artificial intelligence.

The first one concerns expert systems. These are software-designs that simulate not just a normal human person of average intelligence but the expert. They simulate the expert's ability to offer a dependable judgement regarding a specific course of action to be taken. They thus simulate the person who is capable of making a valuable judgement because of his or her long experience. The machine can simulate this by referring to a vast amount of stored data. In a sense, like a human being, it can experience, understand and then judge. This artificial capability became possible primarily when researchers started to model the machine hardware on the brain's neuronal structure[9]. They called the new structure an artificial neural network. We can think of an artificial neural network as a number of points in space with connections between them. The points, or units of the network, are simple processors and are usually situated in a number of distinct layers. The connections between units of one layer with units of the next layer are extremely numerous and are not all of the same strength. Information passes from one layer to another, from processor to processor, but the itinerary of information as it passes through the entire network is not linear. It is not along a single line made up of connections from one node to another node to another node, and so on. On the contrary, information is spread out. It is distributed. It passes along various connection pathways involving many nodes at the same time.

---

[9] Warren McCulloch and Walter Pitts are the recognized pioneers in this area, especially with their paper, W. McCulloch W. Pitts «A Logical Calculus of Ideas Immanent in Nervous Activity», *Bulletin of Mathematical Biophysics* 5/4 (1943) 115–133. DOI:10.1007/BF02478259

This kind of parallel distributed programming has very interesting properties. For instance, the input layer and the output layer are connected to the outside world, precisely because the former receives the information and the latter delivers it. The layers in between, however, are not. Engineers who construct the network never know for sure what is happening within these hidden intermediate layers. Moreover, this kind of distributed programming can simulate high-level human intellectual abilities such as learning. We use the expression «machine learning» when the networks program themselves for some specific task. Engineers will give the network a training period during which they expose it to a sample of input-output pairs. For example, a network may be given the chance to learn how to recognize words as they are spoken by a person with a particular accent and then to write these words on a screen. In this case, the training period consists of the person reading a standard text into the system. The standard text constitutes a set of input-output pairs to standardize the network. The network then can work and expand its detecting ability on its own. It can «learn on its own». The significant point for us in this paper is that the engineers who build the network will never know exactly how its hidden layers become standardized for that specific task[10].

The second point that deserves our attention deals with artificial life. While artificial intelligence simulates our mental functions, artificial life simulates the entire evolutionary process by which various biological species have emerged. In the work of John von Neumann towards the mid-1900s, we already find the hypothesis that not only intelligence but also life itself could be abstracted, as it were, from organisms. Once abstracted, it could be realized elsewhere using artificial structures[11]. The first step is to define life within a computer system. We define it as a set of rules that determine how a particular code self-replicates. When the code is executed, in other words, when the instruction is followed, we can say that the artificial organism lives. The self-replication allows some random changes that, through repeated iterations, could produce new stable codes. These stable mutants could generate other stable mutants, and the process could branch off and continue to self-replicate in various unpredictable directions. Notice that we are here not simulating the functionality of just one organism. We are simulating the species itself, together with the possible transformations it could undergo in the course of time. We can likewise simulate the various kinds of animal social behavior, like the swarming intelligence of flocks of birds or the optimization behavior we see in ant colonies. As in the previous point, we see again how researchers are now building and exploring systems that, in a sense, are unpredictable. We could say that these artificial systems have a life of their own.

---

[10] For a historical overview, see R.M. Harnish, *Minds, Brains, Computers: An Historical Introduction to the Foundations of Cognitive Science*, Malden, Mass. – Oxford 2001, part III.

[11] John von Neumann developed his innovative views in a set of lectures that were eventually published in A.W. Burks, ed., *Theory of Self-Reproducing Automata,* Urbana – London 1966.

II. Machines and persons

Machine expertise and machine life are developments that have greatly encouraged people to describe machines by using specifically human personal attributes. Without any hesitation, we now describe a computer as remembering, thinking, understanding or deciding. Apparently, the fact that such verbs are attributable to humans, and only rarely to some animals, does not worry us. In our vocabulary, machines have now qualified, as it were, from mere things to being autonomous agents. This new status of the machine has important consequences as regards responsibility. I started this paper by recalling the sad incident of the worker who was killed by a robot. Had the man been killed accidentally by a falling branch during a storm, we would not hold anyone responsible. But because the killing was caused by an entity that enjoyed a certain degree of independence, we are tempted, or even obliged, to see the killing as caused by an agent. No one knows what was going on inside its circuitry, not even its creators. The machine therefore enjoyed a certain degree of privacy and autonomy[12]. This seems to indicate that it should be held responsible. We are entitled or even obliged to attribute to it not only human features like remembering, understanding and deciding but also a human status with respect to the law.

What is at stake here? The idea of a legal person is not new. It arose from the awareness that a combined group of people could in certain circumstances act in a way that is impossible for any of its members on his or her own[13]. Because of this, the law could recognize the group as a legal entity or a legal person. On this point, jurisdictions are not in full agreement, and the difference can be somewhat surprising. For instance the juridical system in India recognizes as legal persons not only groups of people or corporations but also some idols

---

[12] Autonomy in this context just means that the machine is capable of establishing its own criteria for choosing to follow one rule rather than another, and to proceed this way at various stages of its functioning. The machine's overall behavior thus becomes something that had not been included in its original programing. This idea of autonomy is very superficial when compared to views that are more in line with human moral experience. Immanuel Kant, for instance, called the rational will autonomous because it operates by responding to what it considers to be reasons. Consciousness is therefore essential. Humans are autonomous because they act under the idea of their own freedom, responding to laws that they themselves lay down for themselves. See especially I. Kant, *Groundwork of the metaphysic of morals*, tr. H.J. Paton, New York 1964.

[13] This description is from p. 133 of an early study: G.F. Deiser, «The Juristic Person I», *University of Pennsylvania Law Review and American Law Register* 57/3 (48 New Series) (1908) 131-142. Examples of well-documented recent studies include D. Fagundes, «Note, what we talk about when we talk about persons: The language of a legal fiction», *Harvard Law Review* 114/6 (2001) 1745-1768; N. Naffine, «Who Are Law's Persons? From Cheshire Cats to Responsible Subjects», *The Modern Law Review* 66/3 (2003) 346-367.

of Hinduism. The idea of idols as legal persons works well because there is a human person who has the idol in his or her charge, who is in law the idol's manager, and who attends to its interests[14]. Of course, a legal person is a legal fiction. We use it because it facilitates legal reasoning. We should not forget however that, as a fiction, it could function well in some areas and not so well in others. The insufficiency of the idea often emerges when things go wrong and we need to ask, «Who is responsible for the damage done?» For lack of space, I will focus on one important area only: the link between responsibility and autonomy.

In what sense could a machine be responsible before the law? Responsibility and autonomy go hand in hand. Consider for instance the case of robots or drones used in war. In the near future, highly sophisticated autonomous war-machines will probably be entrusted with decisions about target-identification and destruction. Increasing machine autonomy in warfare could push humans out of the picture completely. Technological progress is enhancing the speed of machine decision-making. When the autonomous weapon system's decision-making speed exceeds that of humans, the only way to defend oneself against it will be to oppose it with another autonomous weapon system. The fighting will be entirely «in the hands» of machines because the side that decides to retain human control would lose out[15]. We all agree that many humans are involved in the manufacture and the initiation of such intelligent machines. Nevertheless, if we accept that the machine enjoys a degree of autonomy, we cannot automatically hold these humans responsible for the machine's action. The higher the degree of machine-autonomy, the less justification we have for holding one of these people responsible.

Notice how we reason in the same way when dealing with children. We hold parents responsible for minor children, those who enjoy limited autonomy. When children grow up, they become autonomous, and hence the parents are not responsible any longer. The degree of autonomy therefore determines who is responsible, and this principle seems applicable to intelligent machines. Of

---

[14] The idol's manager is analogous to the manager of the estate of an infant heir. See S.M. Solaiman, «Legal personality of robots, corporations, idols and chimpanzees: a quest for legitimacy», *Artificial Intelligence and Law* 25/2 (2017) 155-179.

[15] See R. Sparrow, «Killer Robots», *Journal of Applied Philosophy* 24/1 (2007) 62-77. We may be tempted to say that, if fully autonomous intelligent machines will really pose a danger to humans, «the solution is not to create them in the first place» (quoting from p. 1261 of L.B. Solum, «Legal Personhood for Artificial Intelligences», *North Carolina Law Review* 70/4 [1992] 1231-1287). This is not a convincing option. Human social reality is very complex. Some people will want to create intelligent machines, attribute legal personhood to them, and use them in war, simply to gain power over other humans. The source of evil lies in the human heart, which is opportunistic and resourceful even as regards evil. It is enough to recall how terrorism and corruption have benefitted greatly from enhanced social media.

course, we are not saying here that minor-children are not persons until they grow up. We are exploring how we attribute responsibility, and it seems that we sometimes resort, consciously or unconsciously, to the idea of quasi-agent. Although the law does not explicitly say so, it considers minor children as quasi-agents. They do not enjoy the full rights of personhood. They cannot for instance sign a contract. Nevertheless, they are protected just like adults. Legal agency, therefore, can apply in an attenuated form to entities that fall short of personhood in the full sense. Applying this idea to intelligent machines, some argue that, even if technical progress will never reproduce a fully-fledged person, we may still eventually arrive at a situation in which machines will be quasi-agents in the legal sense[16].

The basic point therefore seems to be that the concepts of personal agency, responsibility and autonomy allow for degrees. Admittedly, we often assume that personal agency is either present or absent. We often assume that it is present when the entities that caused the action are fully autonomous. This allows us to hold them responsible. It is absent when the entities that caused the actions are causally determined, when the entities enjoy no or very little autonomy, like non-human animals or minor children. This assumption however neglects the fact that, even with no reference to intelligent machines, we sometimes acknowledge an intermediate conceptual space between autonomous and non-autonomous entities. This space is occupied by entities that are partially autonomous. An entity can be called partially autonomous when those who launch it into action know clearly what ends the entity is meant to achieve but they cannot foresee the action that the entity will use to achieve that end[17]. Child soldiers are a prime example of such partially autonomous or quasi-autonomous agents. The international criminal court, in determining responsibility for war crimes, distinguishes between child soldiers and adult soldiers precisely because of this quasi-autonomy of children. We

---

[16] For instance P.M. Asaro, «Robots and Responsibility from a Legal Perspective», *Proceedings of the IEEE* (2007) 20-24. On the issue of responsibility in general, and how it could be applicable to animals and very young children, see J.M. Fischer – M. Ravizza, *Responsibility and Control: a theory of moral responsibility*, Cambridge 1998.

[17] The idea of ends is important. In general, when things go wrong as regards a product of human technology, we could argue in terms of the ends of the machine and of negligence on the part of the user or on the part of the manufacturer. For instance, when a toy robot causes harm, we could prosecute the manufacturers not for the ends the robot was meant to achieve but for negligence in so far as they did not warn of potential hazards. With intelligent machines, the issue becomes more complicated because the effects are much more difficult to foresee. See Asaro, «Robots and Responsibility» (cf. nt. 16). On computers and quasi-responsibility, see B.C. Stahl, «Responsible computers? A case for ascribing quasi-responsibility to computers independent of personhood or agency», *Ethics and Information Technology* 8/4 (2006) 205-213; DOI: 10.1007/s10676-006-9112-4

could argue therefore that, to determine the responsibility for damage caused by an intelligent machine used as a weapon, we would need to resort to this intermediate conceptual space. We would consider autonomous weapon systems as cases analogous to child soldiers[18].

Is the idea of punishment relevant here? For some researchers, attributing legal personhood to a machine does not make sense even if we were to hold it responsible. The main reason is that the machine cannot be punished[19]. For punishment to be possible, the subject needs to have a moral psychology that is open to the burdens of duties and temptations. Corporations do qualify but only in the sense that the punishment transfers to the punishment of the owners of the corporation. Corporations have a clear objective, namely to make a profit in line with the needs and aspirations of the owners. Intelligent machines however do not seem to have any real, intrinsic objective that could allow us to apply the idea of punishment. Including a punishment module within the circuitry would not count.

Does this brief overview of the main arguments allows us to detect any general trend? I started my paper by quoting from the 2016 proposal at the European Parliament, where we find the suggestion that the law should start recognizing a specific legal status for robots. The debate so far has not made it clear how the granting of legal personhood to intelligent machines could be of benefit. The 2016 proposal and its eventual adoption as a Resolution in fact created a robust reaction from political leaders, researchers in artificial intelligence and robotics, industry leaders, physical and mental health specialists, and experts in law and ethics. A joint statement with more than 150 signatories expressed a strong opposition to the conferring of any form of legal status to intelligent machines[20]. The argument depends on two main points. First, the proposal to confer legal status to robots was founded on the idea that in the near future «damage liability would be impossible to prove». According to them, this is totally incorrect. Secondly, the proposal errs because of «an overvaluation of actual capacities of even the most advanced robots». This overvaluation is the result of science fiction and sensational press releases[21]. The signatories of this

---

[18] This argument is developed further in R. Sparrow, «Killer Robots» (cf. nt. 15). Notice that relatively recent philosophical reflection in the area of animal rights could perhaps give us some ground for speculation on the attribution of moral rights to intelligent machines. See for example, D.J. Calverley, «Android science and animal rights, does an analogy exit?», *Connection science* 18/4 (2006) 403-417; DOI: 10.1080/09540090600879711. The divergence between animals and intelligent machines however is considerable. Fruitful analogical thinking here is possible only when we clarify the underlying semantic issues.

[19] For an example of such reasoning, see L.B. Solum, «Legal Personhood» (cf. nt. 13).

[20] The open letter is available online, from which the quotations in this paragraph are taken: <http://www.robotics-openletter.eu/>.

[21] Popular literature has grossly exaggerated the advances in machine intelligence. So-called

open letter express their expert judgement that all intelligent machines, even when very sophisticated, make decisions that can always be traced back to human agents as regards responsibility. They claim therefore that «creating a legal status of electronic 'person' would be ideological and non-sensical and non-pragmatic»[22].

This strong criticism reveals how, behind the proposal to attribute legal personhood to robots, there could be a hidden political agenda. Hidden agendas in such situations are not new. The classic author Suetonius mentions in his work *Lives of the Twelve Caesars* that Emperor Caligula planned to make his horse Incitatus a consul. Caligula did this probably to ridicule the senate, realizing thereby what is probably the earliest case of a politically motivated attempt to confer legal personhood to a non-human entity. Today's analogue is Sophia, the humanoid robot, with a pretty woman's face, who was granted citizenship by Saudi Arabia in October 2017[23]. This was politically motivated, just like Caligula's proposal. In spite of the fact that experts have harshly discredited Sophia, the human face and the woman's name have subliminal effects on the public. In such cases, it is always important to uncover the ideological and political undercurrents. Some scientists and electronic engineers seek to create a split between their new use of some key words and the standard use of those same words. They do this because such a split allegedly reveals the brilliance, the novelty and the relevance of their work. For some people, one social desideratum of neuroscience, a condition for its progress and of its public acceptability, seems to be precisely this kind of disturbing novelty. The idea of research impact, so crucial for attracting funding, is related to the degree of revolutionary noise that the research will produce. The hidden political agenda may include other desires as well, for instance the desire to gain worldwide prominence within social media, to generate awe and submission before the power of information technology, to alienate the public from serious issues, or simply to discredit traditional values. The most dangerous element of the hidden political agenda in this

---

expert systems do indeed have some capacity for dealing with novel situations that had not been included in the initial programming. Nevertheless, the capacity for dealing with complex novelty is not within reach. Solum argues in «Legal Personhood» (cf. nt. 13) that an intelligent machine can achieve a human-like competence as regards dealing with novelty, but this competence will be limited to one sector only. Serious novel situations, like those faced by humans, require expertise in many sectors.

[22] See also S. Wettig – E. Zehendner, «A legal analysis of human and electronic agents», *Artificial Intelligence and Law* 12 (2004) 111–135; this paper offers a useful study comparing the idea of electronic person in US, Canadian, and German law.

[23] These cases are discussed in U. Pangallo, «Vital, Sophia, and Co.—The Quest for the Legal Personhood of Robots», *Information* 9 (2018) 230, Special Issue *Roboethics*; DOI: 10.3390/info9090230

context is the desire to create ways for exonerating humans from liability in situations when robots cause harm. When the event involves an intelligent machine, the temptation is to distribute responsibility in a way that avoids the just punishment due to humans. Such a temptation should be resisted. The correct strategy is to determine the human origin of the harm done. When a robot causes harm and we cannot identify a scientific reason for the robot's malfunction or misuse, it makes no sense to blame the machine. We should rather seek to compensate the harm by some kind of mandatory insurance system[24].

III. Semantic issues

Let us now try to unveil what is happening at the deeper semantic level. In these debates, we rightly assume that personhood, at least in the legal sense, means the capacity to enjoy rights and perform duties. We assume also that, for these capacities, one needs awareness and free will. Since intelligent machines have no capacity for awareness, we conclude that attributing legal personhood to them is incorrect[25]. Can a philosophical opponent challenge this reasoning? Suppose engineers will add some new sophisticated module to intelligent machines and then announce that, by convention, the meaning of «capacity for awareness» should include also such machines with this added unit. Should we accept this?

To answer this question in the light of recent philosophical advances regarding the theory of meaning, it seems best to start with the collaborative work of Max R. Bennett, a neuroscientist, and Peter M. S. Hacker, a prominent Wittgenstein scholar. These two authors have raised very serious doubts about the way some cognitive neuroscientists and artificial intelligence engineers are using personal attributes to describe their observations and achievements[26]. The problem is semantic. The basic starting point is that the meaning of a word is not arbitrary. A word's meaning is its specific use within the complex social life of language users. This use is governed by rules. Consider the analogy between language and the game of chess. The chess pieces are used

---

[24] This is Solaiman's recommendation, «Legal personality» (cf. nt. 14). Of course, if we can identify a scientific reason, we should hold liable the legal person who was the manufacturer or the legal person who was the user, or both. For more on the legal problems regarding responsibility attribution in this context, see S. Beck, «The problem of ascribing legal responsibility in the case of robotics», *AI & Society* 31 (2016) 473-481; DOI: 10.1007/s00146-015-0624-5.

[25] This point is made in S.M. Solaiman, «Legal personality» (cf. nt. 14).

[26] M.R. Bennett – P.M.S. Hacker, *Philosophical Foundations of Neuroscience*, Malden, Mass. – Oxford 2003.

according to rules, rules that determine the game. For the game to be possible, the rules need to be fixed and accepted by both players. The same happens in language. The rules of grammar determine the correct use of words. These rules do not determine what is said but are accepted by all language-users for language to be possible. Now the specific problem related to machine intelligence arises because some important verbs like «desiring», «intending», «thinking» and «understanding» function fully and correctly only when their subject is a human person. In spite of this, some cognitive neuroscientists and artificial intelligence engineers freely use these verbs when the subject is just the brain or just a computer. Bennett and Hacker show convincingly that this incorrect use causes confusion. They start from the obvious point that, to attribute a specific thought to someone, we resort to that person's behavior characteristics. When that person's behavior is of a certain kind, for instance, when she picks up her umbrella before going out for a walk, then we rightly say, «She thinks it might rain». These behavioral features are part of the very complex social interaction of the person with the environment, an interaction that might also include role-play and deception. Of course, a brain, as distinct from the person, cannot show any such behavioral features. Brains as such do not move around. They are not socially engaged. Admittedly, the brain may have some features that depend on whether that person thinks it might rain or not. But the one who thinks is the entire person not just the brain. First, we determine what the person thinks and then we study the brain, not the other way round. In spite of these fundamental principles, many cognitive neuroscientists still describe the results of their research by speaking about the brain's thinking and reasoning, about one hemisphere's knowing something and not informing the other hemisphere, and so forth[27].

When cognitive neuroscientists do this, they assume that we could separate the concept of person, with all its connotations regarding thinking, reasoning, and understanding, from the social relationality this concept involves. This assumption however cannot but lead to confusion, precisely because the concept of person is essentially connected to relationality. When faced with the question whether an entity before us is a person or not, we need first to apply biological criteria, to ascertain that we are dealing with a genuine living

---

[27] In P.M.S. Hacker, *Wittgenstein: meaning and mind*, Malden, Mass. – Oxford 2003 1990, 162, we find the following two expressions: «Hush, I'm thinking!» and «Hush, my brain is thinking; in a moment it'll tell me what needs to be done, and then I'll let you know!» Hacker of course condemns the latter expression as non-sense. Indeed, we feel that it is a twisted use of words, intended as a joke. We use it to enjoy the surprising breach in meaning that it produces. Notice how the case of Sophia the intelligent robot is similar. Sophia is a joke, a work of art, an artifact to stimulate an audience, touching onlookers in areas of feeling and experience they had never explored before.

thing. Then we apply intelligence criteria to ascertain that we are dealing with an intelligent organism. Finally, we need to apply personhood criteria, to see whether we are dealing with a personal organism or another form of intelligent organism. And especially this latter step involves observing the kind of social relations that organism has with others[28]. Using Ludwig Wittgenstein's vocabulary, we could express the same point by saying that words like «aware» or «conscious» are attributes that depend on the form of life. A computer does not have a form of life. Someone may object by saying that machines nowadays can indeed include processes that correspond to human behavior. Some machines can make use of biological material together with some form of metabolism. More in line with intelligence, they can answer questions typed on a screen, or spoken into a microphone, and therefore these machines do resemble the human form of life. But this resemblance is minimal when compared to the vast array of social behaviors that we need as a criterion for the correct use of important words like «awareness» or «consciousness».

The idea of resemblance here is very important. How close to ours must the form of life of a non-human entity be for that entity to deserve the status of personhood? Our recognition of resemblance is a complex process. It involves holistic perception and does not always imply prior knowledge of precise necessary and sufficient conditions. Wittgenstein's expression «family-resemblance» is very appropriate because it highlights how we recognize resemblance in a holistic fashion and not point by point. A lack of resemblance as regards one aspect may be compensated by resemblance as regards another[29]. The classic Turing test involves resemblance as regards linguistic skill only. The machine is designed to generate human-like responses for questions set by an evaluator. If the evaluator cannot tell the machine from the human, the machine is said to have passed the test. The machine is said to merit the attribute «intelligent». Is it not obvious however, that this is a very reductionist approach? From the innumerable behavior patterns that characterize human rationality, we pick just one element. And this leads inevitably to bias in our conclusion. To avoid this, we need to adopt a holistic approach, one that considers not only linguistic skills but also all other kinds

[28] Here I am focusing on corporeal persons. I am assuming that we can correctly use the expression «non-corporeal person» to the extent that what we seek to describe shares the core, significant features that we are acquainted with when dealing with corporeal persons, features like individuality, intellect and will. This extended use of «person» beyond material conditions, however, is beyond the scope of this paper.

[29] Hence we must be careful not be too essentialistic in our investigation concerning legal personhood. Solum (cf. nt. 13) for instance adopts a typically essentialistic attitude by seeking the essence of personhood, in terms of necessary and sufficient conditions, and then tries to apply this essence to intelligent machines. This approach may indeed be inevitable for legal reasoning but it does not correspond to what happens in real life.

of behavior, including thereby the entire form of life. This is what «family-resemblance» involves. As things are today, an intelligent machine may indeed pass the Turing Test but still fail to pass the family-resemblance test. It will not pass because it does not behave like an intelligent organism in the broad sense. Consider how our understanding functions as regards living things, for instance dogs. Since dogs can neither read, write nor speak, they will certainly not pass the Turing Test. They are therefore definitely unlike humans. Their outward behavior however compensates for this lack of resemblance. That is why we rightly attribute to them some personal attributes, in a derived sense, and can even understand what they communicate with their various sounds[30]. Of course, if a non-biological entity, like a sophisticated humanoid robot, will start behaving like us not only in a restricted linguistic sense but also over a very broad range of social behavior patterns, then we could perhaps decide to call it a living thing[31]. Notice however that this would not be a discovery but a decision. Moreover, to produce a thinking machine that could deserve our calling it a person we need to start with producing animality first, and then we could perhaps arrive at rationality[32].

---

[30] Self-awareness (or self-consciousness) does not involve awareness of a self. A «self» is not a thing. We call non-human animals conscious in a derived sense, suggested by their non-linguistic behavior. It is incorrect to claim that self-consciousness is nothing more than a kind of self-scanning device. The idea of form of life is now recognized as crucial by a number of researchers, e.g. A. Kern, «Human Life and Self-consciousness. The Idea of 'Our' Form of Life in Hegel and Wittgenstein», in C. Martin, ed., *Language, Form(s) of Life, and Logic: Investigations after Wittgenstein*, Berlin – Boston 2018, 93-112. For more on how our ethical relations with intelligent machines are bound up with our social relations with them and on how the social world is itself enabled and constrained by the physical world, and by the biological features of living social participants, see S. Torrance, «Artificial consciousness and artificial ethics: between realism and social relationism», *Philosophy & Technology* 27 (2014) 9–29; DOI: 10.1007/s13347-013-0136-5.

[31] An important legal point here is that, when we have a high degree of anthropomorphic representation, as in humanoid robots engaged in social life, people are likely to impute to the machine more competence than it deserves. This could make the manufacturers liable because the situation could amount to deceit. The robot could be in fact a misrepresentation made with the express intention of defrauding someone. This point shows how important it is for us to ensure that legal liability standards remain in pace with evolving technology. See C. Heckman – J.O. Wobbrock, «Liability for Autonomous Agent Design», *Autonomous Agents and Multi-Agent Systems* 2/1 (1999) 87-103; DOI: 10.1023/A:1010087325358.

[32] This point is made in P.M.S. Hacker, *Wittgenstein* (cf. nt. 27), 170. Some convincing arguments, however, show that even the first step, namely to produce a mechanical organism, is impossible. See M. Schark, «Synthetic biology and the distinction between organisms and machines», *Environmental Values* 21/1, special issue on *Synthetic Biology*, (2012) 19-41. The way people are ready to attribute emotions and other personal attributes to robots is the central interest of P. Dumouchel – L. Damiano, *Living with Robots* (cf. nt. 4). It is a pity that this book remains detached from the considerable advances made these last decades in the area of philosophy of meaning.

These arguments point clearly to the following conclusion. It is not true that, if an entity functions like a person in some restricted sense, for example as regards linguistic skills only, then it deserves personhood in some sense. This conclusion is an invitation for us to retrieve the richness of pre-Cartesian philosophical anthropology according to which the person is not made up of two separate substances body and soul but is, on the contrary, a unity that can be appreciated both from the material viewpoint and from the intellectual or spiritual viewpoint. This is the valuable heritage of Aristotle and Aquinas. The specifically human form of life, which we use as a criterion for the correct attribution of personhood, includes not only endo-somatic relations, those between one part and another part of the body, but also exo-somatic relations, those connecting the entity to things outside it. The brain functions not on its own but in constant symbiosis with the entire body. Likewise, the person, the unity of mind and body, operates not on its own but in constant symbiosis with the entire social and cultural relational space that makes rationality possible[33].

To appreciate better the benefits of the position defended so far, it is helpful to consider a major philosophical objection that has been levelled against it. Daniel Dennett criticized this position severely by arguing that semantic rules of grammar seem fixed but are not really so[34]. For him, these rules are nothing more than collective habits. They are nothing more than the result of human interaction with the environment over millennia, rules that ensure a stable and useful production of human sounds. He disagrees with those who see grammatical rules as logically compelling. For him, the entire network of rules is just a feature of human social life, a feature that should be the object of study for anthropologists, not logicians. When radical changes occur within society, some of these rules may be revised to achieve better overall efficiency. Dennett is moreover convinced that the emergence of highly intelligent and autonomous machines represents precisely such a radical change that demands a revision. It justifies the use of expressions that we have traditionally banned, expressions like «this machine understands» or «this machine is a person»[35].

---

[33] Christian theology highlights these same fundamental principles. The *Catechism of the Catholic Church*, Vatican City – London 1994, summarizes the theological concept of «soul» in the following way. «In Sacred Scripture, the term *'soul'* often refers to human *life* or the entire human person. But 'soul' also refers to the innermost aspect of man, that which is of greatest value in him» (§363; italics in the original). Also: «spirit and matter, in man, are not two natures united, but rather their union forms a single nature» (§365). These two quotations show how a naïve Cartesian dualism is foreign to the correct theological understanding of the soul.

[34] See M. Bennett – *al*., *Neuroscience and Philosophy: Brain, Mind, and Language*, New York 2007.

[35] For a study on how Dennett's stance is applicable to questions regarding the moral status of robots, see S. Chopra, «Taking the moral stance: morality, robots, and the intentional stance», in B. van den Berg – L. Klaming, ed., *Technologies on the stand: Legal and ethical questions*

This criticism, in its apparent defense of human freedom as regards the use of words, may seem very attractive. Nevertheless, it rests on an oversight. For any kind of scientist or anthropologist, the sense of a hypothesis must be settled before he or she could determine whether that hypothesis is true or false. For a string of words to make sense, one needs to respect a set of rules. Only then can scientists conduct their inquiry to decide about the truth or falsity of what is proposed. If neuroscientists or legal experts want to change the rules regarding the word «person», they need to realize that such a change generates conceptual repercussions elsewhere. When a rule is revised, it will produce a ripple effect that could go right across the entire conceptual landscape, an effect that may cause confusion. The new rules could for instance have detrimental effects on the use of that same word in everyday contexts. Such scientists therefore would do well to be aware of the need to reestablish consistency across the entire range of human experience and not only in their own specialized semantic space.

As an example of how inconsistency and conceptual confusion could emerge, consider the hidden links between personal attributes and ethical concepts. Amelie Rorty in a short paper published more than fifty years ago but still relevant today, exposed how «the word 'think' is not only used descriptively, but also implies an ethical decision or attitude»[36]. Her argument is about the use of «think» as applied to machines. To decide whether we could correctly speak of a machine as thinking, the Turing Test implies that, out of all the possible forms of human behavior, one specific form determines whether a machine thinks or not. Rorty rightly points out however that when I say, «This X thinks», I am not only describing its behavior. I am also saying something about myself. I am saying something about the proper behavior I should have towards X. I am saying something about showing respect towards X, about treating X with dignity. Rorty here is encouraging us to explore the broad conceptual network associated with crucial expressions like «to think», «to respect», and «to treat as a person». These conceptual areas merge into one another. It is incorrect to use «to think» without the ethical overtones that go with it. What lies behind this overlap between conceptual spaces is not just descriptive but also prescriptive. It is prescriptive in the sense that it indicates what I ought to do. For instance, in a geriatric ward, I treat this person in front of me with dignity even though his thinking is confused. Of course, I could neglect this ethical imperative. I could even be brainwashed to treat no one with dignity. The fact that we disapprove of such neglect, however, shows that we

---

in neuroscience and robotics, Nijmegen 2011, 285-295. Chopra argues that, when an artificial agent's behavior is of a certain kind, it merits our ascribing to it a moral sense. The paper however does not consider the broader conceptual issues, e.g. problems that arise when the corresponding responsibility ascribed to machines is used to exonerate humans.

[36] A.O. Rorty, «Slaves and Machines», *Analysis* 22/5 (1962) 118-120; the quote is from p. 120.

indeed recognize the imperative. Rorty rightly exposes therefore the inevitable conceptual links between the two conceptual spaces, the one associated with thinking and the other corresponding to some fundamental and non-negotiable attitudes within human living. The upshot is that claiming that robots think, understand, decide and are indeed persons has wide consequences in our life in general including our moral attitudes.

Probably, philosophers like Dennett would not find this compelling. They would claim that such links between descriptive and normative expressions are not as untouchable as we are assuming. They would probably argue that it is perfectly possible to tear apart the conceptual fabric right there, on the fault line between thinking and respecting as a person. If we start using personal attributes without their ethical overtones, what then?

In that case, the situation will be like changing the rules of an established game. Suppose we start playing chess and then we decide to change the rules. Would we still be playing chess? Given that the game is defined in terms of its rules, the answer is no. We would have introduced a new game or a new type of chess. We would have chess according to the old rules and chess according to the new rules, two separate games. Analogously, a change in semantic and grammatical rules would produce a complex situation in which a word means one thing in one context and another thing in another context. For instance, it would have ethical implications in one context and no ethical implications in another. Both contexts could enjoy internal consistency but the speaker would need to acquire the skill of switching from one context to the other. This leads to fragmentation of language and eventually of culture itself, as described by Jean-François Lyotard in his study of post-modernism[37]. Is this is a healthy way forward for humanity? As faster communication, more efficient travel, and global concerns are helping us to grow into a global community, favoring the conditions for the realization of a truly universal family, it would be self-defeating to encourage fragmentation at the conceptual level.

Could we perhaps just wait and see what happens? In the name of new technology, we inflict conceptual lacerations onto the semantic fabric that was meant to keep our understanding as self-consistent as possible. Could we just wait for these lacerations to heal on their own? Some researchers have indeed advocated such an approach, which we could call pragmatic. Lawrence Solum for instance argues that we do not need to waste our time now to figure out beforehand whether or not we should attribute legal personhood to intelligent machines in the future. When the time comes, when machines will reach a high level of intelligence and autonomy, when they will have a will of their own, then, Solum argues, we will see how society reacts. We will see how society adjusts itself. This approach is pragmatic in the sense that it assumes

[37] J.-F. Lyotard, *La condition postmoderne : rapport sur le savoir*, Paris 1979.

a kind of invisible hand that guides us always towards the optimal situation[38]. It implies that we should concentrate on what works well as regards our immediate needs and should avoid the deeper metaphysical question of whether machines really have awareness, intention, or understanding. This assumption may seem reasonable but, in fact, is misleading. It leads to no solution at all. Trust in providence is a very cheap excuse for a laissez-faire mentality. The invisible hand that alleged guides humanity has not always resulted in the optimal situation. Far from it. Without deliberate, responsible and courageous decisions, based upon the attentive consideration of the future consequences of our action, chances are that humanity will just spin out of control and fall into the hands of demagogues. The pragmatic approach therefore is not convincing. It cannot revoke the duty to think ahead. It cannot substitute a responsible, detailed study of the way shifts of meaning in one area of our conceptual scheme could affect meaning in other areas, causing confused thinking and problematic legal practice[39].

Conclusion

The intricate set of mutually dependent questions discussed in this paper constitute a broad issue that will be with us for decades. The paper has limited itself to a philosophical evaluation of one area only. Philosophy, of course, is no substitute for natural science or for technological innovation. The research efforts of these last decades in the areas of cognitive neuroscience and robotics have been enormous. The resourcefulness, creativity and care represented by such collaborative research undoubtedly demands respect. Philosophy is here a partner with a specific role. It contributes by being attentive to conceptual

---

[38] I use here the expression invisible hand with explicit reference to Adam Smith's famous use of it in the context of economics. Just as Smith assumed that the stability of the system is guaranteed by an invisible hand, even though the individual agents act in their own self interest, so also the pragmatist view described here assumes that logical consistency is guaranteed even if semantic changes are allowed without any concern about the overall picture. There are plausible reasons to hold that Smith used the expression in a religious sense because of his admiration for Isaac Newton. See for instance P. Oslington, «God and the Market: Adam Smith's Invisible Hand», *Journal of Business Ethics* 108/4 (2012) 429-438.

[39] I am not advocating radical semantic conservatism at any price. Cultural, technological and intellectual novelty is to be supported, even though it is nearly always associated with neologisms. What I am advocating is caution and responsibility in this area. Not all neologisms are beneficial or neutral. Sometimes they can have damaging consequences, even when accepted by the majority. As an example think of how the expression *Lebensunwerten Lebens* (a life unworthy of life) was one of the catchphrases during the Nazi era (in use between 1920 and 1944), with devastating legal consequences involving the unrestricted killing of innumerable vulnerable human subjects.

links and by identifying any problematic transgressions of the bounds of sense. In this task, it is concerned not with matters of fact regarding the material world but rather with matters of meaning. It is concerned, in other words, with facts about rules for understanding. Philosophy can help to reveal and correct those conceptual confusions and misconceptions that sometimes emerge because of changes in culture and everyday life.

This paper's overall result can be summarized in three points. First, it showed that a very liberal use of personal attributes to describe non-human entities, attributes involving intelligence, understanding, willing and personhood, is detrimental, especially if we use such words univocally. Negligence in this area undermines conceptual consistency because it disregards very important normative nuances that these attributes have. It should be clear by now therefore that the expression «intelligent machines» carries nuances that could be misguiding. Perhaps a better designation would be one that highlighted the fact that these machines, like all other machines, are instruments for our use. They are sophisticated electro-mechanical instruments, tools to help us with *our* entertainment, *our* intelligence, *our* understanding, *our* willing and *our* personhood.

Secondly, society may indeed persist in its misconceptions and may indeed end up attributing legal personhood to machines. It would thereby neglect how concepts relate to each other and would assume that the conceptual scheme will somehow heal itself through the action of an invisible hand. This paper argued that such a move would be severely irresponsible. Even in this case however, we would still be obliged to distinguish between what is human and what is not. The distinction between human and artificial intelligence, or that between human and artificial person, would remain inevitable. We would probably start using expressions like «human person» as distinct from «machine person». We would probably start saying that the intelligent machine is like us in everything except for being human. There is a deep desire within us to defend our identity[40]. We naturally insist that no amount of simulation would ever be enough to justify complete equality between humans and machines. This deep desire to defend our human specificity reflects the fundamental distinction between the categories of natural and artificial, between φύσις and τέχνη, a distinction that cannot be overruled.

And thirdly, as regards the specific question of machines as legal persons, the paper has made it clear that there could be a hidden agenda. The most dangerous aspect is the possible hidden agenda of wanting a kind of legal machinery to exculpate the real perpetrators when things go wrong. To be prepared for dealing with wrongs committed by machines, we should not create a legal fiction that could serve as a shelter for wrongdoers. We should

[40] Cf. L.B. Solum, «Legal Personhood» (cf. nt. 15).

rather invest the necessary time and energy to clarify the various forms of accountability for complex cases where responsibility is distributed among many agents[41]. Blaming the machine is not the way forward[42].

Pontificia Università Gregoriana, Louis Caruana, S.I.
Piazza della Pilotta, 4
00187 Roma
caruana@unigre.it

ABSTRACT

To describe computers and sophisticated robots, many people today have no problem using personal attributes. Alan Turing published his famous intelligence test in 1950. From that time onwards, computers have gained increasingly higher status in this regard. Computers and robots nowadays are not only intelligent. They perceive, they remember, they understand, they decide, they play and so on. Recently, another such step has occurred but, this time, many researchers are seriously concerned. In February 2017, the European Parliament passed a Resolution to attribute legal personhood to intelligent robots. If this is accepted as law, it will have very serious consequences for our self-understanding and for the way we live together as a community. The EU Resolution has stimulated various studies, arising mainly from the area of legal studies. It is urgent that the response include also a philosophical evaluation regarding the fundamental concepts at play. This paper seeks to make a contribution precisely in this area. It explores the attribution of legal personhood to machines by focusing on what is happening at the level of meaning. It explores crucial concepts like responsibility, autonomy, person and quasi-person by drawing inspiration from the seminal works of Aristotle and L. Wittgenstein and from the ensuing debates between current philosophers like P. Hacker and D. Dennett. The results of this paper indicate what dangers could lie ahead and what could be the right way to avoid them.

Keywords: artificial intelligence, legal person, autonomy, robot, Wittgenstein

RIASSUNTO

Per descrivere computer e robot sofisticati, molte persone oggi usano senza nessun scrupolo attributi personali. Alan Turing pubblicò il suo famoso test di intelligence nel 1950. Da quel momento in poi, i computer hanno acquisito uno status sempre

[41] This is one of the points advocated by U. Pangallo, «Vital, Sophia and Co.» (cf. nt. 23).

[42] Previous drafts of this paper were presented and discussed during the *Congreso Internacional de Filosofía de la Mente*, held at Morelia, Mexico (25-27 March 2019) and during the *Transhumanism Conference*, held at the University of Comillas Madrid, Spain (29-31 May 2019). Thanks are due to all those who helped by their questions and suggestions during these conferences, and also to an anonymous referee for *Gregorianum*.

più elevato in questo senso. I computer e robot al giorno d'oggi non sono soltanto intelligenti ma anche percepiscono, ricordano, comprendono, decidono, giocano e così via. Di recente si è verificato un altro passo del genere, ma questa volta molti ricercatori sono seriamente preoccupati. Nel febbraio 2017, il Parlamento Europeo ha approvato una risoluzione per attribuire la personalità giuridica ai robot intelligenti. Se questo è accettato come legge, avrà conseguenze molto serie riguardanti il modo in cui noi comprendiamo noi stessi e il modo in cui viviamo insieme come comunità. La risoluzione dell'UE ha stimolato vari studi, associati principalmente agli studi giuridici. È urgente che la risposta includa anche una valutazione filosofica relativa ai concetti fondamentali in gioco. Questo articolo cerca di dare un contributo proprio in questo settore. Esplora l'attribuzione della personalità giuridica alle macchine concentrandosi su ciò che sta accadendo al livello del significato. Esplora concetti cruciali come la responsabilità, l'autonomia, la persona e la quasi-persona, traendo ispirazione dalle opere fondamentali di Aristotele e di L. Wittgenstein e dai conseguenti dibattiti tra filosofi attuali come P. Hacker e D. Dennett. I risultati di questa indagine indicano quali pericoli potrebbero presentarsi e quale potrebbe essere il modo giusto per evitarli.

Parole chiave: intelligenza artificiale, personalità giuridica, autonomia, robot, Wittgenstein