ARTIFICIAL INTELLIGENCE, ROBOTS AND THE ETHICS OF THE FUTURE

CONSTANTIN VICĂ, CRISTINA VOINEA

Abstract. The future rests under the sign of technology. Given the prevalence of technological neutrality and inevitabilism, most conceptualizations of the future tend to ignore moral problems. In this paper we argue that every choice about future technologies is a moral choice and even the most technology-dominated scenarios of the future are, in fact, moral provocations we have to imagine solutions to. We begin by explaining the intricate connection between morality and the future. After a short excursion into the history of Artificial Intelligence, we analyse two possible scenarios, which show that building the future with technology is, first and foremost, a moral endeavor.

Keywords: AI, robots, ethics, future, automation.

MORALITY AND THE FUTURE. THE ETHICS OF THE FUTURE

The future rests, as always, under the sign of technology. Karl Marx famously observed that: "Men make their own history, but they do not make it as they please; they do not make it under self-selected circumstances, but under circumstances existing already, given and transmitted from the past"[1]. And, indeed, the future, as it is currently envisaged, seems to be in the hands of almost blind forces that orient it towards the obsolescence of men. According to one of the most

Constantin Vică ✉
Faculty of Philosophy, University of Bucharest
204 Splaiul Independenței, 060024, Bucharest, Romania
e-mail: constantin.vica@filosofie.unibuc.ro

Cristina Voinea ✉
Department of Philosophy and Social Sciences, Faculty of Management
Bucharest University of Economic Studies
2–10 Căderea Bastiliei Str., Bucharest 010374, Romania
e-mail: cristina.voinea@man.ase.ro

---

[1] Karl Marx, *The Eighteenth Brumaire of Louis Bonaparte*, New York, Mondial, 2005, p. 1.

Rev. Roum. Philosophie, **63,** *2*, p. 223–234, București, 2019

prevalent predictions, the machines we have built will become autonomous, taking up more and more of the tasks classically reserved to human beings[2]. This scenario has caught people's attention because it seems plausible: after all, machines do not revolt and nor do they get sick; they can't be exploited, nor can they feel alienated. It would only be rational to prefer them, instead of human beings, as workers. They are the perfect continuators of what Shoshana Zuboff called surveillance capitalism[3], a regime based on the logic of accumulation of data, which is further used for discovery, anticipation and manipulation of human behavior.

But the future cannot be reduced to strictly technological problems, in as much as it is a time of moral duty and action. As Vladimir Jankélévitch observed[4], we conceive our moral reality by reference to the future. This orientation towards the future is inherent in the language of ethics: the imperative clause, used to convey the right or good action to the moral subject – disguised as a moral imperative, moral command or an order, etc. – is, by nature, a species of the future tense. Similarly, when we think about consequences, when we imagine the ramifications of an action, we look towards the future. The time of morality is fixed by performative acts of language, urging action, doing. Through this continuous reference to the future, morality escapes a "negative" temporality.

At the same time, discussions about morality are meaningless if we do not also assume, or even *postulate*, freedom as the main condition of possibility for morality. As Kant observed in his *Critique of Judgement*, the world made possible by freedom (as a "formal condition" and a "general regulative principle"), or the realm of morality, is similar to a virtual world, a world where everything exist "*als ob* dieses geschähe"[5], i.e., as it is for real or as it happened for real. The future is the actualization of an open possibility, which emerges from the physical in determination of human goals, be they good or bad, correct of incorrect, or somewhere in between. As such, even if we accept determinism in its physicalist guise, it does not also imply accepting moral determinism – the moral life, with its beliefs and actions, is governed by principles and rules which are not reducible to, caused by or supervenient on *what it is*, or on plain physical existence. Although predetermination seems to be at work at the atomic and neural levels, and even if neurosciences and psychology depict an image of a human being that resembles an *automaton*, something still seems to escape all these regularities: morality. In fact, the individual process of moral deliberation and evaluation is always articulated in reference not only to our feelings, uneducated intuitions or instinctive reactions, rooted in a complicated neural architecture, but also to principles and norms which are perpetually to be reevaluated, all of them demanding the presence of human will.

---

[2] Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies*, Oxford, Oxford University Press, 2014.

[3] Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* London, Profile Books, 2019.

[4] Vladimir Jankélévitch, *Curs de Filosofie Morală*, București, Polirom, 2011, p. 115.

[5] Immanuel Kant, *Critica Facultății de Judecare*, București, Editura ALL, 2007, section §76.

Secondly, morality presupposes a set of socially shared acts, embedded in a cultural system of reciprocal demands and obligations[6] which are not dependent on another substrate.

On the one hand, thinking about the future means putting our moral imagination at work. But making the future, as Nick Montfort observes[7], is also a moral act. This is the productive perspective on the future, which shifts the emphasis on agents and their powers of building the future they imagine. Montfort thus sees all attempts at predicting the future as a delegation of our duty of thinking and building the future on our own terms, with us as the main agents[8]. Consulting an oracle, divination and prophecy depict an inhuman or non-human future in which we are invited, but we are never autonomous, free agents or *homo faber*.

Today, science and innovation are the new oracles. The accelerating pace of innovation invites us to imagine a future where people will lose their jobs because they will be displaced by robots, where the so-called Singularity moment will enslave human beings or where machines will be better, faster and more productive than human beings. At the same time, there seems to be something implacable about technology (especially when it comes to digital): sooner or later, when everybody on earth is connected, the amount of data and, implicitly, of technological progress, will grow exponentially[9]. The future will be defined and shaped only by technology. This is the ideology of *inevitabilism* specific to Silicon Valley: "As in most accounts of the apparatus, questions of individual autonomy, moral reasoning, social norms and values, privacy, decision rights, politics, and law take the form of afterthoughts and genuflections that can be solved with the correct protocols or addressed with still more technology solutions" [10]. The future seems to be dominated by technological and not by social or political progress, because, in a way, the need for the latter will be eliminated once the former is actualized. As such, it is no wonder we talk more about what the next generation of robots will look like, or about the emergence of an Artificial General Intelligence, than we do about bettering social and political institutions. The future thus seems determined by technology.

Going back to Jankélévitch, our claim is that thinking about the future and making it are moral acts. Making the future is a human being's most natural vocation. As work in progress, the future is only built through human will, which is creative and capable of inventing new possibilities[11]. The future is the time of the will to act which, in its turn, is the result of a sole condition – the possibility of freedom. Even the most technology-dominated scenarios of the future are moral

---

[6] Peter F. Strawson, "Social Morality and Individual Ideal", in *Philosophy* vol. 36, no. 136, 1961, pp. 1–17.

[7] Nick Montfort, *The Future*, Cambridge, Massachusetts, MIT Press, 2017, p. 4.

[8] Idem, p. 22.

[9] Zuboff, *op.cit.*, p. 211.

[10] Idem.

[11] Jankélévitch, *op. cit.*, p. 126.

provocations we have to imagine solutions to because designing the technology of the future is a moral endeavour, as will become evident in the last section. As such, a technology-driven future cannot be reduced to strictly engineering issues – how do we make this or that technology or process more efficient, productive and sustainable – because the future is always a duty we have towards the others. The duty of care towards the future is not necessarily a duty towards precise future individuals, as Parfit shows[12], but a duty towards humanity in general, as Hans Jonas observes[13]. In what follows we will analyse two technology-dominated scenarios of the future, and we will show that the problems we face are, first and foremost, moral problems. But first we will go into a short excursion into the history of Artificial Intelligence – a technology that seems to dominate scenarios of the future.

ARTIFICIAL INTELLIGENCE AND ROBOTS

The field of computer science, generally called Artificial Intelligence (AI), is deeply intertwined with philosophy, although AI is part of the larger scientific program of cognitive sciences. But the origin of AI is dual: both in the world of mathematics and in those of philosophy and psychology.

The first schism in the world of AI researchers took place during the 1960s[14]. The majority remained faithful to logical formalism, and to the attempt to create inferential propositional systems; but a small minority of researchers turned towards observing and emulating human behaviour. The latter position takes into account simple facts: one can build refined languages for AI – through which certitude can be represented and knowledge acquired –, but the entity equipped with that sort of a language system might not ever discover how to take a step on accidental terrain. This leads to some of the most interesting philosophical tensions: symbolic processing vs. neural networks, mind vs. life, symbolic computation vs. cybernetics, formalism vs. connectionism.

But the idea of an artificial intellect, akin to the human one, was born long before the 1960s and, naturally, it comes from a woman. Countess Ada Lovelace was sure, in the 1840s, that machines would be able to compose music, explain facts of nature and inaugurate a new era for science. At the same time, Ada Lovelace is recognized as the first computer programmer in the world, having described an algorithm, now acknowledged as the first model of a computer program[15]. Her *poietic* vision was further developed by a plethora of scientists like,

---

[12] Derek Parfit, *Reasons and Persons*, Oxford, Oxford University Press, 1984, pp. 351–356.

[13] Hans Jonas, *The Imperative of Responsibility: In Search of an Ethics for the Technological Age*, Chicago, University of Chicago Press, 1985.

[14] Margaret A. Boden, *Artificial Intelligence: A Very Short Introduction*, Oxford, Oxford University Press, 2018.

[15] Robin Hammerman and Andrew L. Russell (eds), *Ada's Legacy: Cultures of Computing from the Victorian to the Digital Age*, New York, Morgan & Claypool, 2016.

for example, Alan Turing, who was the first to prove that any possible computation can be, in principle, done by what we now call a Universal Turing Machine[16].

Ever since AI's inception, philosophers have mounted numerous attacks against it. Out of them two stood the test of time. The first critique maintains that it is impossible to create disembodied cognition or an intelligence without a body. Hubert Dreyfus showed that an artificial intellect is a contradiction in terms because intelligence must emerge naturally, alongside other faculties, like sensibility, as such it too must be situated somewhere[17]. In 1980, John Searle explained that an intelligent and powerful computer, let's say like Deep Blue which beat the world chess champion Gary Kasparov in 1997, cannot understand what it does or, in this specific case, the significance of its victory. Deep Blue is just an extremely skilful program which can manipulate data following rules, but it won't ever be able to meaningfully represent the world, people or itself. The semantic level of language will always be an intentional, human affair[18].

Today, AI's germline is represented by all those connected to the network, out of whose data and documents – replicable, just like biological genes – a new edifice of control or liberation is being built. What does it mean to be the germline of AI? Taking into account both machine learning techniques and formal artificial thinking, the sources for AI models, strategies or algorithms spring from the human cognition and action, from how these are represented through data and used by inferential rules (as in logic and mathematics). Some have claimed that, once again, humanity created something radically new, which has the potential to eradicate human beings much like atomic weapons: these potential risks call for humanity's transition to a posthuman era[19]. But unlike the atomic weapon, AI can correct, improve and replicate itself. And unlike the atomic weapon, AI needs us in order to operate, for without data generated by human (and other natural, biological) beings, algorithms are useless.

Just as the basis of our intelligence is biological, so the support of Artificial Intelligence is a program or piece of software. In this sense, AI is both material – it wouldn't be plausible in the absence of a huge amount of data and computational power – and textual. AI is the production of algorithms, or algorithmic strategies, that respond to inputs and produce outputs[20]. For example, recommendation systems (on Amazon, Spotify or Netflix), expert systems in healthcare or systems used for the assessment of insurances, are all based on machine learning, one of the

---

[16] Alan M. Turing, "Computing Machinery and Intelligence", in *Mind*, Vol. 49, pp. 433–460.

[17] Hubert L. Dreyfus, *What Computers Can't Do: A Critique of Artificial Intelligence*, New York, Harper & Row, 1972.

[18] John R. Searle, "Minds, Brains, and Programs", in *Behavioral and Brain Sciences*, Vol. 3, no. 3, 1980, pp. 417–457.

[19] Nick Bostrom, "Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards", in *Journal of Evolution and Technology*, Vol. 9, 2002.

[20] Lucas D. Introna, "Algorithms, Governance, and Governmentality On Governing Academic Writing", in *Science, Technology & Human Values*, Vol. 41, No. 1, 2016, pp. 17–49.

most popular set of techniques for AI. Health, reputation, trust and safety are all evaluated with the use of AI. Even voting options can be influenced through such programs running on digital platforms.

Some might say that with current AI systems, human beings are always in the loop. After all, they must label the data fed into AI systems and also supervise the way algorithms learn. But a new paradigm is gaining importance: unsupervised learning. The connectionist project starts from the simple idea that in order to create a better Artificial Intelligence, one must start by simulating, as much as possible, human neural networks, as a first step in mimicking the architecture of the human brain. The most powerful forms of AI today use unsupervised learning, which essentially bring together Big Data and computational power, sometimes running on virtual machines which accommodate neural networks.

According to Tom Mitchell, machine learning appears whenever: "A computer program is said to learn from experience *E* with respect to some class of tasks *T* and performance measure *P* if its performance at tasks in *t*, as measured by *P*, improves with experience *E*"[21]. Machine learning produces new results in comparison to human analysis and it can even incorporate past experiences. It can be used to generate recommendations – crucial for entertainment services like Netflix or for retail services like Amazon – but it can also be employed in the attempt to discover new forms of treatment for cancer. There is a suit of algorithms produced through machine learning techniques like:

1. supervised learning – the system is provided with data marked as "the correct answer" for each example processed.
2. unsupervised learning – the algorithm looks for structure in data and similarities between the examples fed into it, which are then clustered in groups.
3. reinforcement learning – in which the algorithms are trained by using a system of rewards and punishments.

The common problem of all of these cutting-edge techniques is not that they will develop human-like consciousness, thus setting their own aims and destroying humanity. The real issue has to do with their opacity – these algorithms are "black boxes" even for their trainers[22]. Programmers can understand why a certain system produces a result, the results can be checked, but they cannot understand how that final output is reached.

AI is actually a provocation to rethink morality: if we can train intelligence *in silico*, and intelligence is one of the criteria for granting an entity moral status, then we should grant AI moral status and even rights. Because, in the end, when we claim that something is intelligent, this is not just a psychological or epistemological observation, but a moral judgment[23]. But these are all problems for the far future,

---

[21] Tom Mitchell, *Machine Learning*, New York, McGraw Hill, 1997, p. 2.

[22] Pedro Domingos, *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*, 1 edition, New York, Basic Books, 2015.

[23] Amelie O. Rorty, "Slaves and Machines", in *Analysis*, Vol. 22, No. 5, 1962, pp. 118–20.

when AI becomes conscious of its intelligence. For now, AI produces knowledge without understanding, and for the majority of people these processes happen in a "black box". The nihilist vision which predicts the disappearance of our species because of a malevolent artificial entity is justifiable as a belief, even if it is not quite plausible. The real issue is that we are witnessing a rising tide of algocracy[24], where those that own the data and the algorithms are becoming increasingly powerful. This is due to the fact that AI models are not simply used as a neutral procedure, but as a basis for public policy and even as social and political institutions.

THE ETHICS OF THE FUTURE THROUGH ROBOTS AND AI

Thinking about the future is a moral endeavour. Most of the time we are tempted to believe that the present is the time of morality, because it is the time of doing. But the main concepts of morality, those of duty and of responsibility, are always articulated in relation to the future. For example, moral responsibility, equated with prospective responsibility, refers to acts that have not yet happened and to the duties that we must fulfil[25]. As such, morality regards the future and is deeply connected to freedom and hence contingency. Only inasmuch as man if free, the future appears as a construction site in which we can exercise our powers, or will, and actualize various possibilities[26]. The future is more morally laden than the past or the present; the present will become the past, while the future will always become present. This natural course of time presents us with two options: the first option is to go with the flow and adopt a passive attitude towards the future and let it actualize to see what it has in store for us; in the second option we can "help" time through moral acts – human beings change the face of the future in accordance to their moral will[27].

Technology is here with us and it is highly implausible to imagine a future without it. Our claim is that any scenario of the future involving technology presupposes, first and foremost, a moral dimension. In other words, when we imagine the future with AI or with robots, we must firstly think about what we want technology to do for us in terms of moral good, justice and fairness, and not only in terms of efficiency and productivity. As we claimed earlier, the future is a construction site which invites us to act, after engaging in moral deliberation. In what follows, we reflect on two very plausible future scenarios and highlight their moral implications. The short history of AI laid out in the previous section is helpful in discerning what we can reasonably imagine technology will do in the near and not so near future. As such, we will not talk about the moral implications

[24] Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information*, Cambridge, Harvard University Press, 2015.

[25] Jankélévitch, *op.cit.*, p. 123.

[26] *Ibidem*, p. 128.

[27] *Idem*.

of Artificial General Intelligence or of cyborgization – which even though not quite unreasonable, are still achievable in a very far future (if we are to believe in the present constraints of technological innovation and scientific discovery). In exchange, we will deal with possibilities which already have taken root in the present. More precisely, we will discuss the disappearance of work in a world of automation and algorithmization, and the issue of love and robots. We will show that the main questions of these scenarios are of a moral nature.

SO LONG, AND THANKS FOR ALL THE WORK!

The most pressing issue related to automation and Artificial Intelligence is already with us: artificial agents are replacing more and more people in their jobs, turning the human workforce into something obsolete[28]. Although the introduction of automation will also create some new types of jobs – probably only in the service of AI, such as trainers, explainers or sustainers[29] –, it is expected that the number of people who will be unemployed will exceed the number of those who will have jobs. In *Race against the Machine*[30], Erik Brynjolfsson and Andrew McAfee argue that the introduction of automation looks very much like the agricultural revolution of the last century, when the number of farmers has decreased from 40% of the total workforce, to only 2%. The fear of automation is accentuated by the fact that not only physical, but also cognitive jobs are automated today. Whether we talk about the work of cashiers, postmen, accountants or experts, statistics show that robots are gaining ground compared to human workers. Some even worry that the rising level of unemployment will contribute to the rise of populism[31].

Even if most jobs will disappear, there is still something left for people to do and which only they could do: politics. With enough time on their hands, but also with some means through which a decent living can be assured, people could participate and bring to life a real deliberative, epistemic and participative democracy, in a minimal state. So, we argue that there are still things left for people to do, even in a scenario in which most jobs will be automated. The significant moral question now is: who has a duty of care towards the people who will lose their jobs? Or, even better yet, what sort of institutional arrangement should we conceive to truly bring to life a perpetual participatory democracy, undistorted by economic factors?

---

[28] David Rotman, "How Technology Is Destroying Jobs", in *Technology Review*, Vol. 16, No. 4, 2013, pp. 28–35.

[29] H. James Wilson, Paul Daugherty, and Nicola Bianzino, "The Jobs That Artificial Intelligence Will Create", in *MIT Sloan Management Review*, Vol. 58, No. 4, 2017, pp. 14–19.

[30] Erik Brynjolfsson and Andrew McAfee, *Race Against the Machine: How the Digital Revolution Is Accelerating Innovation, Driving Productivity, and Irreversibly Transforming Employment and the Economy*, Lexington, Massachusetts, Digital Frontier Press, 2011.

[31] Frank Levy, "Computers and Populism: Artificial Intelligence, Jobs, and Politics in the near Term", in *Oxford Review of Economic Policy*, Vol. 34, No. 3, 2018, pp. 393–417.

The only way to actualize this utopia is a total reformation of the state, the institution of private property and markets; all of them should be considerably limited in their scope. A minimal state using artificial intelligence technologies and blockchain could tax robots and owners of AI systems, and the collected funds could be reinvested in a (digital) universal basic income[32]. We are not the first to observe that, even today, the profits of technology companies developing innovations based on Big Data and AI would not be possible without people's data[33]. Through a combination of prejudice and privilege, the data which fuels most technological innovation has been treated as capital rather than labor, meaning that the benefits of its processing have not trickled towards its rightful owners, the users. Just as the work of slaves was seen as capital rather than labor, so the services, apps, sensors trackers used today are seen as entertainment rather than labor. The idea is that nothing, besides free services, returns to the users after the use of digital technologies, while the tech industry is becoming even bigger and more powerful than the oil or weapon industries. The issue is thus a conceptual one; if we change the way we see data – from capital to labor – then it would be theoretically possible to find a justification for users getting a share of the profits, as Lanier[34] or Posner & Weyl[35] so pertinently show.

But even this solution raises some new questions. For example, is freedom possible in a world without work? Labor and human enterprise were, historically, the main way of gaining, preserving or extending freedom[36]. During Antiquity, slaves were liberated as a result of the hard labor they underwent for their masters. Private property (which is a result of either labor or theft) guarantees freedom from undesirable interferences, while labor is emancipatory because it frees human beings from the power of nature.

Another concern is connected to the abuse of Big Data and Artificial Intelligence technologies by authoritarian states for purposes of surveillance and control. But this is just Orwell's scenario. What about Huxley's scenarios – where technologies are used to keep people busy, entertained or even distracted from the gradual deterioration of their rights and freedoms? And what if the states' roles are taken over by private companies, who own the technologies of the present and future? So, with this in mind, would a Universal Basic Income based on users' data

---

[32] The idea of a Digital Universal Basic Income is based on the presentation "Grounding DUBI", by Anda Zahiu, Cristina Voinea, Radu Uszkai, Constantin Vică, at the conference „Etică, tehnologie și imaginație sociologică într-o lume interconectată", 27 June 2019, at the University of Bucharest.

[33] Imanol Arrieta-Ibarra and others, "Should We Treat Data as Labor? Moving beyond 'Free'", in *AEA Papers and Proceedings*, 2018, CVIII, pp. 38–42.

[34] Jaron Lanier, *Who Owns the Future?*, London, Simon and Schuster, 2014.

[35] Eric A. Posner and E. Glen Weyl, *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*, Princeton, Princeton University Press, 2018.

[36] In order to spread any doubt, this use of the term 'freedom' is slightly different from its Kantian use. Still, Kant is an 'emancipationist', for him the work of reason and understanding/ intellect are involved in acquiring the principle and the idea of freedom.

be a means of liberating people or of further subjugating them? Another issue is connected to inequality, be it economical or epistemic: if AI systems are proprietary, how will those that only own their data fare? And what type of power will the technological savvy and experts hold over the others? Every technical solution brings about new moral questions, because technologies are never neutral.

LOVE IN THE TIME OF ROBOTS

Unlike smart devices, social robots are supposed to interact with human beings on various levels of complexity and, just like any autonomous entity, they can initiate or terminate interactions as they see fit[37]. The agency of robots raises a host of new issues, especially in the realm of personal relationships between artificial agents and human beings. For example, sex robots are no longer a fiction. Some people argue that engaging in intercourse with a robot is blamable form a moral point of view, so long as it does not provoke a harm[38]. Some have even argued that using robots for sex could be a form of therapy for those that break the law[39], or for people affected by various medical conditions, such as dementia[40]. But things are different when the discussion moves towards having a romantic relationship – with or without erotism – with a robot with Artificial Intelligence. This type of personal relationship with an artificial agent could be existentially risky, especially for humans, given that the partners are not equal from an ontological point of view. But there are several ways of looking at such relationships and, implicitly, there are several layers of moral analysis:

1. From a robocentric perspective: when and under what conditions is it morally permissible to treat robots as mere instruments/objects we can use in order to achieve our goals? And what are the conditions that robots equipped with Artificial Intelligence must satisfy in order to be treated as equals to human beings?
2. From an individual perspective: will sexual and personal relationships with robots produce more harm than good? Will such relationships result in the dehumanization of human beings? Or will they contribute to the enlargement of the sphere of entities worthy of moral consideration?
3. From a social perspective: what will be the impact of robots on social and political relationships between humans? Will the introduction of robots open new

---

[37] Matthias J. Scheutz, “The Inherent Dangers of Unidirectional Emotional Bonds between Humans and Social Robots”, in *Robot Ethics: The Ethical and Social Implications of Robotics*, eds. Patrick Lin, Keith Abney, George A. Bekey, Cambridge, Massachusetts, 2009, pp. 205–222.

[38] John P. Sullins, “Robots, Love, and Sex: The Ethics of Building a Love Machine”, in *IEEE Transactions on Affective Computing*, Vol. 3, No. 4, 2012, pp. 398–409.

[39] Kate Devlin, “In Defence of Sex Machines: Why Trying to Ban Sex Robots Is Wrong”, in *The Conversation*, 2015.

[40] Oliver Bendel, “Surgical, Therapeutic, Nursing and Sex Robots in Machine and Information Ethics”, in *Machine Medical Ethics*, van Rysewyk, Simon Peter, Pontier, Matthijs (Eds.), Springer, 2015, pp. 17–32.

avenues for abuse or will it contribute to a decrease in the rate of some moral wrongs, such as the objectification of women and sexism?

In fact, there are two different questions involved: can a human love a robot? And can a robot love a human being? Personal relationships, such as friendship and love, are based on reciprocity and equality. While current robots equipped with Artificial Narrow Intelligence cannot reciprocate on an emotional level, it is still unclear whether Artificial General Intelligence will be able to do so. As such, this is largely an empirical question. But in both scenarios, the thorniest question refers to equality: could personal relationships between entities which are not equal from an ontological point of view exist? Does love always presuppose freedom and equality? What if the attraction towards a robot is not the sign of desire for an artificial body, but an attempt to dominate and control other entities with inferior or even equal cognitive and sensory capacities? Should robots be programmed with the capacity to consent? Still, from a more down to earth perspective, the biggest dangers are the development of unidirectional relationships with robots[41], akin to the relationships we develop with deities or imaginary friends. The danger inherent in these unidirectional relationships – where humans invest emotions and feelings, but the robot can't reciprocate – is that they create psychological dependencies that could be exploited. For example, a robot might attempt to convince a person to buy a certain product or to do a certain thing that its manufacturer programed it to promote[42]. So how do we make sure that love robots won't be just another type of technology, beneficial in some respects, but manipulative in others? The response to such a question might refer back to what keeps human beings from constantly manipulating and deceiving others.

CONCLUSIONS

Even if technology ameliorates some aspects of our lives or radically transforms them for the better, there is a dark side to this moon too. The problem of power inherent in technological scenarios has rarely been discussed and, implicitly, issues of justice, fairness or fair distribution of resources are hardly ever addressed. Moral questions connected to technological design and use have gained importance in the past few years, as technologies are interpreted not just as simple functional tools, but as enablers of the good life and good societies, and promoters of well-being, even *eudaimonia*[43]. Still, these debates have mostly been retrospective – meaning that they focus on existing technologies and on mitigating the harms they do now while the technology industry refuses to acknowledge the part it plays in

---

[41] Scheutz, *op.cit.*, p. 216.

[42] Idem.

[43] Cristina Voinea, "Designing for conviviality", in *Technology in Society*, Vol. 52, 2018, pp. 70–78.

shaping the moral world of tomorrow. Getting excited about the latest technological innovations that allow us to go beyond our cognitive and physical limits is understandable as is being optimistic about the future. But when we project and think about the future, and especially when we act to actualize it, we are actually making moral decisions. The ideology of inevitabilism, coupled with the aura of neutrality surrounding our technologies, contribute to obscuring this point. But technologies are never neutral, they embed values and have mediating power[44]. They mediate our perceptions, enhancing what human beings can basically sense – just as in the case of the microscope or seeing glasses; we can learn to see the world through them, and thus we have to learn how to 'read' them; we can interact with/through them and from this point of view technology becomes alterity; and last, but not least, they create the background for human experience[45] – just like the Internet does. As such, technologies create new affordances and constraints which reiterate power structures.

Choosing a particular technology is, implicitly, choosing a certain set of moral values or principles. It is a choice about what we want the world to look like, who we want to gain more power, wealth or status; and, of course, it is also a choice about who will stand to lose from that particular arrangement. Unfortunately, all of these existential choices for the future are obscured either by technological neutrality and inevitabilism, or by very implausible but appealing future scenarios involving evil robots and AIs. Our claim is that thinking and acting for the future always presuppose a moral dimension, even if the future is populated by informational and algorithmic technologies, seemingly neutral artifacts. The Artificial Intelligence research agenda should be driven by the moral duty we have towards the others, be they part of the present or next generations, which is a duty to enhance their lives and to make it possible for them to choose and to be free.

Acknowledgements: This research was funded through the project "Longitudinal analysis of coauthorship networks and citations in academia" (iCoNiC), funded by the Executive Unit for Financing Higher Education, Research, Development and Innovation (UEFISCDI) - code: PN-III-P1-1.1-TE-2016-0362.

---

[44] Don Ihde, *Technology and the Lifeworld: From Garden to Earth*, Indiana University Press, 1990; Albert Borgmann, *Technology and the Character of Contemporary Life: A Philosophical Inquiry*, Chicago, University of Chicago Press, 1987.

[45] Idem.