Ciências & Cognição 2018; Vol 23(2) 237-242
ISSN 1806-5821 - <http://www.cienciasecognicao.org/revista>
Submetido em 12/09/18 | Revisado em 05/11/18 | Aceito em: 15/11/18 | Publicado em 31/12/18

© Ciências & Cognição

# The brain and the robot: bioethical implications in transhumanism

*O cérebro e o robô: implicações bioéticas no transhumanismo*

**Elen Cristina Carvalho Nascimento [1], Rodrigo Siqueira-Batista [1,2]**

*[1] Postgraduate Program in Bioethics, Applied Ethics and Collective Health (PPGBIOS), Federal University of Rio de Janeiro /UFRJ. [2] Laboratory of Epidemiological and Computational Methods in Health (LMECS), Federal University of Viçosa/UFV (Professor Associado); Course of Medicine, Dynamic College of Piranga Valley/ FADIP (Professor Titular).*

**Abstract**
This paper is a critical review of the book "O cérebro e o robô: inteligência artificial, biotecnologia e a nova ética" ("The brain and the robot: artificial intelligence, biotechnology and the new ethics"), by João de Fernandes Teixeira. Publisher: Editora Paulus, 2016.
**Keywords**: brain; artificial intelligence; bioethics.

***Resumo**: O presente texto é uma resenha crítica do livro "O cérebro e o robô: inteligência artificial, biotecnologia e a nova ética", de João de Fernandes Teixeira. Editora Paulus, 2016.*
***Palavras-chave**: cérebro, inteligência artificial, bioética.*

To think of a new ethic for the rampant advancement of knowledge has been the task of many authors in the twentieth and twenty-first centuries, in a context in which there is a marked "impregnation" of contemporary culture by technology. The challenge is to propose references capable of broadening the dialogue between the two spheres – ethic and technoscientific, perhaps as a bet for the construction of a fairer society.

It is in this scenario that the book, "The brain and the robot: artificial intelligence, biotechnology and the new ethics" by João de Fernandes Teixeira, announces a motivation to deal with the theme in the first pages: "The optimistic and comforting certainty that the expansion of technology will always bring benefits to humanity no longer exists" (p.13). The book is organized in an Introduction, a Conclusion and six chapters – 'The Window of Descartes',' Technology and Abyss', 'The Phantom of Singularity', 'The Enigma of Meaning', 'Human, Too Transhuman' and 'Brains in the test tube' – , which address, in different tones, the interrelationships between ethics, biotechnology, neurosciences and artificial intelligence (AI). In this way, to think of technology as a creative becoming, a reflection of the human need to overcome obstacles, is to think, as Fernandes says, that "a life story can be reconstructed in several ways; and often reinventing it is a good strategy against anguish" (p.125).

In the first chapter, *Descartes's window*, Fernandes will articulate aspects between technological revolutions and the need to think ethics in this context. Fears and expectations

Ciências & Cognição 2018; Vol 23(2) 237-242 © Ciências & Cognição
ISSN 1806-5821 - <http://www.cienciasecognicao.org/revista>
Submetido em 12/09/18 | Revisado em 05/11/18 | Aceito em: 15/11/18 | Publicado em 31/12/18

regarding technology and its development are approached from a reflection on Descartes' view, in which a parallel is drawn between the philosopher's statement – that a machine, an automaton, will never have a soul, or it will not be able to acquire consciousness – and the Turing Test, in which a machine, through AI, is expected to create the illusion that there is consciousness. As for the AI to develop consciousness of fact, there is a vacuum, since everything that is understood by subjectivity would have to be quantified and computed.

Neurosciences have contributed to this field of research by studying the brain through different methods, such as EEG (electroencephalogram) and fMRI (functional magnetic resonance imaging). In this context, this field has been investigating the impact that emotional response has on the quality of learning, through observation, in descriptions such as those related to mirror neurons[1]. Thus, the search for answers to improve language and decision-making in an AI has been pushing the studies on the brain, dismantling old notions of the mind with functions organized in separate parts – a localizationist perspective – in favor of a system of emotions, which has important effects on health. In fact, the body is not exactly a machine in which each of its senses, functions, and qualities come with a stored "chip" of knowledge necessary for its functioning. The study of vision, for example, showed a few decades ago, that our ability to see is not only physiological qualities, but we need the experience of seeing through the semantic values we give to what we see (Kandel, Schwartz & Jessel, 2014). The image is formed through a set of perceptions where experience, observation, language, territory, and culture are determining.

In *Technology and Abyss*, Chapter 2, the author suggests that there is a tendency of "neoludism" in those who see the need to eradicate technology. However, it is precisely the creative becoming of technology, the anguish transformed into objects, projects and solutions, that make it not an entity external to the human, but a reflection of its own condition[2]. They are creations that gain identity, own citizenship. They are not simply "things" that can be replaced. They are knowledge and experiences, and their formation process – gestation – condenses energies and information flows. This is a view advocated by the philosopher Gilbert Simondon, quoted by Fernandes, to discuss the relevance of "improvisation" in science to the expected results of an evolution. The improvisation, the improbable, the unknown elements over which there is no control. Non-linear history is reflected in ancient artifacts that continue to be useful side by side with other high-tech artifacts. This is the case with the hammer, the watch and the bicycle. Fernandes also states that Heidegger, in his essay "The Question of Technique", predicted an apocalyptic future for human technology (p. 54)[3]. According to the author, for Heidegger, mathematics is "not only a calculation but a way of interrogating nature, interpreting it" (p.55). What Heidegger seeks to show is that the technique is not really neutral. Thus, from Heidegger to a leap into the present, we can conclude that the world impregnated with binary mathematics is the "one that increasingly rehearses its self-destruction" (p. 56).

In Chapter 3, *The Phantom of Singularity*, he walks on slippery ground: "To what extent these new beings do not threaten the identity of the human species?" (p. 73). To understand such a questioning, should we not ask ourselves what identity we are talking about, how, and, on what precepts does it define itself? When Fernandes presents a new era of "disenchantment of self" (p. 75), would not this be a new age of disenchantment with the morality proposed by the Enlightenment? That is, a "checkmate" on the foundations of humanism, in a game of growing crises of the reason, that was fomented by internal conflicts, still unresolved, in the sphere of scientific knowledge? In this chapter, the author

Ciências & Cognição 2018; Vol 23(2) 237-242 © Ciências & Cognição
ISSN 1806-5821 - <http://www.cienciasecognicao.org/revista>
Submetido em 12/09/18 | Revisado em 05/11/18 | Aceito em: 15/11/18 | Publicado em 31/12/18

gives voice to transhumanist conjectures. Transhumanism is a position that, in theory, pretends to be an alternative to the premise that the human condition is unalterable. However, transhumanism develops future projections that are, many of them, questionable. Such projections involve perspectives of superintelligent machines, human enhancement through biotechnology features and colonization of space, among a list that involves developments in nanotechnology, interconnectivity, and AI.

At the end of the 1980's, the scenario of computational ubiquity was already projected. In the late 1990's, large companies such as HP and IBM sought to foresee a future of the "internet of things," in which it would be possible, for example, to provide relief for an elderly person who was living alone five minutes after fainting[4]. In the period when such predictions were made, there was still a long way for network expansion and nanochips development. Today we have the technology for this, but if two decades ago such a possibility could seem spectacular and welcome, the facts demonstrate that there are security issues that jeopardize the effectiveness of such applications, insofar as permitting vital data of a patient traveling on the network can leave him, or her, vulnerable to interference and invasion, other than those related to the assistance. Technology, therefore, reflects human relationships and their conflicts. Close security doors, architect layers of firewalls and other measures have not solved the problem of how to advance in idealized progress without considering the need for collaboration between all parts of this body, this structure we call progress.

In Chapter 4, *The Enigma of Meaning*, the author discusses the problem of language in the development of an AI: "Language is a symbolic representation of what mental states represent, it is a representation of a representation, which, in no way resembles the binary code of 0's and 1's of digital computers" (page 96). Thus, Fernandes rethinks the Turing Test from the "Imitation Game" [5], as, like the "Chinese Room" argument of John Searle, are experiments that demonstrate that the computational models are learned from patterns of information and repetitions by codes. The argument of the Chinese Room, in turn, constitutes a criticism of what is to be called AI, since for Searle an intelligence without consciousness is not possible. In this sense, Fernandes concludes: "The basis of intentionality and consciousness is life. Without a living brain, they would not be possible" (97). Authors like Paula Sibilia will amplify this perception with the statement: "The brain exists in the body, and the body exists in the world" (Sibilia, 2014). There are conclusions that are based on the foundation that the processes of mind, such as rationalization, for example, are totally dependent on corporeal sensory experience, as demonstrated by Damásio (2003). With the science of such facts, to believe that the human mind can be replicated in machines to be able to respond to unpredictable situations by making use of the symbolic and creative exercise of language, becomes a challenge with small chances of obtaining satisfactory results. Making life a computable phenomenon does not mean being able to reliably imitate it. The point at the end of this chapter is that if humans sought to adapt more and more to a mechanical worldview, which is based on the presuppositions of modern scientific rationality, the copy of the human is a copy of what would be the idealization of the human, as a machine.

Thus, if the model that generates binary computation advances in the artificial replication of vital processes, resulting in new discoveries that inspire transhumanist ambitions, in Chapter 5 – "Human, Too Transhuman" – we are talking about the development of minds without a biological brain, rather, without a body. One can even

Ciências & Cognição 2018; Vol 23(2) 237-242 © Ciências & Cognição
ISSN 1806-5821 - <http://www.cienciasecognicao.org/revista>
Submetido em 12/09/18 | Revisado em 05/11/18 | Aceito em: 15/11/18 | Publicado em 31/12/18

assign an "advantage" of these machines within human conviviality, as "ethical agents" capable of discerning better, that is, without the intervention of emotions and any possible malaise coming from them. The concept is presented by Fernandes in citing the philosopher Nick Bostrom, according to him, one of the greatest contemporary transhumanists who "supports that superintelligent machines can be ethical agents superior to human beings" (p. 112). If we consider what Antonio Damasio (Damásio, 2003) tries to show – those brain phenomena, alone, do not explain the mind, taking into consideration the physical and social environment – we are facing a conflict that is prior to technological developments, its benefits, dangers and ambitions (Esperidião-Antônio et al., 2017). It is a mechanistic view we use as method, to explain existence, which isolates parts and functions, separates nature and culture, treats the body as an integrated organism, but independent of its environment.

Such a view – which for Damásio (2003) is a Cartesian view of the human condition, it's focused on the physiology and the pathology –, and that tends to break empathy, generating less respect for life. For him, the mechanistic, Cartesian view, in which the vital processes are regarded as the mechanics of a clock, separates the thinking thing, the *res cogitans*, from a non-thinking body, the *res extensa*, which is nothing more than a geometric extension, subtended as a minor. This set, within the imagery of caring for the other, assists, according to him, a break of empathy and generates distance. This argument is the result of research that reveals the inability of certain attributes of the mind, or of the body, to have an independent constitution of the senses.

Within this understanding, the discussions proposed by Fernandes in this chapter are about to think devices, in the programming of AIs, that could protect humans from the possible damages the machines could cause them. The question is how to design ethics by thinking in our relations with these machines, and even, if emotions are simulated in the machines, as well as making them similar to humans, that may include them in the sphere of moral responsibility. However, such questions tend to be empty, since there are previous problems of epistemological order. If the transhumanists quoted by Fernandes announce the "reformation in human nature" in a "World Declaration" (p. 120), they are possibly referring to a "human nature" that still needs to best suit a mode of being machine, a plan of Modern Science that was inconclusive. On that sense, can we interpret that genetic interventions and manipulations (p. 122), drugs to enhance cognition and memory (pp. 125-126) are only new devices to make bodies docile, according to Foucault's concept of biopower?

In *Brains in the Test Tube*, Chapter 6, the author introduces the concept of "zombie robots", which would be humans that replace neurons with electronic chips, and that would tend to gradually reduce consciousness. Fernandes wonders if replacing elements in a person's brain would not make them stop being who they are. On the other hand, if people often have their behaviours adjusted and transformed by culture and social demands, what would be the difference in the case of a targeted intervention? A differential presented would be the concept of qualia, which constitutes the human sensory experience that intervenes in its general biological constitution [6]. This is a process that has not yet been unveiled, just as the ethical dimension about what technologies are indeed acceptable and desirable, constitute an open debate. When one tries to think about such phenomena using the (bio)ethics references, it is also considered that a new ethic should shift the Anthropos from the center to a wider scenario where the human is part of a whole, and where it is necessary to attribute moral value to other beings, and not only to humans.

Ciências & Cognição 2018; Vol 23(2) 237-242 © Ciências & Cognição
ISSN 1806-5821 - <http://www.cienciasecognicao.org/revista>
Submetido em 12/09/18 | Revisado em 05/11/18 | Aceito em: 15/11/18 | Publicado em 31/12/18

The chapter *Conclusion* presents the questioning of whether technology is not only self-replicating and generating a loss of the meaning of life, without actually bringing benefits to human life. Would that also be an anthropocentric view? Our argument suggests that to elect technology – or virus or disease – as villains, keeps the human at the Centre, as an independent organ of the whole, frightened and vulnerable. What are its responsibilities? Is not the human, part of an integrated system of life that goes beyond the limits of its own body? Fernandes suggests that the "narcissistic projection of the improved man" is presented as a "solution to the dead-end we think we are in" (p.150).

So, when speaking of AI as a "superior" intelligence, it seems that we still need to confront the origins of our beliefs and how much humanism it is necessary to preserve, and how much it is necessary to leave in the past, as a social and cultural phenomenon of a historical period. Thus, it is possible that if the transhumanist project has many misconceptions, they may be on the same foundations of those who attack them, with bioconservative arguments.

The philosopher Rose Braidotti (2013) presents a third way to think of a *posthumanism*, which would be a process of becoming, a constant search of a critical thinking, after the shock and recognition of the uncertainty that technological revolutions have been emphasizing[7]. For her, embedding the ethical precepts that are important to the community in technology, may be the opportunity to rebuild a collective sense that is defined by affinities and that can be accountable, based on a new social contract. However, what is still not decided among us, will not make computers decide better. If an autonomous vehicle needs to choose, in case of an inevitable accident, between killing ten people or one, it seems that it would be better if people were simply using bicycles to get around. A drone flying over war territory will be able to know more accurately where there are more civilians, women, children, and the elderly, and thus not decide to launch a bomb. But this does not eliminate the obvious paradox of wanting to embed ethics in a drone scheduled to launch bombs.

As long as the transhumanist project continues to be governed by the desire to "control destiny, to restrict the enormous range of possibilities contained in the data game of the future, by directing the options, especially in the sense of prolonging life and annulling finitude" (Sibilia, 2014), we are still trapped in anthropocentrism and their mistakes. In that sense, it would not be exactly the sphere of technology which is capable of extinguishing humans, but their own desires, beliefs and inconsistencies.

**Funding Statement**

This work was supported by CAPES (Higher Education Coordination of Brazilian Government) and CNPq (National Council for Scientific and Technological Development).

**References**

Braidotti, R. (2013). *The Posthuman*. Cambridge, Massachussets: Polity Press.

Damásio, A. (2003). *O erro de Descartes: emoção, razão e o cérebro humano*. São Paulo: Companhia das Letras.

**Ciências & Cognição 2018; Vol 23(2) 237-242** **© Ciências & Cognição**
**ISSN 1806-5821 -** <http://www.cienciasecognicao.org/revista>
Submetido em 12/09/18 | Revisado em 05/11/18 | Aceito em: 15/11/18 | Publicado em 31/12/18

Esperidião-Antônio, V., Majeski-Colombo, M., Toledo-Monteverde, D., Moraes-Martins, G., Fernandes, J. J., Assis, M. B. D., Montenegro, S. & Siqueira-Batista, R. (2017). Neurobiology of emotions: an update. *International Review of Psychiatry*, v. 29, p. 293-307.

Kandel, E. R., Schwartz, J.H., Jessel, T. M. (2014). 'Processamento Visual de Nível Superior: Influências Cognitivas'. In: *Princípios de Neurociências*. Trad. 5ª. Edição. Porto Alegre: AMGH, p. 539-553.

Sibilia, P. (2014). *O homem pós-orgânico: A alquimia dos corpos e das almas à luz das tecnologias digitais.* Rio de Janeiro: Contraponto.

Teixeira, J. de F. (2016). *O cérebro e o robô: inteligência artificial, biotecnologia e a nova ética*. São Paulo: Paulus.

**Notes**

(1) For a brief understanding of “mirror neurons”, please check Wiedermann, J. (2012). Mirror neurons, embodied cognitive agents and imitation learning. *Computing and Informatics*, *22*(6), 545–559.

(2) This concept is well developed by Verbeek, P.-P. (2015). “Cover story Beyond interaction: a short introduction to mediation theory.” *Interactions 22.3*, 26–31.

(3) In fact, Heidegger, in the essay, states that “Technique is not dangerous. There is no demonic technique; on the contrary, there is the mystery of its essence. The essence of technique, as a destiny of desolation, is the danger.” Heidegger, M. (2007). The question of technics. *Scientiae Studia*, *5*(3), 375–398.

(4) Video: HP. (2000). Cool Town. YouTube. Retrieved from https://youtu.be/U2AkkuIVV-I

(5) “The imitation game” is a model created by Alan Turing and became the title of the film by Morten Tyldum about the mathematician, released in 2015.

(6) Fernandes quotes, on page 142, the neuroscientist Ramachandram, who elucidates issues related to the *qualia* concept at: Ramachandran, V. S., & Hirstein, W. (1997). Three laws of qualia: What neurology tells us about the biological functions of consciousness. *Journal of consciousness studies*, *4* (5-6), 429-457.