



ETHICAL EDUCATION AND ARTIFICIAL INTELLIGENCE

^{1,2,3}Oswaldo Jesus Rodrigues da Motta, ^{4,5}Eugênio Silva and ^{2,3}Rodrigo Siqueira-Batista

¹School of Medicine, Faculdade Dinâmica do Vale do Piranga (FADIP), Ponte Nova – MG – Brazil; ² Department of Medicine and Nursing, Universidade Federal de Viçosa (UFV), Viçosa, MG – Brazil; ³ Department of Public Health Sciences and Pediatrics, Università degli Studi di Torino (UNITO), Torino – Italy; ⁴ Computing Unit, Fundação Centro Universitário Estadual da Zona Oeste (UEZO), Rio de Janeiro, RJ – Brazil; ⁵School of Computer Science, Centro Universitário Serra dos Órgãos (UNIFESO), Teresópolis, RJ – Brazil.

ARTICLE INFO

Article History:

Received 20th September, 2020
Received in revised form
28th October, 2020
Accepted 10th November, 2020
Published online 31st December, 2020

Key Words:

Ethics, Ethic Education, Artificial Intelligence.

*Corresponding author:

Oswaldo Jesus Rodrigues da Motta.

ABSTRACT

It is considered that any teaching-learning activity must be ethically conducted, and for that, efforts are necessary to create an environment conducive to the assimilation of the concepts, theories and methods considered indispensable for good coexistence in a democratic, secular and plural society. This perspective is extremely important, as current research aimed at machine learning tries to build intelligent systems, without prejudice, preventing them, for example, from behaving in a racist, sexist or speciesist manner. The discussion of these questions is the aim of this essay.

Copyright©2020, Oswaldo Jesus Rodrigues da Motta et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Citation: Oswaldo Jesus Rodrigues da Motta, 2020. "Ethical education and artificial intelligence", *International Journal of Development Research*, 10, (12), 43111-43112.

INTRODUCTION

The current developments in Artificial Intelligence (AI), especially in the scope of decision support, have revealed the importance of considering ethical aspects in machine learning processes (Dennis, 2020; Siqueira-Batista and Silva, 2019). Such elements gain special relevance when recognizing that ethical decision-making is, most likely, one of the most challenging tasks in the cognitive realm of *Homo sapiens*, which makes its "transposition", for devices endowed with AI, an endeavor of the greatest difficulty (Motta et al., 2020a; Cervantes et al., 2020). In fact, in view of the prospect of developing autonomous artificial agents capable of ethical judgment — perhaps of legitimate choices — the following question arises: *would it be possible to teach machines ethics?* (Motta et al., 2020b). The question now becomes even more problematic when taking a step back and considering the problem of teaching ethics to humans. For this, returning to the very dawn of philosophical reflection in the West becomes necessary. In these terms, the possibility of teaching *arete* is discussed in the Platonic dialogue of Meno (PLATÃO, 2011). In the dialogue, from the beginning, Socrates replies to Meno who does not know whether *arete* can be taught or not and,

continually, redirects the conversation to the following question: *what is arete?*

Arete, according to ancient Greeks, was understood as excellence; the passage to Latin, as *virtus*, term directed to the most usual translation in contemporary times: virtue. The articulation *arete* ↔ *virtus* ↔ *virtue* allows for the interesting exercise of reorienting the Platonic question to the ethical sphere: is it possible to teach it (ethics)?

It is a question that spans throughout centuries and remains unanswered until the present day. In fact, it is not known whether it is possible to teach ethics, but there is some conviction in affirming that learning ethics is possible. Indeed, ethical learning, within the scope of moral education, moves not only along the paths of the educational process, but also through the experiences of individuals, based on the fundamentals that must prioritize autonomous, critical and reflective training, due to coexistence in secular and contemporary education. The task of teaching ethics, if considered possible, is — clearly — not an easy one, especially in the current context of moral obscurantism, in which anti-democratic and fascist ideals flourish on social media.

When applying such a discussion to the scope of AI, so that ethical content is assimilated by machines, science has recently focused on the following statement: allow them to learn to be independent, that is, to be able to make decisions based on a set of data, considering the other and the laws of society. However, how is it possible that they, the machines, learn ethics? In March 2016, Microsoft launched a bot — a software for simulating human activities — named Tay, which, thanks to AI techniques, was able to talk to people. However, very quickly the bot learned exactly what is unethical: it started to dialogue based on prejudices, opting for its deactivation on the same day it was put into action.

On the other hand, researchers at the University of Technology in Darmstadt, Germany, provided data to an AI-based system, extracted from books, news and religious literature so that the machine learned the associations between different words and phrases. This time the system assimilated moral values instructed after training, corroborating for the understanding and analysis of the context of sentences and not word for word. As a result, the AI system understood that killing time is not necessarily a problem, but killing living things is highly reprehensible. (PATRICK S.*et al.*, 2020) Therefore, the machine would show a more adequate judgment, in relation to right and wrong, when analyzing content produced by users.

It is considered that any teaching-learning activity must be ethically conducted, and for that, efforts are necessary to create an environment conducive to the assimilation of the concepts, theories and methods considered indispensable for good coexistence in a democratic, secular and plural society. This perspective is extremely important, as current research aimed at machine learning tries to build intelligent systems, without prejudice, preventing them, for example, from behaving in a racist, sexist or speciesist manner.

Thus, only time and technological evolution will be able to answer whether there is a possibility of creating machines with AI and that are capable of ethical, rational and impartial assessments, perhaps surpassing — in moral terms — humans themselves.

REFERENCES

- Cervantes, J. A.; López, S.; Rodríguez, I. F.; Cervantes, S.; Cervantes, F.; Ramos, F. 2020. Artificial Moral Agents: A Survey of the Current Status. *Science and Engineering Ethics*, v. 26, n. 2, p. 501-532.
- Dennis, L. A. 2020. Computational Goals, Values and Decision-Making. *Science and Engineering Ethics*, v. 26, n. 5, p. 2487-2495.
- Motta, O. J. R.; Silva, E.; Siqueira-Batista, R. 2020a. Artificial intelligence & COVID-19: (bio)ethical aspects of end of life. *Revista da Associação Médica Brasileira*, v. 66, p. 5-6.
- Motta, O. J. R.; Silva, E.; Siqueira-Batista, R. 2020b. “Ensinar” ética às máquinas dotadas de inteligência artificial. *Jornal da Ciência*. Available in: <<http://www.jornaldaciencia.org.br/edicoes/?url=http://jcnovicias.jornaldaciencia.org.br/30-ensinar-etica-as-maquinas-dotadas-de-inteligencia-artificial/>> Access in: Dec 27.
- Platão. 2001. *Mênon*. Texto estabelecido e anotado por John Burnet; tradução de Maura Iglésias, Rio de Janeiro; Ed. PUC-Rio; Loyola, 2001.
- Patrick, S.; Cigdem, T.; Sophie, J.; Constantin, R.; Kristian, K. 2020. The Moral Choice Machine. *Frontiers in Artificial Intelligence*, v.3, p. 01-15.
- Siqueira-Batista, R.; Silva, E. 2019. Notas sobre os fundamentos matemáticos da Inteligência Artificial. *RevistadeCiência, TecnologiaeInovação*, v. 4, p. 44-54.
