

# Consciousness of intelligent machine entities: a critique of the Philosophy of Technology

Mariah Brochado

**ABSTRACT:** In the last century we never used expressions in our everyday jargon like ‘thinking machines’, ‘intelligent entities’, ‘conscious machines’, ‘machine responsibility’. The use of these expressions in the 21st century has been as new as it has been imprecise. The question that arises in this regard is: what is the reference adopted by computer scientists to attribute to machines faculties that have always been attributed exclusively to humans? The research program in this article proposes a reflection on the adoption of human adjectives for machinic reality, bringing in counterpoints from philosophical analysis in this regard. The proposed starting point is a dialogical imbrication between epistemology and ethics to discuss the unnecessary and dangerous analogies established between humans and machines, from the perspective of the philosophy of technology. This is an important issue to bring up for debate about the so-called ‘technological age’, in the context of social science reflections on the challenges of our century.

**KEYWORDS:** consciousness, philosophy of technology, human-machine

## I. INTRODUCTION

The prodigious evolution that our civilization has undergone - already known as the "Civilization of Technique" - imposes on the Human Sciences reflections on topics that have never been considered, either in the sciences or in the field of technology, because they have never been put in the terms in which they are announced today.

In this first quarter of the century a new symbology comes to inhabit the imaginary of our civilization: we are talking about autonomous entities, independent artificial entities, conjecturing about the responsibilities that might fall on such entities. Artificial Intelligence, an area of Cognitive Computing, has come to occupy special prominence in these discussions. The expression ‘intelligence’ here refers to machinic performance, not to what we conceive of as the rational capacity of humans. Roughly speaking, it's a machine simulation of the constitutive elements of human cognition.

This achievement ends up triggering a hypothesis, which had already been considered by Turing and continues to cause a stir: can machines think? More: this is associated with the expectation that machines with high cognitive performance will be able to autonomize themselves independently of what we call intentionality, an exclusive characteristic of humans. In this article we propose a reflection on the concept of consciousness in technological jargon and in philosophical jargon, criticizing it from the perspective of the Philosophy of Technology, as well as the indiscriminate use of human qualifiers in machines, unnecessary use of this for media purposes.

## II. PHENOMENAL CONSCIOUSNESS AND FUNCTIONAL CONSCIOUSNESS

I am introducing this article with a provocation from Blay Whitby, considered one of the current references in Philosophy of Technology, with extensive experience in computing and robotics, professor at the University of Sussex, traditionally one of the most fruitful centers of research in technology in the world. Whitby writes when introducing the theme of artificial consciousness: "Some years ago, my duties at a British university included turning off the computers in the AI lab. There was one machine that, whenever I went to do so, would scream, 'No! Don't turn me off, I have become conscious'-I never had a moment's hesitation to turn that machine off, certain that what I was hearing was a consequence of student humor and not progress in AI". [1]

This passage reveals much about the whole futurological zeal to assert an artificial consciousness, which finds increasing echoes in the media, and Whitby himself is a researcher quite concerned with presenting consistent data in Ethics applied to technology.

The conjectures about the possibility of acquisition of conscious knowledge by a machine in Artificial Intelligence start from a distinction that impoverishes the entire theoretical path of the philosophy of Western consciousness: AI scientists distinguish phenomenal consciousness from functional consciousness. The latter is the consciousness referred to intelligent machines; the former, with its intrinsic characteristics, is human consciousness, capable of the experience called qualia<sup>2</sup> [2].

In fact, the reference to a "functional consciousness" starts from the assumption that it is an entity destined to exercise functions, to execute tasks, not being what we of philosophy understand by consciousness in any sense, be it intellectual, psychic or moral, since we are self-referential beings in a thinking unit that knows itself by thinking (I think, in Kant), which has nothing to do with functionality, a characteristic attributed, in the Western philosophical tradition, to objects, not subjects.

On the other hand, the fragility of the philosophical approach to the subject of the consciousness is pointed out by the sciences of the brain (empirical sciences that submits phenomena to verification) argument that it is an undemonstrable phenomenon, which has always been referred to a hidden space of mystery. This is a typical philosophical analysis: it always tries to link highly complex phenomena to transcendence, as a constitutive attribute of human existence, when there is no satisfactory scientific support to explain such phenomena.

Not without purpose, this is the title that John Searle gave to one of his main works: The Mystery of Consciousness. In fact, as Whitby provokes, human beings seem to have a predilection for mystery and tend "to like a little mystery in their stories about the world and particularly like a little mystery in the way they make their dearest decisions and come up with their best ideas" [3]

Against this status of mysterious data are the researchers of Cognitive Computing, but contrary to what we might expect, the "such mystery" is not being unveiled, but rather, disregarded. Certainly, when "artificial consciousness" is mentioned, it is not intended to present an artificially simulated hypothesis of the functioning of human consciousness<sup>3</sup> [4] to reveal its intricacies, and therefore, its mysteries.

We note, first of all, that the weak use of the word conscience, in this techno context, does not intend to go deeper into the traditional discussions of Ethics since Socrates about the ontology of the human conscience, which starts from the distinction between the species sapiens and all the other terrestrial species, since sapiens is the only living being endowed with conscience - even though there are animal practices of construction by instinct that resemble conscious practices. In view of this, Henri Bergson poses the question:

[...] does an unintelligent animal also possess tools or machines? Yes, certainly, but

here the instrument forms a part of the body that uses it; and, corresponding to this instrument, there is an instinct that knows how to use it. True, it cannot be maintained that all instincts consist in a natural ability to use an inborn mechanism. Instinct is therefore necessarily specialized, being nothing but the utilization of a specific instrument for a specific object. The instrument constructed intelligently, on the contrary, is an imperfect instrument. It costs an effort. It is generally troublesome to handle. But, as it is made of unorganized matter, it can take any form whatsoever, serve any purpose, free the living being from every new difficulty that arises and bestow on it an unlimited number of powers. Whilst it is inferior to the natural instrument for the satisfaction of immediate wants, its advantage over the greater, the less urgent the need. Above all, it reacts on the nature of the being that constructs it. For incalling on him to exercise a new function, it confers to him a richer organization, being an artificial organ by which the natural organism is extended. [5]

It seems that the most respected computer scientists are aware of this *petitio principii* on the definition of human faculties and their extension to machines, and even denounce the media exploitation of the theme, influenced by the very need for projection of AI researchers, eager for publicity, who use megalomaniac tricks to raise huge amounts of money for their endeavors and the millionaire sale of their products. Boden brings the example of a prototype that "'fooled' psychiatrists into thinking they were reading interviews with paranoiacs - because they naturally assumed they were dealing with human patients" [6]. There are countless examples of "imitation games" using sophisticated computer systems.

### **III. UNNECESSARY ANALOGIES: WORDS STABILISED IN PHILOSOPHIE ARE USED INDISCRIMINATELY**

First of all, we need to understand what these researchers in the field of AI intend when they use the word intelligence: it does not mean, in the usage of the field, exclusively human faculty, but information processing, as seen. When it is said that a machine learns, obviously the word learning cannot be taken as a free activity of humans. For the rest, whenever words like intelligence, learning, consciousness, neural evolution or similar are used metaphorically, we need to delimit in what sense the terms are being employed.

Technosciences currently use these words by analogy with schemes that already work to describe human behavior by the Human Sciences. And these sciences derive this behavior from a structure understood as human nature, a philosophical category that survives time, from the Greeks to contemporary philosophy, as an unfathomable instance that transcends the biological structure of the animal man.

But the attempts to adapt these words by technology aim to adapt certain concepts of traditional philosophy may seek, from an empiricist reading, to extend phenomena attributable exclusively to human experience to hyperpowered anthropomorphized machines for the purpose of interpreting extremely ingenious and useful technological constructs. Assuming that humans are not defined by their utility, but by their intrinsic and untouchable dignity (Kant's second formulation of the categorical imperative), any analogy that is drawn between human experience and copies of human ways of being in the world in machine format is yet another technical sophistication that we need to understand.

We must try to understand what this new way of influencing human life may bring in the long run as a differentiator in the "shape" of human nature. After all, this way of understanding humanity has remained impervious to the changes in pre-industrial and post-industrial societies and technical revolutions have not significantly touched philosophers' conviction about our essence. Will the machine evolution of this century promote radical changes

in the cognitive and behavioral structures of humans? This is the central question facing the philosophy of technology today, and this question is essentially ethical, not epistemological [7].

In the context of this article, it is important to discuss the notion of intentionality as a reference that gives unity to the "I think" of the epistemic agent and the moral agent (the unity in the notion of person), an interface that does not exist in machine intelligence. This will be the object of investigation of the recently emerged Computational Ethics and Roboethics, attempts to transpose traditional ethical systems (especially the eudaimonistic ethics - of virtues) to the field of Cognitive Computing. This is a difficult composition within the Philosophy of Technology, since it requires relating approaches with different and even excluding *parti pris*, since Cognitive Computing is already born from a psychological reading of human intelligence (and consciousness), while traditional rationalist philosophy assumes an intellectual-moral instance of the human being irreducible to psychological states (who will say to causal brain functions) [8].

This impasse generated between traditional philosophy and philosophy of mind reveals, once again, the difficult assimilation of the program of a philosophy of technology by the rationalist-based philosophy, since the latter does not retreat to seek in the plane of scientific understanding the brain functions in order to legitimize its theoretical status. On the contrary, the movement is in another direction: that of ascending to the plane of reason (inflection) in the form of practical reason, directing the discussion to the field of Ethics.

As we have seen, the computational cognitive theory of Artificial Intelligence employs the word consciousness in two basic senses: phenomenal consciousness, studied by philosophy and the human sciences, and functional consciousness, an expression adopted in the field to refer to computational models considered "machine consciousness" (CM) or, still, to comprehensive analyses of computational systems, independently of addressing specific models, as they explain, without doing any specific functional modeling; that is to say: it is a comprehensive use of the word consciousness to refer to machine-installable computational processes. [9]

By addressing machine consciousness in these terms, computer theorists have in view the discussion about the possibility of developing a general artificial intelligence (GIA) that could be considered truly intelligent to the extent that it exhibits characteristics equivalent to human consciousness, such as the abilities to concentrate, to pay attention, to be aware of different things at different times.

Here the word consciousness indicates only that a machine (virtual or embodied) could have a constitutive self-referential refinement of distinguishing and choosing performances diverse in quantity and complexity; that is: "being aware" is not knowing that it knows; it is knowing how to do, knowing how to function in various ways, without immediate interventions in its programming. From here comes the idea of deep learning. This is where Robotics plays the most important role in machine simulation. It is here that artificial intelligence programs can present themselves as similar to humans by embodying the human shape: humanoid robots.

Today there are discussions about the epistemological autonomy of Robotics: some consider it a subfield of AI; and some already consider it an autonomous science. In any case, it is a technology that aims to design perceptual and motor tasks in order to produce physical and animated behavior. While AI programs generally deal with cognition and operate in computer-simulated worlds, robotics programs aim to physically place their commands in a machine that interacts with the real world, acting on it with perception. This involves complex engineering of sensors (perceptual interface between robots and the environment) and effectors (legs, wheels, joints, etc.) to act physically. [10]

According to Boden: "a human-level system would also be capable of deliberation and self-reflection. It could generate creative ideas and even evaluate them deliberately. Without these capabilities, it would not be able to

generate seemingly intelligent performance" [11].

Since dealing with AI is to deal with appearance, imitation, machine simulation that reproduces human faculties and characteristics, the attempt to copy human consciousness is nothing new. It happens that this imitation is not a superficial one, as is that of doing math and cross-referencing data; it should be an imitation that convinces humans that they are dealing with an entity that acts on its own, that deliberates on whether or not to perform a task, that opposes a programming for which it was predestined.

For the advocates of artificial consciousness, this kind of machine performance is a matter of time, because, just as the natural algorithms that make up the human brain have been "trained" by nature for thousands of years, it will be possible to train thousands of artificial algorithms that allow machines, armed with trillions of data, to behave in the manner of human freedom, which is always the exercise of choices among data arranged in neurons (brain algorithms). If the machine obtains, therefore, an extraordinary amount of information, it will be able to exercise control over it, choosing among the most sophisticated combinations and decisions, as occurs with the human brain machine. It is a question of time, because, as cybernetics has surmised, the control of information is a universal datum that can be replicated in machine organisms.

Boden refers to an apparent intelligence. This is a much-debated question among computer scientists and goes back to the famous Turing Test, which is actually a test to measure how much imitation of human intelligence a computer program achieves in order to fool a human being. In fact, Turing never seriously referred to a test in his famous essay published in the philosophy journal *Mind* in 1950. In it, the scientist had intended to make a hypothetical provocation about machine intelligence being able to pass itself off as human, a convincing, ludicrous imitation that would make a human believe that he was dealing with another human, which would not really reveal whether such intelligence had surpassed the human, but only deceived it. [12]

This test was taken so seriously by intellectuals that it was transformed into a computing slogan, more like a joke or anecdote of Turing, a provocation that eventually caught the attention of philosophers, especially John Searle, who was responsible for rekindling the debate based on Turing, opposing to this test another one, the one already exhausted in literature, "The Chinese Room", as we have seen. It was about a trick that a computer program could play on a human, cheating him if he could not tell whether or not he was dealing with a machine (or another human). As Boden reveals:

That was tongue in cheek. Although it featured in the opening pages, the Turing Test was an adjunct within a paper primarily intended as a manifesto for a future AI. Indeed, Turing described it to his friend Robin Gandy as light-hearted 'propaganda', inviting giggles rather than serious critique. Nevertheless, the philosophers pounced. Most argued that even if a program's responses were indistinguishable from a human's, this wouldn't prove its intelligence. The most common objection was—and still is—that the Turing Test concerns only observable behavior, so could be passed by a zombie: something behaving exactly like us, but lacking consciousness. [13]

Every year there is a competition run by inventor and industrialist Hugh Loebner that offers \$100,000 ("Loebner Prize") to anyone who comes up with a computer program that can meet the challenge set by Turing (what is called "passing the Turing test"), as well as offering \$2,000 for the "most human program in the competition. Even though no candidate has managed to win the competition and take the top prize [14], the smaller amount is disputed among programmers of the illusion, who are busy creating what are known today as chatbots (a corrupt of chat and robot), which are nothing more than simple programs that have a number of responses to the various inputs entered by the interrogator, giving the illusion of holding a conversation.

#### IV. MACHINE AUTONOMY IS NOT HUMAN AUTONOMY

This technique had already been tested in Weizenbaum's Eliza program in 1966 [15] to give the illusion that the machine could speak, and has only been improved upon by illusionist programmers, the digital magicians of our age, who rely on the technological arsenal developed thus far to create fake prototypes of human-like intelligence. "This reveals a serious [problem] with regard to treating the Turing test as the primary goal of AI. It causes researchers to produce programs that are primarily directed at fooling humans, rather than a more fundamental approach to the problem of intelligence" [16]. A real Turing Test establishes that the interrogator can ask any question, which makes it very difficult, and the pragmatic point raised by Whitby is that imitating human intelligence is of no benefit to the evolution of science or humanity: we already have plenty of human intelligence available, and imitation games to test our intelligence are unnecessary, which is a given. [17]

Searle will later challenge the idea of "understanding" launched by Turing, distinguishing machine performance from understanding reality. [18] The words consciousness and intelligence have been brought in an analogous sense to human faculties. The quarrel arises when we believe that there are machines that know themselves to be machines. Nobody is surprised by the use of the expression artificial intelligence because we have already assumed (and the culture has already absorbed the idea in general) that intelligent machines are those that have an abundance of data that allow it to solve a series of questions, even though it feeds itself back with more data generated from the crossing of data, according to sophisticated systems of operation. These improved executions of tasks do not bring us discomfort, because we know that they are functional systems, and the word intelligence is being used as a way to reprogram itself, or to know how to exponentiate the data implanted in the machines.

It is a giant step when it comes to computational systems, because these are sophisticated programs with high functional performance. Here machine learning means the feedback from the program itself, which has already been architected to "create" outputs from a functional logic that allows this machine-like procedure. The distinction proposed by Brian Cantwell Smith between "accounting" and "judgment", terms that he uses to distinguish machine from human capability, is quite appropriate.

According to him, the machine is capable of reckoning, that is, of "calculative rationality," by manipulating representations that are semantically interpretable, but these systems of interpretation are not capable of understanding what those representations are, that are not themselves capable of holding the content of their representations to account, that do not authentically engage with the world's being the way in which their representations represent it as being.

Cantwell Smith reserves the term judgment to refer to understanding that is able to take objects for objects, to distinguish appearance from reality, that is existentially committed to its own existence and to the integrity of the world as world [understanding the difference of its existence and that of the world], that is subject to objects and bound by them (that is existentially committed to its own existence and to the integrity of the world as world, that is beholden to objects and bound by them). [19]

Calculating rationality lacks prudence, common sense (as it is commonly called), and reason is not just calculation, as some AI researchers believe. To be capable of judgment is to take full account of the consequences of records, and not just records. Records are essential, but they abstract from the details of the world, approximate, violate, and privilege some things over others, since our brain capacity is finite and we live in partial disconnection from the world. Without the records, we would be lost in the world, unable to orient ourselves to something that is not within our effective [real, concrete] reach.

The most important thing is not the records, but what we record and what we take responsibility for in the process, knowing the difference between the two and committing to the latter. In Smith's words,

[...] registrations (even nonconceptual ones) abstract from the world's details - they approximate, do violence, privilege some things at the expense of others. That is not to deny that registrations are essential. Without them, given our finite brain (system) capacity and the world's partial disconnection, we would be at a loss, unable to orient to anything beyond our effective grasp. We would have no world. Yet no matter their necessity, our registrations are not what matters - what matters is that which we register. To be accountable, to hold things to account, is to know the difference between the two - and to be committed to the latter, not the former. [20]

This reference brings us to the Greek notion of *dianoia*, a term used in Greek philosophy to designate the type of discursive thought that proceeds by reasoning, that is, starting from launched premises, which differs from *noesis*, which is the immediate intellectual capture of the intelligible reality, that is, a thought that does not proceed by reasoning, being closer to an intuition: hence we speak of two modes of thought, the *noetic* and the *dianoetic*. On the purely intellectual level of the very way in which human thought is unleashed, one can already detect a subtle difference between thinking by conduction from premises and having an idea "installed" in a totally unpredictable way, as if it were being "generated" at another level of rational activity and that comes to light without our control.

This distinction will be expressed in other terms by contemporary psychology: the so-called insights. If there is no control of thought even over its own intellectual connections, how could there be control over the awareness that they are happening? The thinking subject witnesses the intellectual processes while thinking, as if he were an external observer of his own thinking, that is, he experiences both mental phenomena, that of thinking and that of knowing himself thinking. This double perception of the intellectual process is what makes us *con-scious*, knowing that we know; hence the word *conscious* means witness: to be conscious is to be witness of one's own thoughts and to be responsible for the actions chosen according to what one thinks. [21]

The balance between the conception of human being as rational and free (brought by classical philosophy of rationalist basis) and the radically opposite version that freedom and rationality are a result of physical-chemical processes of the brain (as brought to us by Neuroscience) can be found, with unusual originality, in the thought of Gilbert Simondon, who, starting from categories proper of the theses of the philosophy of technique, and launching elements of Cybernetics, brings an argumentative arsenal coherent with the assumptions of technosciences by establishing a parallel between machine autonomy and human autonomy.

Confirming the diverse states of consciousness, Simondon starts from the distinction between moral consciousness and psychological consciousness and, from there, characterizes human autonomy and machinic autonomy.

Psychological consciousness differs substantially from moral consciousness, according to Simondon, in that it expresses the repercussion of acts or events in the present state of subjectivity, that is, it is a referential state of current occurrences, effective in the individual's experience. It is about a judgment (to be conscious is to be one's own observer) according to a current determination, while moral consciousness, complementary to psychological consciousness, "reports the acts or the beginnings of acts to what the subject tends to be at the end of this act; it can only do this in an extremely precarious way, somehow 'extrapolating' to account for the subject's current transformation". This is why there is relative indeterminacy in the domain of moral consciousness, which initially

establishes one type of reactivity, as simple psychological consciousness (reaction as return of causality), and only in a second moment does it unleash another type of reactivity, which depends on the regime of action that they control (both forms of consciousness).

This is why psychological consciousness is already regulatory, because it imposes (self-) regulation on itself; as we have seen in cybernetic theory, this control is not external to it, its equilibrium depends on regulation and self-regulation. According to Simondon's cybernetic jargon,

[...] in this recurrence of information, the subject is not a being endowed only with a simple internal teleology, but also with a teleology that is itself submitted to a self-regulation: psychological consciousness is already regulative; one's conscience is a regulative consciousness submitted to an internal self-regulation; this doubly regulative consciousness can be called normative consciousness. It is free because it elaborates on its own regime of regulation itself. [22]

This distinction is extremely valuable to reflect on the notion of "functional consciousness" proposed by strong AI theorists, who argue for the possibility of a consciousness emerging from artificial intellectual machinic activity. A merely functional consciousness is one that exercises reckoning-type rationality, as presented by Cantwell Smith, which is merely accounting or recording, being a computational construct to mean information processing. The idea of consciousness supposes self-reflection, attributed exclusively to humans and referring to psychological/emotional states according to "sensations (such as melancholy or affliction) or 'qualia' (a technical term used by philosophers) [whose] existence, in an essentially material universe, is an evident metaphysical enigma" [23]

Ending with Simondon, when he distinguishes informational regulation of consciousness on the psychological and moral planes:

A teleological mechanism can imitate the functioning of psychological consciousness, which can be instantaneous; but teleological mechanism cannot imitate conscience, for it never has a twofold and simultaneous conditioning; the organic and the technical must already be present, close to being put into relation, in order for conscience to be able to exist. Valorizing consciousness therefore defines a level of teleological activity that cannot be reduced to any automatism. The solution to the moral problem cannot be sought by a computer. [24]

Only the human being is capable of conversion, in the sense of being able to totally change the directions of his existence, unlike the automaton, constituted by a teleology proper to technical beings, whose universality is always causal. The machine is precise in its teleological mechanisms, reason why it will be "the more perfect the more its automatism allows it, according to its predetermined purpose, to regulate itself. But the machine is not self-creating," whereas the human has as part of the problems to be solved by him the fact that he himself is a problem. The human has an open faculty of acquiring information beyond his present structure, so that the information acquired is recast in the present structure, and he can always question himself.

This capacity to be itself one of the terms of the problem one has to solve does not exist for the machine. [...] the machine cannot have any veritable exteriority, because it has no veritable interiority within itself [...] The machine can lose its regularity and then present functional characteristics analogous to wild behavior in a living being. But it cannot revolt. Revolt in fact implies a profound transformation of finalized

behaviors and not a malfunctioning of behavior. The machine is capable of self-adaptive behaviors; but there is nevertheless a big difference between a self-adaptive behavior and a conversion that no external resemblance can disguise: man is capable of conversion in the sense that he can change goals [fins] throughout the course on his existence. [25]

Here it would be appropriate to bring, yet, the senses of the experience of the drive, as thought by Freud, since the machine can be endowed of performance to execute the most remarkable functions (to generate impulses), surpassing, in this strict sense, the human brain activity; but it will never be endowed of the impulse that triggers its performance, this *causa sui* that does not derive from the engendering, neither from parts of the body, nor from parts of the machine, nor even from constitutive elements of the whole universe. According to Freudian theory:

Pulse (Trieb) is a stimulus that arises within the body, from which, therefore, one cannot escape; unlike external stimuli, Freud describes. Drive is a border concept between the psychic and the somatic. Its origin is a state of arousal, its constitution is an amount of energy pressing in a certain direction, its purpose is to put an end to arousal, and its objects vary. Obscure intersection between the mental and the physical, the primary drive is to drain the excitation, to release, to keep constant, or as low as possible, the amount of excitation in the psychic system. [26]

Here it is also important to distinguish drive and instinct, like the animal automata (Aristotle, Descartes) endowed, as humans are, with unreflective conditionings instanced by nature in the constitution of their bodies, which act according to natural causation. There is a discussion about the use of the word pulse in Freudian thought, because he prefers the word Trieb as Instinkt, since the latter would be linked to the idea of inherited and unchangeable model, which is not verified in the idea of pulse. Among Freud's translators, Ernst Jones proposed the translation of Trieb by impulse or instinct, Laplanche understood that Freud only used the word Instinkt few times to refer to animal behavior, and Pontalis suggested that the Freudian theory of drives cannot be confused with psychological readings of animal instinct. But, as Kangussu ponders, arguing with Paulo César de Souza, the distinctive proposals of the last two suggest a "rupture or censure between the human and the animal, despising what there would be of the latter in the animal". [27]

Although it is possible to equate animal and man, since animality is also part of the latter, the task of philosophical reflection has always been the search for something beyond the animal in the human, a *telos* that transcends his situated finitude, a *telos* that projects functional similarities in the form of mimetic machines of human intelligence, the human not even being able to reproduce animal instinct.

Human fabrication has never been able to reproduce even an entity endowed with instinct (something similar to an animal), a being that could have elementary perceptions through the senses (sentient). The most sophisticated software, by its very composition and way of functioning, no matter how skillful it may be in obtaining computational results similar to the activities of the human intellect, is not capable of sentience, because it is not capable of sensitive affections, a condition for the access of conscious and sentient beings to the world.

Any artificial intelligence can be more than a man (and an animal, of course) in function, because it imitates activities that have been reproduced from the observation of bodies, in the case of AI, from the functioning of the brain organ in the designed computing machine (In this sense, see Ernst Kapp's Organ Projection Thesis). But artificial beings will never imitate reaction, which presupposes either freedom, or a "programming" that we are far from designing, as a mechanism lacking empirical realization might wish: the programming (natural, divine, demiurgical, it doesn't matter) of the reactive capacity, however primary, because no machine escapes its

programmed destiny.

Even machine learning is a program millimetrically thought out to fill algorithmic gaps, according to the teaching left by Cybernetics that any machine is nothing more than the selection of components in a set of diverse occurrences, each machine being defined by the presence of a regulator of varieties of occurrences in relatable sets, which defines the way to put components together correctly. Without this joining of occurrences, there is no machine performance, and instinctive and intelligent occurrences are still far from being joined together in an artificial entity design.

And here we have a practical issue: there are no algorithms that can reproduce the complex web of human activities, always related to other humans and referred to ends that transcend experience (such as the realization of the good and the just), which generates a countless range of expectations unrealized by unpredictable actions. From the point of view of practical execution, we are far from delimiting a set of occurrences to systematize them by thousands of algorithms.

Margaret Boden treats the subject as a "philosophical imbroglio", using some references, for example, David Chalmers' thesis on phenomenal consciousness, as "an irreducible property of the universe, analogous to mass or charge"; the position of Colin McGinn, who "has come to claim that human beings are intrinsically incapable of understanding the causal connection between brain and qualia, just as dogs cannot understand arithmetic"; and the thesis of Jerry Fodor, "a leading philosopher of cognitive science, [who] believes that no one has the slightest idea how anything material could be conscious". [28]

There is a chapter in Boden's introductory work on Artificial Intelligence, entitled "But is this really intelligence?" It is with this rhetorical question that we direct the closing of this chapter, establishing the understanding that the concept of intelligence was taken in a sense quite different from what we consider in philosophy, a simple metaphorical transposition of the functioning (and not the origin, foundation or nature) of intelligent human activity. And here it is important to note the understanding manifested in the European Ethical Charter:

The "strong" AIs of science fiction literature do not exist. This type of AI, which would be equipped not only with intelligence but also with conscience, remains purely fictional. The machine learning systems currently being developed are described as "weak" AIs and are capable of extracting complex patterns and learning from large volumes of data efficiently and often with high levels of predictive accuracy. To overcome any instinctive or fantasized considerations, these information processing and analysis technologies must be understood on the basis of the following three concepts. - AI is not a single, homogeneous object: it is actually an assemblage of sciences and techniques (mathematics, statistics and computer science) capable of processing data to design very complex computer processing tasks; - AI engines do not produce intelligence per se but proceed using an inductive approach: the idea is to associate in a nearly automated way a set of observations (inputs) with a set of possible results (outputs) using various preconfigured properties. [29]

See that this is a fallacy built in the imaginary of this century around the area of computing called Artificial Intelligence: that artificial intelligence is intelligent and that artificial life is living, nothing more than anthropomorphizations of systems and processes, as occurred in other moments of technical experience, or of the manifestation of human technicality, such as the figure of the automaton.

The computational area has no pretension to engage in discussions of the classical notion of human nature upon

which Practical Philosophy was built, simply because the purposes of this research are merely instrumental, to improve the performance of machines. This technical-scientific approach cannot be confused with the validity of the philosophical approach, which is absolutely unscathed by these instrumental metaphors.

The assumption of Practical Philosophy is a metaphysical human instance untouched by the conditionings of animality in man, who stands out from all others by his ability as sapiens, and this fact has not been contested by any demonstrative science or by any technique of simulation of cognitive activity to date. It is a value upon which Anthropology and Ethics have risen, which is the core of human dignity, namely, the unique greatness of the human being as the only animal endowed with logos. In the words of Brazilian philosopher Henrique Cláudio de Lima Vaz:

The greatness of the human being as the only animal endowed with logos, which makes him absolutely unique, is celebrated by the Western philosophical tradition, from Plato to modernity, through the biblical tradition exhorting the genesis narrative of creation, being thus received in the Christian tradition and becoming a privileged theme in Renaissance thought which extends from Descartes to Hegel. [29] Dignity is the concept that synthesizes this uniqueness of the human being, by designating an attribute that is essential and inalienable, appearing in art. 1 of the Declaration of Human Rights of 1948. As a manifestation of reason and freedom, human dignity is, however, an essential predicate of the I am in the entirety of its expressions: body, psyche, spirit. This observation is directed against the dualistic conception according to which dignity belongs properly to soul and spirit, and against the monistic conceptions which, reducing the human being to the simple organic, deprive the attribute of dignity of any *raison d'être* [...It took a long evolution in the social, legal, political and philosophical fields for the concept of human dignity to obtain the recognition of universality already achieved in the philosophical and religious tradition by the idea of the human being as a being endowed with a unique nature and greatness manifested in reason and freedom and that make him unique in the history of life. [30]

Since Greek philosophy this greatness is recognized, although justified politically in other terms. Aristotle, for example, according to Hannah Arendt, never denied the condition of humanity to slaves. In fact, what he did not attribute to them was the condition of man, because they had no possibility of freeing themselves from the conditioning imposed on them by human needs, which resembled animals that act upon nature only to provide for their own subsistence. Human beings are not free to act in this way: the product of their actions/activities is totally integrated into the body, like the food they need to prepare, just like an animal does, but in a more elaborate way, the result of their intellectual capacity. Aristotle, who freed his slaves when he was on his deathbed, "denied the use of the word 'man' to designate members of the human species totally subject to necessity," to whom the expression *animal laborans* is more adequate than *animal rationale*, as we shall see below. Now, the human animal is "really only one of the animal species living on earth - at best the most developed. But, still, an animal - never denied Aristotle". [31]

It happens that this animal has a peculiarity: besides moving, idealizing things and projecting itself in the production of artifacts and various symbologies, it carries the responsibility of acting ethically, a greatness that allows it to act freely towards the simplest to the highest purposes, but which also impels it to give the reasons for its actions, that is, to take responsibility for them. Western construction of the idea of "man" (human being), the basis of the dignity of the human person. This dignity supposes what the Brazilian philosopher Henrique Cláudio de Lima Vaz calls human spiritual metabolism, which is conformed by freedom and will, and therefore, by

responsibility.

## V. CONCLUSION

The eschatological promises of machine domination and control over our lives find fertile ground in the field of AI and Robotics. Robotics produces humanoids and these are machinic artefacts apparently endowed with a mysterious self-sufficiency. This expectation is perhaps the most persuasive thrust of the futurological discourse of computer scientists and programmers today, a contagious rhetoric that foreshadows the revolution that will be promoted in the field of Neurosciences, as this science can be associate with the reverse engineering of the brain to move towards the humanoid incorporation of so-called strong artificial intelligence (John Searle). This strong AI was made possible by the computational power applied to the field of Robotics. And it is expected that this achievement will soon be the objective of Robotics.

To think of human action is to attribute to it the condition of ethical action, admitting that the animal sapiens carries within itself an instance of decision that differs from the physical functionalities of its body in the form of a conscious mind, which is not to be confused with the brain, which carries it without, however, causing it.

To close this argument and move on to the symbolic effects that the belief in a machine intelligence can provoke in the symbology of the human of this century, impacting his beliefs and praxis, we conclude by the obvious, a synthesis that dispenses with any discussion about machine autonomy: no human being can be attributed the condition of functionality, being always an end in himself (hence his intrinsic and inalienable dignity), his ontology totally distinguishes him from the ontology of machines, which, not being subjects, are, since always, objects destined to perform functions.

## REFERENCES

- [1] B. Whitby, *Inteligência artificial: um guia para iniciantes*. Tradução de Claudio Blanc. São Paulo: Madras, 2004, 117.
- [2] D. Dennet, *Quining Qualia*. In: GOLDMAN, A. (Ed.). *Readings in Philosophy and cognitive science*. Cambridge: MIT Press, 1993, 2,4.
- [3] B. Whitby, *Inteligência artificial: um guia para iniciantes*. Tradução de Claudio Blanc. São Paulo: Madras, 2004, 20.
- [4] M. Brochado, *Ética e direito. Pelas trilhas de Padre Vaz*. Curitiba: CRV, 2021b.
- [5] H. Bergson, *Creative Evolution*. New York: Dover Publications, 1998, 139,141.
- [6] M. Boden, *Inteligência artificial: uma brevíssima introdução*. Tradução de Fernando Santos. São Paulo: Unesp, 2020, 167.
- [7] M. Brochado, *Inteligência artificial no horizonte da filosofia da tecnologia: técnica, ética e direito na era cibernética*. São Paulo: Editora Dialética, 2023.
- [8] M. Brochado, *Epistemological and Ethical Reflections on Artificial Intelligence as an Area of Cognitive Computing*. In: Great Britan Journals Press. London Journal Press, 2024. Available at: <https://journalspress.uk/index.php/LJRHSS/article/view/718>. Accessed 22 Ma 2025.

- 
- [9] M. Brochado, *Inteligência artificial no horizonte da filosofia da tecnologia: técnica, ética e direito na era cibernética*. São Paulo: Editora Dialética, 2023.
- [10] E. Rich; K. Knight, *Inteligência artificial*. Tradução de Maria Cláudia Santos Ribeiro Ratto. São Paulo: Makron Books, 1993, 665-666, 645; S. Russel; P. Norvig, *Inteligência artificial*. Rio de Janeiro: Elsevier, 2004, 870, 872; S. Russel; P. Norvig, *Artificial Intelligence: a modern approach*. 4. ed. Hoboken: Pearson Education, 2021, 1690-1693.
- [11] M. Boden, *Inteligência artificial: uma brevíssima introdução*. Tradução de Fernando Santos. São Paulo: Unesp, 2020, 170.
- [12] M. Brochado, *Inteligência artificial no horizonte da filosofia da tecnologia: técnica, ética e direito na era cibernética*. São Paulo: Editora Dialética, 2023.
- [13] M. Boden, *AI: its nature and future*. New York: Oxford University Press, 2016, 120.
- [14] M. Boden, *AI: its nature and future*. New York: Oxford University Press, 2016, 121.
- [15] J. Weizenbaum, *Computer power and human reason: from judgement to calculation*. New York, W H Freeman & Co, 1976.
- [16] B. Whitby, *Inteligência artificial: um guia para iniciantes*. Tradução de Claudio Blanc. São Paulo: Madras, 2004, 33.
- [17] B. Whitby, *Inteligência artificial: um guia para iniciantes*. Tradução de Claudio Blanc. São Paulo: Madras, 2004, 33.
- [18] B. Whitby, *Inteligência artificial: um guia para iniciantes*. Tradução de Claudio Blanc. São Paulo: Madras, 2004, 89.
- [19] B. C. Smith, *The promise of artificial intelligence: reckoning and judgment*. Cambridge: MIT Press, 2019, 110.
- [20] B. C. Smith, *The promise of artificial intelligence: reckoning and judgment*. Cambridge: MIT Press, 2019, 111-112.
- [21] M. Brochado, *Ética e direito. Pelas trilhas de Padre Vaz*. Curitiba: CRV, 2021b, 56-57.
- [22] G. Simondon, *A individuação à luz das noções de forma e de informação*. Tradução de Luís Eduardo Ponciano Aragon e Guilherme Ivo. São Paulo: Editora 34, 2020, 407.
- [23] M. Boden, *Inteligência artificial: uma brevíssima introdução*. Tradução de Fernando Santos. São Paulo: Unesp, 2020, 168.
- [24] G. Simondon, *A individuação à luz das noções de forma e de informação*. Tradução de Luís Eduardo Ponciano Aragon e Guilherme Ivo. São Paulo: Editora 34, 2020, 407.

- [25] G. Simondon, *Individual in Light of Notions of Form and Information*. Translated by T. Adkins. London, University of Minnesota Press, 2020, 419, 418.
- [26] I. Kangussu, *Sobre eros*. Belo Horizonte: Scriptum Livros, 2007, 88.
- [27] I. Kangussu, *Sobre eros*. Belo Horizonte: Scriptum Livros, 2007, 89.
- [28] M. Boden, *Inteligência artificial: uma brevíssima introdução*. Tradução de Fernando Santos. São Paulo: Unesp, 2020, 169.
- [29] European Commission For the Efficiency of Justice (CEPEJ). *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment*. Adopted at the 31st plenary meeting of CEPEJ, 2018, 31-32. Available at: <https://www.europarl.europa.eu/cmsdata/196205/COUNCIL%20OF%20EUROPE%20-%20European%20Ethical%20Charter%20on%20the%20use%20of%20AI%20in%20judicial%20systems.pdf>. Accessed on 22 May 2025.
- [30] H. C. VAZ, *Escritos de filosofia V: introdução à ética filosófica 2*. São Paulo: Edições Loyola, 2000, 201-202.
- [31] H. Arendt, *A condição humana*. Rio de Janeiro: Forense Universitária, 2007, 95.

1 Full Professor of Philosophy of Technology and Philosophy of Law at the Law School of the Federal University of Minas Gerais, Brazil, with research in the Foundations of Philosophy of Technology and Ethics applied to the human-machine connection, in the regulation of Artificial Intelligence, and in Digital Law Theory. Member of INCT-IACiber - Artificial Intelligence for Cybersecurity: a Technological and Social Approach- Department of Computer Science/Federal University of Minas Gerais, Brazil. Coordinator of the PhiloTech Chair: Philosophy of Technology applied to Digital Law/Jean Monnet Center of Excellence (European Union/Federal University of Minas Gerais, Brazil). Visiting Professor at the Leibniz Institut für Medienforschung - Universität Hamburg, Germany. PhD in Law from the Federal University of Minas Gerais, Brazil, with a Postdoctorate in Philosophy from Ruprecht-Karls-Universität Heidelberg, Germany. Associate of The Society for Philosophy and Technology-SPT (<https://www.spt.org/>). President of the Commission on Artificial Intelligence in Law of the Brazilian Lawyers Association of Minas Gerais, Brazil. Director of Human Rights at the Institute of Law and Artificial Intelligence (IDEIA) of Minas Gerais, Brazil. Collaborator for PACED (Project for the Consolidation of the Rule of Law)/PALOP-TL/European Union. Honorary member of the Ibero-American Union of Judges – UIJ.

ORCID: 0000-0001-5863-7360

2 From quale, in Latin, which means quality, qualia (plural) is a word adopted by Daniel Dennet and others who dialogue with him to refer to absolutely unique ineffable properties of the human mind, directly accessed by the consciousness of each individual, being these inner sensations that allow us to access our conscious states, such as the unique experience of pain, taste, perception of colors and shapes, etc., which are not universally definable or even subject to precise definition because they are not communicable. Dennet exemplifies with how orange juice tastes to him at breakfast, before and after tasting other items, and how cloying the smell of cauliflower seems to him, which to many people is appetizing. In short, the idea is that there are conscious experiences that are not definable. In his words: "the very properties the appreciation of which permits us to identify our conscious states. So, to summarize the tradition, qualia are supposed to be properties of a subject's mental states that are: ineffable, intrinsic, private, directly or immediately apprehensible in consciousness".

3 The discussion on artificial conscience would, of itself, require a major comparative digression among the expressive theses on human conscience, from Socrates (gnothi seauton), through St. Thomas Aquinas (synderesis/conscientia), and on to other contemporary approaches, such as that offered by the work of Henrique Cláudio de Lima Vaz. The deepening in this specific theme does not fit in the program of this book, reason why, for the deepening in the genealogy of moral conscience, I indicate the reading of these book: *Ética e Direito: pelas trilhas de Padre Vaz*.

4 “The current rules prescribe twenty-five-minute interactions, using twenty pre-selected questions designed to test memory, reasoning, general knowledge, and personality. The judges consider relevance, correctness, and the clarity and plausibility of expression/grammar. As yet, no program has fooled the Loebner judges for 30% of the time”