filosofia unisinos | Unisinos Journal of Philosophy

ISSN 1984-8234

Filosofia Unisinos
Unisinos Journal of Philosophy
26(1): 1-13, 2025 | e26109
Nome dos editores responsáveis pela avaliação:
Inácio Helfer
Leonardo Marques Kussler
Luís Miguel Rechiki Meirelles

Unisinos – doi: 10.4013/fsu.2025.261.09

Article

Artificial intelligence, extended cognition, and the narratives of cyberimmortality

Inteligência artificial, cognição estendida e as narrativas da cyberimortalidade

Léo Peruzzo Júnior
https://orcid.org/0000-0003-3084-5170
Pontifícia Universidade Católica do Paraná - PUCPR, Programa de Pós-Graduação em Filosofia, Curitiba, PR, Brasil.
E-mail: leoperuzzo@hotmail.com

ABSTRACT

The article shows how the extended mind theory, which suggests that cognition is not confined to the brain but extends beyond the body, is used to redefine our understanding of the nature of the mind and legitimize narratives of cyberimmortality. Additionally, it explores how this concept has been considered within the field of Artificial Intelligence (AI). By exploring the interactions between the environment, technological devices, and mental processes, the extended mind theory challenges the traditional boundaries of historically constructed epistemology upon the mind-world dichotomy. The paper analyzes various features of this new cognitive topography and how this perspective can transform human experience into a narrative that transcends organic limitations. Finally, it highlights a few consequences and critical challenges of the extended mind proposal in the recovery of the body and the environment.

Keywords: extended cognition, cyberimmortality, environment, artificial intelligence (AI), technological artifacts.

RESUMO

O artigo pretende mostrar de que modo a tese da mente estendida (*extended mind theory*), segundo a qual a cognição não está confinada ao cérebro, mas se estende para além do corpo é usada, por um lado, para redefinir nossa compreensão sobre a natureza da mente e conferir legitimidade às narrativas da chamada *cyberimortalidade* e, por outro, como isso tem sido pensado pela Inteligência Artificial (IA). Ao explorar as interações entre o ambiente, os dispositivos tecnológicos e os processos mentais, a tese da mente estendida desafia as fronteiras tradicionais da epistemologia que fora historicamente construída sobre a dicotomia mente-mundo. Para isso, o trabalho analisa os vários traços dessa nova topografia do cognitivo e como essa posição é capaz de transformar a experiência humana em uma narrativa que transcende as limitações orgânicas. Por fim, aponta algumas consequências e desafios críticos da proposta da mente estendida na recuperação do corpo e do ambiente.

Palavras-chaves: cognição estendida, cyberimortalidade, ambiente, inteligência artificial (IA), artefatos tecnológicos.

1 Introduction

In the mid-1950s, the project led by John McCarthy, Marvin Minsky, and their colleagues at Dartmouth College in New Hampshire was responsible for crafting one of the most antagonistic, appealing, and dangerous metaphors of our era: to build computers capable of "all aspects of learning — or any other feature of intelligence — that can, in principle, be so precisely described that a machine can simulate them" (A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, 1956, p.2). McCarthy had popularized the term "artificial intelligence" in the previous decade, implying that it would be possible to produce mathematical models that would simulate the functioning of brain neurons. The fact is that this metaphor established many other questions, and, more than a technical or conceptual definition, the term "Artificial Intelligence" became yet another of the unique objects produced by technology. But what exactly is the design proposed by Artificial Intelligence? Is Artificial Intelligence a construct in the style of Non-artificial Intelligence — that is, Human Intelligence? Or is it an attempt to reach a completely different stage from the set of solutions that we can traditionally think of and solve?

In his essay *Minds are simply what Brains Do*, Minsky signals a decisive answer to the previous questions: brains and minds are not different; they do not exist in separate worlds, since they are different viewpoints to describe the same things. In other words, minds are simply what brains do, although brains are immensely complex machines. When we talk about a mind, we are referring to the processes that move our brains from one state to another, he concludes. Minsky (1986, p.287) also offers one of the most seductive descriptions of future Artificial Intelligence: "When the mind is considered, in principle, in terms of what the brain can do, many issues that are generally considered philosophical can now be recognized as merely psychological — because the long-sought connections between mind and brain do not involve two separate worlds, but simply relate two viewpoints." Therefore, minds are seen only as relations between states, a society of agents that can function without knowing the physical constitution of the other agents they are connected to.

The program built to support Artificial Intelligence, therefore, detached itself from a material analysis of this society of agents (colors, sizes, shapes, or any other individual properties of the agents), because "it doesn't matter what the agents are; only what they do and what they are connected to matters" (Minsky, 1986, p. 287). However, brains would not manufacture thoughts in the same way factories manufacture cars. According to Minsky (1986), brains use processes that modify themselves, indicating,

in turn, that we cannot separate such processes from the products they produce. The main challenge for science, therefore, would be to understand the activities that brains perform to changes themselves, especially because they are machines with vast amounts of parts that operate perfectly in accordance with the laws of physics. And machines of such complexity, as we know, would depend on better instruments and theories than those we have at the moment.

In the early 1990s, Rich and Knight (1991, pp.3-4) proposed that the goal of Artificial Intelligence is to develop systems to perform tasks that, at present, on the one hand, are better performed by humans than by machines and, on the other, have no viable algorithmic solution through conventional computing. Insisting on the classical project, Rich and Knight (1991) understand that some of these problems could be solved by a defined algorithm model, thus having causal and exact solutions. Others, largely due to their complexity and the relationships that do not have a viable algorithmic solution through conventional computing, seem to cluster the way human intentionality functions. For example, choosing the least congested street among the available data and streets or the file that presents more words according to those typed in the search engine seems to be a less complex task than the ability to create a poem or make a decision regarding a mundane life event. Based on such hypotheses, we are therefore led to believe that thinking is nothing more than a symbolic, mathematically programmable, causal structure, and, in most cases, so efficient that the moment, now no longer time, is capable of producing a decision.

It is at this moment that the connectionist paradigm comes onto the scene to advocate its thesis: language is represented as a network of basic components modeled after the function of the brain. Artificial neurons, interconnected, can learn and extrapolate based on examples, since reasoning basically consists in learning the input-output function. According to this approach, then, it would not take long to obtain an answer for a problem with possible solutions, since every choice is nothing more than the solution according to a given criterion. In other words, it means that the domain of Artificial Intelligence is not only conceived an algorithmic solution to thought but also assumed that the collection of models, techniques, and technologies can infer a priori the effect of all choices.

Therefore, if there is a philosophical project that underpins Artificial Intelligence, such project appears as the algorithmization of thought, that is, the reduction of the cognitive structure to finite, causal, and linear descriptions. After all, an intelligent process is nothing more than knowing how to think correctly; and thinking correctly is the conditional distribution that can occur in the topology written by such algorithms. It is exactly at this point that cognition distanced itself from the environment, decision mechanisms, natural language processing, and handling uncertainties. After all, the trump card of modernity is still the aspiration that technological progress, now identified with the domain of human consciousness, may be capable of telling us how we think and what exactly are the limits of what we can know (Muller, 2013).

The argument behind this debate, therefore, is not an attack on the hypothesis that we will soon share the world with a new type of highly capable entity, trained but not controlled by us, which could represent a catastrophic risk for humanity (Bales; D´Alessandro; Kirk-Giannini, 2024). On the contrary; it is about analyzing how projects that imagine Artificial Intelligence systems with human cognitive abilities assume that an intelligent agent is an encapsulated computational system, i.e., its actions are autonomous in relation to the environment where certain objectives can be achieved. This hypothesis ignores the fact that "not only is there no commonly accepted definition of AI, but what the term refers to will, at the very least, depend on whether one is talking about AI as (a) a scientific field, (b) a technology or method, or (c) concrete applications of AI systems" (Konig et al., 2022, p. 18).

Faced with these possible approaches and readings, there are at least two epistemologically critical ways to evaluate Artificial Intelligence: the first is to analyze the assumption that human cognition can house or instantiate something like a digital cognition, now disembodied from the world and body; the second, based on the first, is how such a hypothesis is capable of constructing new narratives, among

which is cyberimmortality and its argument that this may overcome the conventional boundaries between life and machine.

2 Encapsulating artificial intelligence

While it is broadly and technically accepted that Artificial Intelligence can be defined as the computational aspect of goal achievement in the world, that is, the ability to find the most appropriate course of action, this does not yet imply an adaptive capacity to acquire new strategies and actions. According to (Konig et al., 2022, p. 25), when it comes to implementing AI as a technology, this can take various forms. An important distinction is found between decision rules derived manually and rules learned by an AI system, introducing greater complexity and new challenges.

Another central distinction is between general AI (a form of intelligence that can understand, learn, and perform any cognitive task that a human could do) and narrow AI (an Artificial Intelligence system designed to perform a specific task that lacks the ability to automatically transfer its knowledge or skills to different domains). AI has sought, since its inception, to achieve a general intelligence similar to humans, capable of handling a variety of tasks and adapting to new situations. However, most efforts are directed towards narrow AI solutions, which simply deal with specific tasks to achieve well-defined goals. Additionally, AI systems depend on a broader context in which they are implemented: "a specific technological solution, such as a lip-reading device, may be harmless or even highly beneficial in one environment, for example, for deaf people, but may cause profound ethical and regulatory problems as part of a public video surveillance system..." (Konig et al., 2022, p. 28). Thus, the question of whether human-level Artificial Intelligence is possible has turned into the question of whether it is technologically viable to replicate it without the human body and without any ethical regulations. However, there are important philosophical questions that remain unsolved and deserve closer scrutiny.

It is precisely here that we need to note three brief issues: the first being that the design of Artificial Intelligence ignores the environment and the body and focuses specifically on the idea of a system. An encapsulated system is enough to cooperate and coordinate problem-solving. Moreover, this same algorithmic system can only communicate in its own language. It is automaton-like in that its world is a calculation produced to fulfill its own objectives. Daniel Susser (2013, p. 278), for example, in revisiting and expanding on Dreyfus' arguments about the reproduction of Artificial Intelligence, argues that there are much broader questions to consider regarding what it means for the body to be fundamental to all aspects of intelligent life: "What would an artificial non-human body be like? What is sufficient to constitute it? Indeed, what is common to all the different types of intelligent creatures found in nature? What is common to human bodies, dog bodies, and octopus' bodies?" It seems evident that, contrary to the formalistic desires that nurture AI projects, the meanings in which certain definitions fluctuate and the context dependency are inherently indeterminate. In short, what aspects of the context should, in principle, be considered during the formalization of intelligence?

In summary, Susser (2013, p. 285) argues that body and intelligence are not distinct things, as the body is a fundamental piece for all aspects of intelligent life and is therefore coextensive wherever there is intelligence. Agreeing with Susser's position, the more or less discrete physical systems we call bodies are exactly the type of physical systems endowed with the ability to interact skillfully with their environments. Hence, the analytical distinction between bodies and intelligence, artificial and organic, and many others, refers to two aspects of the same phenomenon. This leads us to an even greater philosophical problem: the symptom that future AI research should continue to pursue a disembodied model of intelligence and, once found, may re-embody it or attach it to humanoid robot bodies and complex computers.

The second issue, perhaps as strange as the first, is attaching the monolithic idea of intelligence to the system, even though we have entered or constructed, especially in recent years, the concept that our

platforms and systems "will have enough intelligence to learn", as some areas of Computer Science have been interested in, including Machine Learning, Deep Learning, Neural Networks, Cognitive Computing, and Natural Language Processing. Intelligence, therefore, would be the ability to "deeply understand" the way of creating itself, without implying, again, an opening horizon beyond the elements that make up the system. Authors like Peter Konigs (2022) and Lauwaert (2021), for example, consider that the problem is not related to intelligence, but to the ability of machines to be considered sufficiently sophisticated or agent-like to be themselves possible bearers of responsibility. Ignoring the pessimistic reading regarding the gaps in responsibility and possible technological catastrophism, the authors believe that, when people interact with intelligent systems — producing them, programming them, using them, etc., it may be difficult to determine to what extent these people should be held responsible for a result caused by an intelligent system. Konigs (2022, p. 9), for instance, states that "this problem is epistemic rather than metaphysical. It does not consist of the real (metaphysical) absence of a responsible agent, but of the (epistemic) difficulty of correctly determining how responsible people are."

Now, the argument seemingly at stake is not the consistency of hypotheses about what it means to be an intelligent agent, but the alleged responsibility that could not be attributed to such intelligent agents because it could violate the genuine ability to constantly make new decisions. Thus, a complete explanation of what autonomy consists of and whether machines could possess it would also require us to be able to answer a set of questions about the nature and limits of will, as Sparrow (2007, p. 65) points out. The fact is that combining the autonomy of a system with the possibility of these same systems having a significant capacity to form and revise their own beliefs, allowing their actions to become unpredictable, is still a nebulous issue and raises a series of difficult ethical questions.

Finally, there is a third issue that needs to be addressed when considering the encapsulation of Artificial Intelligence and the exclusion of the body and the environment. This question is related to how the idea of artificiality is conceived when, in fact, what is meant is only the formal replication of human cognitive structure. Indeed, the artificial is not a new epistemic category, but merely the continuity of the "natural," now instrumentalized on the assumption that non-organic elements can perform the same functions as the organic. And this, agreeing with Taylor's position (2024), leads us to an even bigger problem: the use of Artificial Intelligence to make high-risk decisions prevents anyone from being considered morally responsible for the outcomes achieved. After all, how should we attribute moral responsibility if we are faced with a disembodied intelligent agent?

According to Taylor (2024, p. 1), technological developments could take the burden of certain moral decisions out of human hands. We could have less biased, irrational, or incompetent artificial decisions. However, the shift to algorithmic decision-making would entail other moral costs, among which would be the "responsibility gap," where no one is morally responsible for the behavior of these systems or the results they bring about. Taylor (2024, p. 14) further states that even collective responsibility models that presuppose individual responsibilities could not be applied to many cases of AI development and implementation due to the autonomous nature of the systems. This reading, however, relies on the gaps in attributing collective responsibility to mitigate the attribution of responsibility to the actors behind the programming lines of intelligent artificial agents.

In any case, the encapsulation of the artificial represents even more: it is the assertion of a boundary imposed on the neural network, as it would process information independently of the surrounding environment; the possibility of surpassing time and space, making any content a constantly present memory; and, furthermore, of being able to have autonomy and decision-making power, the same utopia that philosophers and professionals have historically pursued and do not know exactly what it means. Artificiality, therefore, is the project of this carefully crafted new image, whose identity is everywhere and nowhere, where meaning is purely definable in terms of computational semantics and whose narrative of moral responsibility would be naive.

Obviously, the idea that AI could function as a form of cognitive extension and thus help to cognitively enhance human users has been gaining more and more adherents. However, the question would not only be whether to entrust AI technologies with tasks for which we use our intelligence could potentially be a way of making ourselves less intelligent, but why we would be leaving ourselves tasks that do not require us to be very intelligent (Nyholm, 2024, p. 80).

However, before referring to the old idea of cognition that has been assumed by all these projects, we would like to remind, finally, that the literature on Artificial Intelligence has assumed various definitions for the term autonomy, one of them being the concept of "autonomous agents." By autonomous agents, then, are artificial systems capable of "having their own existence" independently of other agents, being autonomous in relation to the environment, being able to work in dynamic and uncertain environments, achieve their goals on their own and without the need for cooperation with other agents, and also separate their decision-making capacity from motivations. Therefore, what is at stake is not simply human-agent interaction, but the theoretical and political architecture that mitigates our understanding of what can be produced as an intelligent, artificial, autonomous, and generative system in what it can create.

3 Artificial intelligence in extended minds?

Since the 1980s, we have witnessed an epistemological shift in cognition studies (Maturana, Varela, 1980; Varela, Thompson, Rosch, 1992). For instance, the concept of autopoiesis begins to indicate the ability of a system to produce and maintain its own components, with such autopoietic systems being closed in terms of their components but open in their interaction with the environment. Thus, cognition ceases to be merely a functional, connectionist, and formal system. Instead, it comes to be viewed as a dynamic, situated system that extends to the artifacts surrounding it.

However, the traditional distinction between mind and machine, natural and artificial, organic and non-organic, human and transhuman, body and consciousness, still preserved the idea of demarcation and independence of the cognitive from the "lived-in world," i.e., from the environmental artifacts that are intrinsically linked to each other. The metaphor inaugurated by Cognitive Science in the 1960s was that the body and the lived-in world (or the environment) no longer occupied the same space in this new topology. There are organs without bodies, as Deleuze would write today (*corps sans organes*); there are cognitive structures displaced from the world and situated in the empty space of mathematical solipsism. Or rather, there is a shift in thinking to the idea that the artificial subject could violate all physical laws that govern bodies, as the software would become arbitrary, entirely generative, and thus both the subject and product of itself.

While theories of embodied cognition moved towards solidifying the inseparability of intelligent agents from the environment and their bodies, technological nihilism forged a total division: intelligent systems should be able to interact with the environment, but change their internal states in such a way that they are independent of external interventions. Thus, on the one hand, robots and any other forms of Artificial Intelligence should be able to mimic human intelligence in general; on the other hand, considering AIs as moral agents or patients would be conceptually inappropriate and morally dubious (Pellegrino; Garasic, 2020, p. 150). However, although objections can be raised to attributing intentionality to AI minds and the possibility of moral responsibility, part of the machine's behavior depends on other software, which often operates according to interfaces and external factors.

At the same time, adopting a comparative perspective between Human Intelligence and AI requires us to be able to ask ourselves what we understand by basic or fundamental subjectivity. This question does not require us to be able to define whether AI has self-awareness, but whether there is a minimal point of view about the world that allows it to define itself in terms of something or someone. Northoff

and Gouveia (2024), for example, adopt a neurophilosophical strategy to argue that, at its current state, AI does not exhibit a basic or fundamental subjectivity, since the main characteristic of what the authors call the ecological base of the point of view is its relational nature. Additionally, there is a difference in cognitive flexibility between current AI models and the human brain, since these models can only specialize in specific tasks, lacking the kind of cognitive flexibility observed in humans. Thus, while the neural network architecture in AI models is fixed after training, the human brain requires neuroplasticity to adapt to new information from the world (Northoff; Gouveia, 2024, p. 13).

However, such scientific evidence is insufficient and faces a discussion that runs through the development of AI. Communication about AI innovation shapes certain expectations and relies on certain imaginaries, which play a fundamental role in the concrete development of AI and its implementation in society, as highlighted by Romele (2022, p. 2). The visual communication of AI, such as the excessive use of the color blue, recurring themes like androgynous faces, brains half flesh and half circuit, among others, seem to promote the necessary scientific confidence for certain hypotheses to be socially approved and therefore tested (Romele, 2022, p. 2-3).

In the work Digital Habitus: a critique of the imaginaries of Artificial Intelligence, Romele formulates three main theses regarding the relationship between AI and the formation of new imaginaries about its role in the social environment: the first, that current technologies are formidable habitus machines, since they offer increasingly personalized services, but are indifferent to individuals and their personalities; the second, that the concrete capacity of these technologies also depends on the expectations, hopes, and fears we have regarding them and their capabilities; and finally, that we should not only analyze things in themselves, but also the symbolic conditions of possibilities in which individual technological artifacts are always embedded. Romele (2023, p. 3) points out, for example, that digital bubbles reduce our global environment to a few stimuli and make us good consumers: "The certainty that our digitally mediated behaviors are predictable is more important for machines and their owners than wealth and variability. But this, one might say, ends up impoverishing our perception of the global environment and the worlds of others (Mitwelt) and the self (Selbstwelt)".

Digital habitus, in turn, is the moment when individuals are systematically reduced to general classes of action and preference, from which targeted contents are offered. According to Romele (2023, p. 3), the effects of this subjectification, built from repeated contact with these technologies, would end up flattening the self to these generic tendencies. In this sense, digital machines are habitus machines "because they actively and autonomously produce social classifications and categories – usually based on previous human classifications – that, to the extent that they are translated into forms of algorithmic curation, are incorporated and embodied in individuals" (Romele, 2023, p. 3-4). In other words, Romele's position is interesting because it precisely indicates what we are trying to show in this work, namely, that digital technologies produce both a new topography of cognition, as well as, when incorporated into the body, begin to act and produce new forms of interaction with the world.

Therefore, digital machines, or rather, all technology built on the sign of AI, has an ethical bias and a type of scientific responsibility that simply transcends technical and methodological questions. Automated intelligent decisions are decisions that impact equity, privacy, and justice and can only be thought of in this way because they extend over them. Keeping this in mind, it is possible to visualize that algorithms are not transparent mirrors of reality, just as they are not representational artifacts, disembodied and dislocated from the world. They are extensions of how we, over time, have been able to incorporate perceptions, values, and social ramifications.

At this point, then, we need to note that AI, assuming that cognition is radically disembodied from the body and the world, proposes to be capable of building intelligent systems completely neutral to ethical and social implications. By assuming this, it asserts the exclusion of the richness and complexity of human experience, reducing it to data and patterns that can be interpreted and processed by algorithms. However, AI and algorithms reflect and perpetuate the social, cultural, and political structures that inform them.

Readings of extended cognition (*Extended Mind Theory*) have shown, here specifically the pioneering work of Clark (2007), a broad range of arguments against the classical view that has shaped Cognitive Sciences and AI in recent decades, mainly by arguing that cognition is not merely a result of personal preferences, suggesting that external elements such as tools, language, and even digital devices play an active role in cognition.

Clark (2007) proposed a conception of the self based on the premise that human beings are "open ecological control systems", i.e., they seek opportunities both within the body and in the environment for solving their problems. The self, then, is all this "cognitive machinery", a "larger problem-solving set", formed by these "selves", agents, and cognitive engines. It is in this sense that Clark understands what he calls a "biotechnologically hybrid self", since our selves would always be an accumulative effect of the various resources that make up that problem-solving set. As he states, the self, i.e., its user, "is what we see (in others and ourselves) when all of this is working properly: a more or less rational being seeking a more or less unified set of goals and projects" (Clark, 2007, p. 112).

Our sense of unity, then, arises simultaneously, on the one hand, as a kind of hallucination that gives us a sense of cohesion beyond concrete reality and, on the other hand, as a sense whose construction is derived from the development of the body's sense of boundary, spatiality, and agency, from which, after reaching a certain level of stability, the organism begins to work to protect itself. Our tendency to narrate our own history and actions, or "narrative impulse", in turn, emerges both as a tool for agency and for the aforementioned sense of unity, and as another factor that contributes to nurturing the illusory image of a "central user".

Thus, considering Clark's arguments, the manifest self that emerges, a story told in first person, is only the functional movement carried out by the way the brain, body, and environment have been evolutionarily coupled. This new image, therefore, repositions the mark of cognition, among which the production of a central consciousness and even the role of emotions. This means that the story encapsulated by the absence of a central self does not need to be told by a clever homunculus; as our biological understanding of the brain increases, more reasons appear for the self to be taken only as the autobiographical drawing that appears to organize our substantial capacity for memory and reasoning. Therefore, if cognition is a process that extends beyond the limits of the body and is socially engaged with the environment, then building AI models that disregard such assumptions seems to be completely distant from the human way of perceiving things and experiencing the world.

Although the thesis of extended cognition is considered a radical form of externalism about the mind (Wilson, 2013), since for cognition to occur properly what is inside is often complemented by what is outside, we need to consider that the addition of resources (artifacts) to a cognitive system, whether human or artificial, is not enough to dissolve the problem we are tackling. There is an important methodological difference between, on the one hand, being an extended cognitive agent and, on the other hand, adding functional elements to an intelligent system in order to functionally improve it. In the first case, we can add more intelligent technologies so that the agent can perform specific tasks, with the execution of a task enhanced by the dynamic interaction between the agent itself and the expanded environment. In the second case, the addition of intelligent technologies continues to confine other forms of bordering with the world to the artifact itself (Peruzzo, 2022; Peruzzo; Stroparo, 2023; Peruzzo; Karasinski, 2023).

It is precisely from the previous debate that one can visualize the existence of a much larger epistemic problem: the distribution of cognition to the body and the environment, therefore, represents a challenge to how AI should be able to position itself taking into account both agents and the active role of artifacts and the world. And this, in turn, can have significant ethical implications if we analyze issues such as agency, responsibility, and decision-making. A deep understanding of this redistribution of cognition not only redefines the traditional limits of Artificial Intelligence but also demands a critical reflection on the ethical foundations that guide its operation and are capable of producing a variety of narratives.

4 Exploring narratives of cyberimmortality: between bytes and eternity

While extended cognition retrieves the body and the environment in defense of its own self-constitution, narratives of the new cognitive — or the digital cognitive — that emerge among the various possible readings of the idea of AI (what McCarthy defined back in 1956 as the science studying the emulation of human intelligence behavior through machines) follow another direction: it is a cognition that is not at a point in space, but throughout it; it does not possess a body, except when it needs to perform movement in the world; its learning capacity is a function whose design is to provide any response as quickly as possible. This is a cognitive characteristic that now transcends sex or gender, group or language, being above biological movements and cultural situation.

Digital cognition, then, can be compared to a chameleon capable of being the environment but not reducing itself to it, as it assumes a form as long as that same form is coupled with some practical functionality. In this regard, then, the following question arises: what is the gender of the numbers or algorithmic functions that make up a computational line responsible for sustaining AI projects? This is an example of a question whose semantics make no sense, akin to what Carnap expressed in Overcoming Metaphysics through the logical analysis of language: "What is the average weight of people in Vienna whose phone number ends in 3?" The absurdity of this latter question reflects the significant difference between describing events in the environment and incorporating the environment and its artifacts into the processes of producing the description itself. It seems clear that, to date, AI projects have not been able to observe the possible implications of the subtlety of this difference.

Among the various features of this new topography of digital cognition are the narratives of Cyberimmortality, that is, the idea that traces left in the digital world may continue to be present and tangible in the future. More than an aporia, Cyberimmortality is the direct consequence of this cognition without organs and without a world, since the project is based on the motto that technology will lead us without a body to a life without end. Cyberimmortality, therefore, reveals itself as a horizon where existence extends beyond the contours of flesh and the world, completely challenging the dichotomy between the organic and the algorithmic. In this panorama, Cyberimmortality emerges as an epistemic unfolding that dissolves the boundaries of corporeality and the very idea of virtuality, transforming human experience into a narrative that transcends physical limitations. But what is the basis for this project?

As Pablo García-Barranquero (2021) argues, the advance of science and technology in the last five decades is opening unprecedented horizons. The advent of the Singularity, that is, the moment when all advances in science and technology would cause unimaginable biological, cultural, and social changes, would be pushing back the boundary of death. In the Singularity, then, there would be no distinction between humans and machines, or between the physical world and the virtual one. For this reason, many advocates of the Singularity would endorse some version of transhumanism. Thus, "for some transhumanists, while the body is simply 'jelly' (...), our minds can be transferred to a computer (which is eternally functional) and thus achieve digital immortality" (García-Barranquero, 2021, p. 180).

Seeking to introduce an agenda to the immortalist fallacy, García-Barranquero (2021) points out that there is a distinction between the fact that a person normally does not want to die and that the same person never wants to die. According to him, the term "normally", used by authors like Bostrom and Hauskeller, does not clarify whether it means "ideally," "generally," "most of the time," or "most people". In any case, the idea of digital immortality is a form of life extension in which we would live forever and never die, that is, a life that would no longer be tied to biological limitations. This transhumanist project, then, focuses on the thesis that our hardware, like the human body, will be disposable, as what will remain forever is our software, further enhanced by technology.

According to Swan and Howard (2012), the idea of digital immortality would also imply the existence of new ethical dilemmas that are complex in new ways. For example, even though murder is wrong in our society, but not in a future society where replicas abound, would they have the same legal, ethical, social rights? Furthermore, if we consider a society where mind uploads are as "valuable as genuine human beings, how could we distinguish someone from their duplicate?" (SWAN; Howard, 2012, p. 248). Obviously, the issue of digital immortality is not only a question linked to technological advances, as it also raises fundamental questions about identity, ethics, and social justice. Thus, if I were duplicated, and if each of us (me and my replica) had our own memories, existential qualities, desires, and emotions, could we assert that this replica still perceives itself as me? This portrayal of the identity problem, among many others, seems to evoke a profound reflection on what it really means to be human and how our conceptions of identity can be challenged and redefined in the face of new narratives produced by technology.

For instance, in October 2023, we saw the disclosure that an American startup was able to reproduce the voice of the late actor Edward Herrmann, who passed away 10 years ago, to be used in new audiobooks. Around the same time, the Disney company, on the other hand, released a statement affirming the recreation of Robin Williams' voice as the genie from the lamp, originally voiced by the author, to celebrate a 100-year special of the company. But what does all this really show us? Perhaps the answer lies in the constitution of another question: why is life as such, naked in all its aspects, no longer able to seduce us? Why does the real no longer enchant us like the digital narratives produced by technology? Or better yet, why do we become disenchanted with the world even before understanding the locus we occupy in it and the role the environment plays upon us?

The narratives of Cyberimmortality are endorsed, on the one hand, by the belief that the Singularity will once and for all surpass the mitigated human nature and, on the other hand, that the transhuman future will lead us to the moment of adjusting our bodies with new non-biological antennas (Kurzweil, 2005). Martine Rothblatt, in the work Virtually Human: The Promises—and Perils—of Digital Immortality (2016), for example, reminds us that one of the issues of this scenario is the fact that these new technological instruments have profound and direct implications on the nature of identity and on the transgressions of biology, although the creation of mindclones may provide a form of transcendence of biology and offer the possibility of continuous life in a digital medium. Such transhumanist metaphors, therefore, seem to have captured both public imagination and futuristic circles that see in the expression "We are the Dr. Frankenstein of our lives" the radical transformation of our existing present with technology (Lorrimar, 2019, p. 192).

A final point deserves to be revisited for closure. The protagonism of transhumanist narratives of Cyberimmortality still lies in the belief that each of us will have a personal "exocortex" in the cloud, that is, a kind of third non-biological cerebral hemisphere. The exocortex would be in continuous communication with the other two biological hemispheres, creating a symbiotic network between the human mind and its digital counterpart. The optimistic hypothesis about the coexistence between biological and Artificial Intelligence, however, faces challenges related to the harmonious integration and control of this interface and ignores potential risks of dependence, manipulation, or even loss of individual autonomy.

It is precisely at this moment that the narratives of Cyberimmortality quickly appear and suggest that a network of devices or computational systems may infinitely expand our human cognitive capacities, ignoring, for example, that the very ideas of progress and technological development are trapped in the pitfalls of realism and scientific instrumentalism. Obviously, this debate is at the heart of the transhumanist agenda, even though this movement is formed by different schools of thought (democratic transhumanism, libertarian transhumanism, extropianism, etc.), as mentioned by Francesca Ferrando (2024, p. 32). In all of them, human enhancement is seen not only as a human impulse, "but as the human duty per excellence: to be human means constantly overcoming limits and opening up new possibilities" (Ferrando, 2024, p. 32).

5 Final considerations

The discussion on AI encapsulation reveals the tendency to ignore the crucial role of the body and the environment in human cognition. By reducing intelligence to an encapsulated system, there is a risk of underestimating the complexity of the interaction between mind, body, and environment. This raises questions about the nature of intelligence and autonomy in artificial systems, especially regarding moral responsibility, agency, and the incorporation of technological artifacts.

Moreover, the conception of Artificial Intelligence as an encapsulated system raises questions about the attribution of responsibility in contexts where the autonomy of artificial agents is highlighted. The lack of clarity about who is responsible for the actions of these autonomous systems presents significant ethical challenges, especially when it comes to making high-risk decisions that affect individuals and societies. Another important aspect is the relationship between AI and human cognition. While AI is often seen as an extension of human intelligence, it is crucial to question whether delegating tasks to AI can actually enhance our own intelligence or whether, on the contrary, it can make us less intelligent by removing cognitive challenges.

The emergence of the concept of autopoiesis and the rise of extended cognition theories challenge the boundaries between the cognitive, the body, and the environment, suggesting a more dynamic and situated understanding of intelligence. However, while AI seeks to achieve ethical and social neutrality, it inevitably reflects and perpetuates the cultural and political structures that inform it. The work of authors such as Romele (2023) and Bonini and Treré (2024), for example, offer important insights into the imaginaries and philosophical foundations of AI and the role played by algorithms, highlighting the importance of considering both technical and ethical aspects in the design and implementation of these technologies.

In this sense, transhumanist narratives of Cyberimmortality not only envision the future of humanity, where the boundaries between the organic and the digital become increasingly blurred, but also serve to articulate the Promethean utopia where technology can infinitely expand our cognitive capacities. The fact is that the pursuit of digital immortality, by ignoring issues such as identity, autonomy, body, and the environment jeopardizes the very constitution of a fusion between the biological and the technological. And this question, in the 21st century, as Francesca Ferrando rightly recalls (2024, p. 32), needs to consider our own behavior that is shaping the planet Earth and affecting all forms of life, including ourselves: "One of the main existential risks for humans as a species is our own behavior, such as uncontrolled anthropocentric habits that are leading the planet to an ecological collapse."

With that in mind, if there are any boundaries to the narratives of Cyberimmortality, they certainly lie beyond the poetic (not to say scientific) imagination that was able to reduce cognition to a monolithic block dissociated from the world and insensitive to the political transformations that surround it. Against this, once again, Heidegger's exhortation is valid — that thought dies where science is born, indeed, this kind of science that was able to turn human experience into a conceptualized, mathematical, and cold artifact.

Referências

BALES, A.; D´ALESSANDRO, W.; KRIK, G.; CAMERON, D. 2024. Artificial Intelligence: Arguments for Catastrophic Risk. In: *Philosophy Compass*, pp. 1-13.

BONINI, T.; TRERÉ, E. 2024. *Algorithms of Resistance: The everyday fight against platform power.* Cambridge: The MIT PRESS.

CHALMERS, D. 2010. The Singularity: A Philosophical Analysis. In: *Journal of Consciousness Studies*, **17**(9-10): pp. 1-56.

CLARK, A.; CHALMERS, D. 1998. The Extended Mind. In: *Analysis*, **58**(1): pp. 7-19.

CLARK, A. 2007. Soft selves and ecological control. In: ROSS, D.; SPURRETT, D.; KINCAID, H.; STEPHENS, G., L. *Distributed cognition and the will: individual volition and social context*. Cambridge: The MIT Press, p. 101-122.

CLARK, A. 2004. *Natural-born Cyborgs*. Minds, Technologies, and the future of human intelligence. New York: OUP.

CLARK, A. 2010. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford: Oxford University Press.

FERRANDO, F. 2024. To be or not to be enhanced? Just ask me Moon – in posthuman terms. In: JOTTERAND, F.; IENCA, M. (Eds.). 2024. *The Routledge Handbook of the Ethics of Human Enhancement*. New York; London: Routledge.

GARCÍA-BARRANQUERO, P. 2021. Transhumanist Immortality: Understanding the Dream as a Nightmare. In: *Scientia et Fides*, **9**(1): pp. 177-196.

KONIG, P. D. *et al*. 2022. Esse of AI. Wha tis AI? In: DIMATTEO, L. A.; PONCIBÒ, C.; CANNARSA, M. (Eds.). *The Cambridge Handbook of Artificial Intelligence: Global Perspectives on Law and Ethics*. Cambridge: Cambridge University Press.

KONIGS, P. 2022. Artificial Intelligence and responsibility gaps: what is the problem? In: *Ethics and Information Technology*, **24**(36): pp. 1-11.

KURZWEIL, R. 2005. *The Singularity is Near: When Humans Transcend Biology*. New York: Penguin Books.

LAUWAERT, L. 2021. Artificial Intelligence and Responsability. In: *AI & Society*, **36**(3): pp.1001-1009.

LICKLIDER, J. C. R. 1960. Man-Computer Symbiosis. In: *IRE Transactions on Human Factors in Electronics*, **1**: pp. 4-11.

LORRIMAR, V. 2019. Mind uploading and Embodied Cognition: a theological responde. In: *Zygon*, **54**(1): pp. 191-206.

MATURAMA, H.; VARELA, F. 1980. *Autopoiesis and Cognition*: the realization of the living. London: D. Reidel Publishing Company.

MINSKY, M. 1986. *The Society of Mind*. New York: Simon & Schuster.

MULLER, V. C. (Ed.). 2013. *Philosophy and Theory of Artificial Intelligence*. Berlin: Springer.

NORTHOFF, G.; GOUVEIA, S. S. 2024. Does Artificial Intelligence exhibit basic fundamental subjectivity? A neurophilosophical argument. In: *Phenomenology and the Cognitive Sciences*, pp. 1-22.

NYHOLM, S. 2024. Artificial Intelligence and Human Enhancement: Can AI Technologies Make US More (Artificially) Intelligent? In: *Cambridge Quarterly of Healthcare Ethics*, **33**(1): pp. 76-88.

PELLEGRINO, G.; GARASIC, M. D. 2020. Artificial Intelligence as extended minds. Why not? In: *Rivista Internazionale di Filosofia e Psicologia*, **11**(2) pp. 150-168.

PERUZZO JÚNIOR, L. 2022. *As Múltiplas faces da Realidade: Percepção, Linguagem e Cognição*. Curitiba: Editora CRV.

PERUZZO JÚNIOR, L.; STROPARO, A. 2023. Dissolving the Self: The Cognitive Turn of the Extended Mind Theory. In: *Trans/Form/Ação*, Marília, **46**(2): pp. 193-214.

PERUZZO JÚNIOR, L.; KARASINSKI, M. 2023. Cognitive Artifacts and Human Enhancement. In: *Ethics in Science and Environmental Politics*, **23**: pp. 45-52.

RICH, E.; KNIGHT, K. 1991. *Artificial Intelligence*. 2ª Ed. McGraw-Hill.

ROMELE, A. 2022. Images of Artificial Intelligence: a Blind Spot in AI Ethics. In: *Philosophy and Technology*, **35**(4): pp. 3-19.

ROMELE, A. 2023. *Digital Habitus: a critique of the imaginaries of Artificial Intelligence.* London: Routledge.

ROTHBLATT, M. 2016. *Virtualmente Humanos: As promessas – e os perigos – da imortalidade digital.* São Paulo: Cultrix.

SAVIN-BADEN, M; BURDEN, D; TAYLOR, H. 2017. The Ethics and impact of digital immortality. In: *Knowledge Cultures*, **5**(2): pp. 11-29.

SUSSER, D. 2013. Artificial Intelligence and the Body: Dreyfus, Bickhard, and the Future of AI. In: MULLER, V. C. (Ed.). *Philosophy and Theory of Artificial Intelligence.* Berlin: Springer, pp. 277-287.

SWAN, L. S.; HOWARD, J. 2012. Digital Immortality: Self or 0010110? In: *International Journal of Machine Consciousness*, **4**(1): pp. 245-256.

TAYLOR, I. 2024. Collective Responsibility and Artificial Intelligence. In: *Philosophy and Technology*, **37**(1): pp. 1-18.

VARELA, F; THOMPSON, E; ROSCH, E. 1992. *The Embodied Mind*: Cognitive Science and Human Experience. Cambridge: MIT PRESS.

WILSON, R. A. 2014. Ten questions concerning Extended Cognition. In: *Philosophical Psychology*, **27**(1): pp. 19-33.

Submetido em 31 de março de 2024.
Aceito em 19 de dezembro de 2024.