General artificial intelligence: singularity versus anti-singularity and a way out of the theoretical impasse

Inteligência artificial geral: singularidade versus anti singularidade e uma saída para o impasse teórico

Luís Filipe Estevinha Lourenço RODRIGUES

Professor Adjunto C na Universidade Federal do Ceará.
Doutorado em Epistemologia e Filosofia da Ciência.
Universidade de Lisboa, UL, Portugal.
com período sanduíche em University of Edinburgh
E-mail: luisestevinha@ufc.br

ABSTRACT

This article briefly describes central ideas for and against the generic notion of the singularity, an alleged moment in time when artificial intelligence could equal or even supplant human intelligence. It is argued that these opposing theoretical positions tend to cancel each other out and that current technological progress supports the prospect of a very powerful type of artificial intelligence, but one that is still functionally distant from the type proposed by the heralds of the singularity.
KEYWORDS: Artificial Intelligence, Paradigm, Singularity, Anti-singularity.

RESUMO

Descrevem-se resumidamente neste artigo ideias centrais a favor e contra a noção genérica de singularidade, um alegado ponto no tempo no qual a inteligência artificial poderá igualar ou mesmo suplantar a humana. Defende-se que esses posicionamentos teóricos contrários tendem a anular-se mutuamente e que o atual progresso tecnológico apoia a perspectiva de um tipo de inteligência artificial muito potente, contudo ainda funcionalmente distante do tipo proposto pelos arautos da singularidade.
PALAVRAS-CHAVE: Inteligência Artificial, Paradigma, Singularidade, Anti-singularidade.

Kalagatos
Revista de Filosofia

EDIÇÃO ESPECIAL
2024

V.21, N.2.
e-ISSN: 1984-9206

1. artificial intelligence and paradigms: singularity versus anti-singularity

Almost 75 years after the publication of Alan Turing's seminal paper (Turing, 1950) that opened the contemporary discussion on the alleged thinking capacity of computational artifacts (Saygin et al., 2000), the debate pitting partisans against skeptics of strong artificial intelligence continues to rage in our century (Russell *et al.* 2010, Eden *et al.* 2012, Frankish *et al.,* 2014). The formulations of the primary questions surrounding artificial intelligence date back to the famous Dartmouth Research Project (1956), where the term artificial intelligence and the generic concept associated with it seem to have been elaborated.[1] However, the human intuition about intelligent artefacts and the debate about this possibility seems to go back a long way, being traceable, for example, to the famous conceptual experiment of the mill proposed by Leibniz.[2] The philosopher intuited the impossibility of intelligent mechanisms, no matter how well they could be engineered. No mechanical articulation could create consciousness and other mental properties that he believed belonged exclusively to our species.

A long road has been traveled in this field since Turing's and Dartmouth's proposals. Today, artificial general intelligence (AGI) is based on artificial neural networks with an enormous capacity for probabilistic and recursive selection and interpretation of information. Following the so-called connectionist model of neural intelligence (Minsky, 1991), these entities are as complex as they are opaque, increasingly hiding the processes, neural layers and recursive algorithms on which they are based from the eyes of their creators (Domingos, 2015). Based on machine learning and deep machine learning systems (Kaplan, 2016), based on information from so-called *Big Data* (Mayer-Schonberger, 2013), an obscure, undefined and abstract entity, today's AGIs achieve results that are both extraordinary and worrying in terms of *replicating the* functionalities of the human mind and human behavior. Contemporary AGI has already surpassed in every way the digital discrete state machines envisioned by Turing.

The technological progress that has taken place over the last four decades has led some thinkers to predict the so-called advent of the *singularity*. In the opinion of some of the most ardent oracles, it would occur at a *singular* point in the not-too-distant future. In it, machines would acquire intelligence equal to or greater than that of human beings, with all that this would imply in terms of capacity for thought, consciousness, cognition, creativity, etc. The literature that examines these AGI phenomena, the singularity of machines and their derivatives (*e.g.,* Saimoli *et al.,* 2020; Frana *et al.,* 2020), refers to three

---

[1] The founding fathers of the project were philosophers and scientists J. McCarthy, M. Minsky, N. Rochester and C. Shannon.
[2] *Discourse on metaphysics,* and other texts, 2004, p. 133.

general types of problems. Firstly, can AGI become as intelligent, conscious and autonomous as *sapiens*? Secondly, what are the foreseeable consequences for humanity and its way of life if the singularity occurs? And thirdly, how can the AGI and its builders be regulated in the present with a view to the future, especially with the singularity hypothesis on the table? None of these problems are new. All three are subdivided, creating a vast array of other conceptual and practical problems that are interconnected and generally inseparable.

The focus of this article is on the first question and the stance taken on singularity. But its central purpose is not to carry out a detailed inspection of the debate surrounding it, much less to provide an answer. The aim of the exercise is simply to show how the most prominent answers that have been tried to these questions form two opposing explanatory paradigms. On the one hand, the singularity paradigm (hereafter PS), and on the other, the anti-singularity paradigm (henceforth PAS).[3] These parent paradigms are subdivided into explanatory sub-paradigms whose central ideas, I will argue, tend to cancel each other out in terms of the explanation they propose for singularity in the sense described above.

In the next section, the second of this article, I establish a transtemporal analogy between a classical parable and the current state of the art on the topic of singularity. The purpose of this analogy is to point out stereotypes and similarities in attitudes towards singularity. In the next section, the third, I list the most relevant sub-paradigms of PS and PAS, highlighting the key ideas they contain. In the fourth and final section, I argue that these tend to cancel each other out from the point of view of demonstrating the future. I conclude by presenting a vision of the problem and the future of AGI that rejects both singularity and anti-singularity. This glimpse of a possible future is closer to prudential than assertive.

2. A transtemporal analogy.

It is a huge triviality to say that we learn to deal with the future by taking into account lessons from the past. We incur in another triviality, perhaps even greater than the previous one, when we say that our theoretical positions depend on the social, historical, scientific, philosophical etc., contexts in which we are inserted. But the triviality of these observations shouldn't stop us from using them, especially when they turn out to be an appropriate *modus operandi* for understanding complex situations with unpredictable outcomes. *Pace* Hume[4], causal inferences drawn from facts and historical lessons are valuable tools for conceptually assessing the future, at least humanly speaking. And all the more so because we live in a world articulated by increasingly undeniable ontological connections because they

---

[3] Examples can be found in Awret et al (2016). Born (1989/2018). In Section 3 we will analyze specific cases of each paradigm.
[4] David Hume, *An Enquiry concerning Human Understanding,* p. 44.

Kalógatos
Revista de Filosofia

EDIÇÃO ESPECIAL
2024

V.21, N.2.
e-ISSN: 1984-9206

3

are more scientifically discernible. Today's world can no longer be seen as causally disconnected, as the illustrious Scottish philosopher proposed. However, and agreeing with Hume on this point, such recognition does not authorize us to take causal inferences for granted. So, assuming that we are here on the right methodological track, a reasonable balance between the past, present facts, and the merely possible, we move on to the first side of the analogy promised above.

Luís Vaz de Camões (1524-1580) eternalized in Portugal's greatest epic work, Os Lusíadas, the antagonistic ideals of two stereotypes regarding Portuguese maritime ventures aimed at discovering new routes and territories from the 15th century onwards. On the one hand, the masterful poet placed the stereotype of the Old Man of Restelo. The term "old man" metaphorically denoted a group of individuals who were old-fashioned, culturally outdated, and, above all, skeptical of the success of maritime ventures. I'll call this stereotype *old men of the castle*. In Camões view, the old men of the castle were representative of a sector of the population that cursed Portuguese maritime ventures; in essence, people who prophesied great misfortune and misfortune for the entrepreneurs and the Lusitanian homeland. These old men were presented by the poet as the bearers of an old-fashioned ideology that supported the closure of the Portuguese nation to the world. The poet gives us the idea that these skeptics were narrow-minded, but not stupid. We imagine them owning vegetable gardens in "the rest of Lisbon" (hence the name "Restelo"). They can also perhaps be understood as symbols representing a feudalistic or manorial class clinging to their properties and fiefdoms, ancestrally obtained through charters. This stylistic figure of the old men of the castle therefore symbolized a considerable slice of Lusitania (according to Bacchus, the mythological founder of Lusitania, according to André de Resende) that was backward and lacked vision for the future. As Camões explains in his neoclassical epic, endangered by the fear of hecatombs and territorial and financial losses, not to mention fears and beliefs about falling off the world when they reached the extremities of an Earth that was still flat in the minds of many, these figures with their heads stuck in the sand, like ostriches, but vociferously, criticized those who dared to set off to discover unknown or poorly known areas, such as the India of those times.

On the other side of the ideological barricade, Camões placed the young barons who embarked on voyages of discovery: captains and other officers. For the poet, the term "barons" symbolizes, among other things, a group of noble adventurers imbued with a desire for maritime exploration, as well as an avidity for trade and obtaining wealth. Some, not a few, were fearless commanders, instrumentalized by the Portuguese Crown of the time, belonging to the higher social classes, but sometimes bankrupt. The aforementioned heroes generically portrayed by Camões form a second stereotype. The second and third children of families that were moderately well-off, but from whom they would inherit little or nothing, neither land, nor money, nor outstanding social status, these daring explorers, by their less favorable birth

EDIÇÃO ESPECIAL
2024

V.21, N.2.
e-ISSN: 1984-9206

Kalsgatos
Revista de Filosofia

4

circumstances, had little or nothing to expect from a lapsed country. Fed up with a still medieval society, hostage to patrician, monastic, and petty mentalities, they were willing to take a risk. These barons whose coats of arms were later marked were emigrants from the Atlantic, the Indies, and even the Pacific. They were daring but reckless. They were heroes, but also murderers. They were victorious but also defeated. Many couldn't imagine the impact their actions would have on the future, although Camões was well aware of it, we believe. Perhaps the world would be much worse, or perhaps it would have been better, if they hadn't emigrated out of necessity, greed and a desire for glory.[7] But we can only imagine, we can't know. These are counterfactuals whose plausibility or implausibility is not easy to discern.

This is one side of the analogy: skeptical conservatives *versus* daring discoverers. This side of the analogy serves as a structuring starting point, as the theme is recurrent and trans-temporal. Although the objective content of this first side of the analogy differs greatly from the content of the second side, we already have the clear idea that, however brilliant they may be, even poets only know the future *a posteriori*. Camões embellishes his past more than his present.

Another extraordinary phenomenon of discovery is currently taking place. We are witnessing a time of searching for new paths and mysterious territories, both conceptual and practical. We are traveling through never-before-seen seas of science, technology and philosophy. We do so at the whim of driving winds that we can hardly control, but which inexorably move our theoretical and technological ships towards new possible worlds, areas that are as unknown as they are uncertain. Our current instruments are very different from those used by our ancestors. The "sea-charts", "astrolabes" and "sails" used today in scientific and technological research are undoubtedly very different in structure, capabilities, and purposes from those of the ancients. So are the nature of our discoveries and objectives. Nevertheless, the spirit of inquisitiveness and discovery remains the same, albeit greatly amplified.

On the other hand, the criticism of navigation is not very different in form and content from what it was in earlier times. The contemporary advent of the AGI, something that many of us still don't dare to describe, is a sea that is as unknown, if not more so, in terms of its dangers and potential as the Indian and Atlantic oceans were in the days of the first Portuguese discovery expeditions when people sailed in caravels. The curious thing is that, just as in those ancient times, the confrontation with dangers and unknowns, with the adamastors erected by the AGI, generates exactly the same kind of fears in modern times as it did in those remote times. Many still fear falling off the Earth, pushed into the abyss by the AGI.

So yes, on one side of the barricade we still have old men of the castle, fierce skeptics of new conceptual and technical possibilities and their consequences. On the other side, we still have bold barons, but probably just as reckless as the originals. However, as the social, economic, and cultural state

of affairs has changed greatly since those early days of discovery, it no longer makes sense to use this typology and its nomenclature. We now call each other helmsmen. We do this for two reasons. The first is that they steer the course of the ship of public opinion. The second is that, just like the ancient helmsmen of the sea, who, despite having an approximate view of the course, were unable to avoid pitfalls, storms, pirates, hostile populations, beliefs and other contingencies that could jeopardize their sailing, today's helmsmen cannot guarantee that the outcome of their predictions, opinions, theories and desires will or will not be as they advocate.

So, to start again with the conservatives, the old men of the castle have now become conceptual helmsmen, philosophers and scientists, theoretically skeptical of the powers and potential of the strong AGI. These old men are not necessarily as old as the original allegorical figure of the old man of the castle in the *Lusíadas*. They are just prudent thinkers who reject technological figures they consider almost mythological and unattainable, in other words, they reject *strong AGI, in* the *Serlian* sense of the attribute of fortitude (Searle, 1980). Some fear the possible current and future impacts caused by these still fuzzy figures, characters that some skeptical helmsmen consider conceptual illusions that are very unlikely to be realized and implemented.

These philosophical helmsmen constitute a family or group of venerable and illustrious thinkers, a think tank *of* the philosophical elite, tacit and sometimes organized, an *intelligentsia* that militates around theoretical-conceptual ideals against a strong AGI. For theoretical and prudential reasons, they adopt a pessimistic outlook, especially in opinion pieces published on social networks or in articles in the traditional *media.* There, they "vociferate", in the manner of contemporary old-timers, their criticisms, especially of generative AGI technologies (henceforth AI-gen), denying them human-like intelligence.[5] From the perspective of these illustrious conceptual helmsmen, a brave new world is now being discovered and explored, with a *Huxelian* dystopian critique of the future inserted *a la carte*. It's a world that they understand like few others from a theoretical point of view, but which they reject from a pragmatic point of view. Later on, when we look at the sub-paradigms of the PAS, we'll see concrete examples of this type of skeptical and pessimistic helmsman concerning the AGI. In other words, examples of skepticism and pessimism about whether the AGI "ship" will take *homo sapiens* to a safe port.

On the other side of the barricade, sailing away from these skeptical waters, we find two versions of the audacious barons. The first version is made up of captains or chief mates of the AGI industry or adjacent industries. They are commanders of the industries of science and technology, as well as the

---

[5] Two examples stand out. The first is Noam Chomsky's critical article in the New York Times on March 8, 2023. Available at: https://www.nytimes.com/2023/03/08/opinion/noamchomsky-chatgpt-ai.html. A well-known critic and skeptic in Brazil of the IAG is the famous neuroscientist Miguel Nicolelis. Some of his opinions are available at: https://www.youtube.com/watch?v=pb4b4_MlNwo&t=14s. (2023)

contemporary, capitalist and globalized free market.[6] They are hopeful discoverers and trailblazers of new technological or entrepreneurial worlds. They are as fearless as they are reckless, both in action and opinion. Many of them don't measure, publicly at least, the possible consequences of their navigations. And those who do, often do so hypocritically. Not a few of them are *ethically* libertine sailors who sail on bumpy business seas, doing so by sight or guided by ideological stars, always hoping to achieve wealth, prestige and fame, progress and happiness. Not infrequently, these captains of technology, of the ideal of progress and opinion, conditioned by the boards of directors and shareholders of their corporate ships, fail to consider, philosophically speaking at least, the dangers to be expected from their sailing. Worse still, sometimes they do, but they ignore what they learn in their intimate reflections.

Famous captains of this strain of almost unconditional supporters of this AGI progress are Sam Altman, Bill Gates, Mark Zuckerberg, Elon Musk, Sundar Pichai, Tim Cook, Larry Page, for example. Some of these *personas* harbor the duplicity of pioneering barons and captains, on the one hand, and old men of the castle, on the other.[7] Yes, these barons also wear the skin of the latter, as here and there they draw our attention to the dangers posed by AGI technology, while the companies they work for or are close to continue to produce AGI every minute, shamelessly and unchecked.[8] This new technological nobility is, so to speak, made up of specialists disguised by a very nefarious form of *akrasia.*[9] They know it's unwise to carry on, as they always do, with semi-transparent personal or corporate interest sales pitches blocking their clear and objective vision, replacing this desirable vision with diaphanous and often illusory perspectives. They do this shamelessly, without real self-censorship, in the name of economic, business, and unbridled progress ideologies; ideologies they pursue because they are motivated by capitalist economic ideals and globalization, the *sacrosanct path of* contemporary societies, both in the West and in the East. They do so boldly, yes, but recklessly. They hypocritically mask these procedures with pseudo-criticism of their products, a well-known *marketing* and advertising device more aimed at

---

[6] A list of this intelligence even appears in a famous volume of Time magazine:
Time100/AI, The 100 most influential people in artificial intelligence (2023). Available at:
https://time.com/6311323/how-we-chose-time100-ai.

[7] Sam Altman, for example, perhaps the leading figure and frontman of contemporary AGI, recently stated (May 2024), in a clear exercise of disclaimer, that the ethical control of IAG depends on the definitions of values at stake and the very definition of the term "value" that society adopts: https://www.youtube.com/watch?v=_Vik6UJTFyk.

[8] Floridi states this duplicity in his interview to Vilaça et al., 2024.

[9] See the contradiction of the combination of these stereotypes in the famous manifesto warning ordinary citizens about the concerns for our species because of IAG, referring to a pause in the production of AI-gen (March 2023). Available at: https://futureoflife.org/open-letter/pause-giant-ai-experiments. As far as we know, nothing has been halted in the process of researching and developing new iterations of AI-gen, quite the opposite.

strengthening their public images, which are often narcissistic, than at educating and guiding ordinary citizens.[10]

This is a problem that plagues many of those interested in thinking *philosophically about* the topic of AGI. As I see it, against the general view of a society that is so often uncritical and clouded by its new technological Pinocchio, AGI problems do not *conceptually* belong to computer and information science, or even to the technological industry that produces AGI. In other words, the topic doesn't belong to the aforementioned captains of industry, even though this idea is deeply ingrained in public opinion, especially on social media. Of course, against this generally uncritical opinion, founded more on soundbites and *clickbytes* than on facts or valid arguments, it is obvious that experts in execution do not necessarily constitute experts in thinking about the specialty executed. For example, it's doubtful that most soccer players can think philosophically about the sport. In the same way, it is questionable whether electronics and computing technicians, entrepreneurs, and the like, are automatically philosophers of artificial intelligence just because they do what they do.

The reasoning leads us to the second side of the analogy. The latter highlights facts that show a contemporary positional duality regarding the AGI and its expected consequences. This dualism is rejected by some (Dahlin, 2023), but we cannot and should not ignore it, as it reflects choices and defenses of paths to be taken. Given this framework, it is important to clarify what separates skeptical conceptual helmsmen, thinkers competent in the topic of AGI who reject and abhor the idea of singularity, and believing conceptual helmsmen, thinkers also competent in this topic who adopt or embrace this idea. The first point of division between them is descriptive: *can strong* AGI be realized? As I have suggested, not a few believing conceptual helmsmen who are optimistic about technological constructs accept this possibility, while others reject it outright.[11] The second point of division is normative: *should* a *strong,* superintelligent AGI be produced by the human species? Here too, the answers vary. On the *yes* side are those who believe that a super AGI can be kept under human control, as long as the builders and products comply with certain ethical rules. On the *no side* are those who believe that the control of a truly conscious, intelligent and free AGI is unlikely or even impossible. They claim that if such a superintelligent AGI were possible, it should not be built, let alone implemented.

3. Subparadigms for and against singularity: opposition and annulment of ideas.

---

[10] Famous and exemplary is the following statement by Elon Musk as reported by a Guardian journalist: "With artificial intelligence we are invoking the devil. In all the stories where there's the guy with the pentagram and the holy water, it's like, yeah, he's sure he can control the demon. It doesn't work."
https://www.theguardian.com/technology/2014/oct/27/elon-musk-artificial-intelligence-ai-biggest existential-threat.
[11] We will see cases of explicit acceptance in the next section. A clear example of this rejection, despite all the advances in technology, can be found in a joint intervention by Searle and Floridi in 2017: https://www.youtube.com/watch?v=b6o_7HeowY8.

The singularity, in a *broad* use of the term, which some attribute to Von Newman (Ulam, 1958; *apud* Eden, 2012, p. 4), but perhaps more based on physics than mathematics, is the idea of a point in time when technology will reach capacities that will radically and irreversibly change the way of life and the nature of *Homo sapiens*. In a *stricter* sense, the term indicates the arrival of a moment when AGI will eventually acquire full intelligence and consciousness, as well as semantic understanding, introspection, reflection, imagination and creativity that are equal, equivalent or superior in terms of functionality and results to those of human beings. We're interested in the latter.

It is in this sense of the expression "AGI *fortitude*" that artificial superintelligence has been classified (Bostrom, 2014, p. 50). AGI that does not meet the conditions listed in the previous paragraph is called weak AGI (Searle, 1980).[12] Enraptured by the conceptual impression of similarity and emulation of computational processes and results and neuro-cerebral processes and results, the believing conceptual helmsmen ended up predicting the equalization and even surpassing by AGI of human intellectual, cognitive, expository and artistic capacities. Although this vision of the parity or even intellectual superiority of machines is far from uncriticized, the idea influences popular culture about how these tools will shape the future (Broderick, 2001; Barrat, 2013; Frischman *et al.*, 2018; Harari, 2018).

A clear example of the modeling and adaptation of postmodern W.E.I.R.D.D.[13] societies to weak AGI are the so-called human natural language generative artificial intelligence tools (AGI-gen)[14], as well as artificial intelligence-derived devices for image, video and sound manipulation. True, there is rarely a day in our lives when we don't come across news of some new technological advance, the iteration of a process, its insertion into a portable *gadget,* and so on.[15] Nowadays, these devices and this state of affairs divide opinion within the philosophical *métier*.

With regard to strong AGI and its predictable evolution, PS and PAS are two major theoretical umbrellas. Both contain subdivisions, axes that I will call subparadigms. I will now mention at least three on each side of the barricade that divides perspectives on singularity. These sub-paradigms seem to be the most representative or closest to the main suggestion of the two parent paradigms that comprise them. I will now give a very brief overview of the central ideas of each axis or sub-paradigm, highlighting their opposition to the ideas of the opposing parent paradigm. I argue that this meeting of opposing ideas gives rise to an annulment of the *theoretical* power of each idea for or against singularity.

---

[12] See Max Tegmark's public and concerned manifestation of an alleged singularity currently underway in:
https://www.youtube.com/watch?v=_-Xdkzi8H_o
https://www.youtube.com/watch?v=xUNx_PxNHrY
[13] Western, Educated, Industrialized, Rich, Democratic, Digital.
[14] Large language models.
[15] A worrying example by Sundar Pichai, CEO of Google, can be found at: https://futurism.com/the-byte/ceo-google-ai-hallucinations. (5/2024).

KalōgatoS
Revista de Filosofia

EDIÇÃO ESPECIAL
2024

V.21, N.2.
e-ISSN: 1984-9206

9

Warning! It's important to note that the divisions and classifications presented here are very rudimentary, and probably don't do justice to the variety of theories that exist in the philosophical "market" about the possibility or impossibility of a super AGI.[16] For this reason, I will only mention the most classic and prominent authors who defend the central ideas of each sub-paradigm. Finally, something that should also be emphasized is the fact that the conceptual boundaries separating the subparadigms are blurred and diffuse. This is certainly helped by the fact that the thinking of the conceptual helmsmen behind these subparadigms cuts across various topics and subtopics related to AGI. However, as our purpose is only to present a minimal conceptual taxonomy of the most representative sub-frameworks in the debate on the question of singularity for the purposes of minimalist comparison, these divisions must be taken with a grain of salt by the reader. We have selected for discussion the theoretical axes that seem most emblematic of the essential ideas of the respective parent paradigm, and it is up to the reader to delve deeper into each of them according to their needs and interests.

In my opinion, the most prominent sub-paradigms of PS are the following: (I) Evolutionary sub-paradigm: progressivity, replicative capacity and functional improvement of artificial intelligence via emulation of the human brain (Turing 1950, Good 1966, Minsky 1991, Vinge 1993, Kurzweil 1999, 2005; Bostrom 2014, Tegmark, 2017). (II) Subparadigm of machine learning towards intelligence and consciousness (Turing 1950, Moravec 1988, Minsky 1994, Bostrom, 2014 Kurzweil 1999, 2005). (III) Subparadigm of the multiple realizability of thought in different physical media (Chalmers, 1996, mainly in section 9; or 2016).

Still from my perspective, the most prominent sub-paradigms of PAS are the following: (IV) Subparadigm of the absence of real learning, creativity, causal thinking, tacit thinking, counterfactual thinking or thinking about the future (Dreyfus 1972, Floridi 2014, Penrose 1994, Chomsky 2023, Nicolelis 2023, online[17]). (V) Subparadigm of the distinction between symbolic processing *and* semantic comprehension (Searle, 1980; 2014). (VI) Subparadigm of ontological and functional mind-machine dissimilarity (Born *et al.* 2018, Floridi 2015, Fjelland 2020, Bishop 2021, Landgrebe *et al.* 2023).

At a time of great fervor for the AGI and what it can give us individually and collectively, but also a time of great fear about the possible harmful consequences of its ever-increasing potential, what should be the theoretical stance towards the singularity given this amalgamation of ideas for and against the future realization of a super AGI? In other words, are the arguments on either side sufficient to

---

[16] For example, Weizenbaum (1975, p. 78) seems to have been one of the first ones to suspend his judgment.
[17] Miguel Nicolelis explains why AI-gen is neither intelligence or artificial. https://www.youtube.com/watch?v=Fw8fJxWhQX8&t=951s.

demonstrate that the singularity is inevitable or, on the contrary, that it will not happen (at least anytime soon)? I reject the alleged sufficiency of the ideas, alone or together, enunciated by any of the sub-paradigms listed to demonstrate the point of the parent paradigm. The premise that I believe supports this claim of insufficiency is the following: ideas that are *indirectly* contrary to each other tend to nullify the primacy or superiority of each one.

This is what I mean by the expression "indirect opposition". Think of the age-old idea that God is not the cause of evil, since it has to be attributed to humans because they use the freedom He has given them to cause evil. Based on this idea, some philosophers (e.g., Swinburne, 2004, p. 122). try to make the concept of a supremely good, omniscient, all-powerful Being, etc. incompatible with the idea of evil, referring the responsibility for moral evil to the (phenomenally) conscious, responsible and intentional *sapiens*. But, of course, we can think that God, although good, might have wanted humans to use their freedom to build the best of all possible worlds by doing moral evil on certain occasions, thus generating a better world. Then, if this were the case, and I'm not saying that it was, God's actions would not be absolutely incompatible, from a logical point of view at least, with moral evil; with him being the cause of this moral evil in the final instance, this to achieve a greater good.

I will apply this principle of reasoning by indirect opposition here. It is certainly not possible to show here the indirect opposition of most of the ideas in the subparadigms described above. But it is possible to use a revealing example whose form of opposition can be extrapolated to many other cases in this list. Here's an example. Minsky famously wrote:

> Many thinkers firmly maintain that machines will never have thoughts like ours because, no matter how we build them, they will always lack a vital ingredient. They call this essence various names - such as sentience, consciousness, spirit or soul. Philosophers write entire books to prove that, because of this deficiency, machines will never be able to feel or understand the kind of things that people feel. However, all the proofs in each of these books are flawed, because they assume, in one way or another, the very thing they intend to prove - the existence of a magical spark that has no detectable properties. (...) I have no patience for such arguments (Minsky, 1994).

Minsky's idea is in favor of sub-paradigm I of PS. Here's what Landgrebe *et al.* suggest:

> (...) the functioning of the mind resists mathematical modeling. Therefore, we cannot emulate the mind using a machine, nor can we design other types of complex non-machine systems to obtain hitherto undescribed types of non-human intelligence. (...) The aim of AGI research, and of those who fund it, is to obtain something useful, and this will imply that an AGI must meet certain requirements - in general terms, that it be able to deal with the reality in which humans live with a level of competence that is at least equivalent to that of human beings. We have shown that this is not possible, because there is an upper limit to what can be achieved by machines. This limit is set, not by the technical limitations of computers, but by the limits of what can be achieved by machines. This limit is set, not by technical limitations of computers, but rather by the limits of the possibilities of mathematical modeling. (Landgrebe *et al.*, 2023, p. 13).

This passage militates in favor of sub-paradigm VI of the PAS.

Note that the passages make equally plausible assumptions, unrelated to each other, but plausibly incompatible concerning the capacity of machines and the very possibility of a singularity. Minsky complains that critics of intelligent machines appeal to (mental?) properties unknown to human beings that machines allegedly cannot have. Landgrebe *et al.* rightly claim that intelligent machines cannot cross a functional barrier, one that humans do not possess, which allows humans, but not machines, to be conscious and intelligent. Thus, the proposals on both sides are not in direct contradiction: the phantom properties of humans that Minsky points out do not even relate to the barrier that Landgrebe *et al* impose on the machine. The hypotheses are both plausible and tend to cancel each other out indirectly. In other words, there is no apodictic way of declaring one a winner and the other a loser, only that they epistemically cancel each other out in terms of persuasive value concerning the topic they dispute.

A second image of this indirect opposition can be drawn from the antagonistic positions of Chalmers (1996, 2016, subparadigm III of PS) and Chomsky (2023, subparadigm IV of PAS). The first philosopher argues that intelligence and other intellectual properties can be realized in multiple physical supports, whether biological or non-biological, the famous principle of organizational invariance of the support for mind, intelligence, *qualia, etc.*:

> I argue that conscious experience results from a finely tuned functional organization. More specifically, I will defend a principle of organization in which consciousness is an organizational invariant: a property that remains constant in all functional isomorphs of a given system. It doesn't matter whether the organization takes place in silicon chips, in the population of China or in beer cans and ping-pong balls. As long as the functional organization is correct, conscious experience will be determined (Chalmers 1996, p. 245-246).

Chomsky doesn't *directly* contradict this idea, but he does point to a limitation in the replicability of human thought on the part of AI-gen. In his own words:

> ChatGPT and similar programs are, by definition, unlimited in what they
> (...) these programs are stuck in a pre-human or non-human phase of cognitive evolution. Their deepest flaw is the absence of the most critical capacity of any intelligence: to say not only what is the case, what was the case and what will be the case - that is description and prediction - but also what is not the case and what could and could not be the case. (Chomsky, 2023).

This implies that certain mental properties, such as the modal-mental elasticity and mental-conceptual plasticity of human thought and language, are not yet replicable in any non-biological so-called intelligent artifact.

Although indirectly antagonistic, but not necessarily contradictory, these two perspectives presented above tend to cancel each other out when it decides singularity, its occurrence or non-occurrence. This is for the following reasons. Firstly, the principle of organizational invariance advocated by Chalmers is still (in 2024) only an unproven conceptual construct. The conceptual plausibility of the principle does not guarantee its scientific truth, as he admits. If Chomsky and others are correct, the

emulation of human thought by today's AI-gen reveals that some mental properties of humans are not yet instantiated, if ever, by non-biological systems. Secondly, even if Chomsky's claim is in order and tends to nullify Chalmers' construct now, it turns out that the idea of functional replication of intelligence contributes to nullifying the idea that machines will not be able to replicate human mental functions, as left open by Chomsky himself.

The sample presented here of indirect opposition of arguments and notions is certainly too short to apodictically demonstrate the mutual annulment of the ideas of the PS and the PAS. Nonetheless, it is a possible way to mitigate abusive theoretical speculation and its attempt to guess the singularity, or to support *ex ante* the necessary dismissal of this hypothesis and the advent itself. I leave it to the reader to carry out other exercises of comparison and indirect opposition based on the general framework presented and inspection of the relevant literature. Now, given the opinions present in that literature, I lean towards the idea that a forceful statement about this possible phenomenon or its denial cannot be vindicated based on any of the subparadigms. The reason is simple: there are equally plausible ideas about the possibility and non-possibility of singularity, and they tend to cancel each other out in terms of persuasive force.

4. A conceptual way out: weak-strong AGI

We see every day (at least those most interested in the subject) that the crucial question has long ceased to be what we want the AGI to be, or even what it will be. The question is how we can classify it for the near future and what that answer will mean in terms of consequences for us. The deadline given by Kurztweil for the singularity is worrying: 2030. Others, like Tegmark, reduce this *timeframe* even further, claiming that non-biological but intelligent entities like Hall 2000, *Terminator*, artificial intelligence from The *Matrix,* etc. are already here. But these apocalyptic predictions with exaggerated prognoses and extrapolations are just that for now. The time has come to stick to the facts and make a moderate, non-binding induction about the near future of the AGI with them in mind, avoiding excessive and dangerous generalizations. Thus, two intuitions, that are not conclusive, immediately stand out through inductive reasoning. The first is that a certain technological progress is undeniable - whether it is correct or desirable is another story. The second is the replication of natural evolutionary conditions, but now manipulated by *sapiens* - total emulation or total instantiation - and by the AGI itself for the replication of itself and the creation of conscious artificial intelligence of the same type as that of human beings seems to be an extremely complex and even unrealizable task in the near future. So, if this is correct, the dates sometimes put forward for the singularity may have been exaggerated, proving premature or even incorrect.

If our suggestion of mutual annulment is correct, does it mean the end of the singularity hypothesis? I believe that the above is by no means sufficient. More recently, some authors such as Bostrom (2024) and others have argued that artificial superintelligence and consciousness do not have to be of the same physical or structural type as their human counterparts, and can even be realized in very different ways. For example, virtually (Watanabe, 2022) through virtual neurons and connections with functionality similar to biological neurons; or even through artificial intelligence social networks. We won't explore these hypotheses or their feasibility here. However, the traditional image of hardware that runs symbols and cannot generate "semantic" intelligence (the so-called reductionist and computational view of AI) seems to us to be outdated, rejecting the concept of non-biological entities with high-level consciousness and understanding. More and more we seem to be moving towards these high levels of intelligence and understanding on the part of virtual devices. More and more we are getting the impression that we are facing a weak AGI, in the traditional computational sense of the term, but an increasingly strong one in terms of its capabilities and potential. It remains to be seen whether this weak AGI will prevent us from bending this cape of storms any further. Let's hope that the trivialization of human indifference, on the one hand, or panic, on the other, doesn't become too real.

5. Final considerations

In this article I have tried to address the controversy between philosophical positions favorable to the singularity hypothesis and philosophical positions opposed to it. I argued that current positions, ideas and arguments that try to support each side tend to cancel each other out, at least contingently, in terms of reasons and persuasive force. Instead of an obligatory choice between singularity or non-singularity, I chose to propose an unquestionable fact: the current progress of AGI militates in favor of an AGI that is still weak in terms of awareness and versatility, given the yardstick imposed by the thought and consciousness instantiated by human beings, but increasingly powerful in many respects.

References

AWRET, Uziel. CHALMERS, David (editors). *The Singularity Could artificial intelligence really out-think us* (and would we want it to? Imprint Academic Ltd. 2016 BARRAT, James. *Our final invention: artificial intelligence and the end of the human era.* NewYork, St. Martin's Press, 2013. eISBN 9781250032263.

BISHOP, John. Artificial Intelligence Is Stupid and Causal Reasoning Will Not Fix It.

Frontiers in Psychology, 2021. 11, 1-18 (online. doi: 10.3389/fpsyg.2020.513474)

BORN, Rainer (ed.). *Artificial intelligence: the case against.* New York. Routledge, 2018 (Reprinted. Original 1989).

BOSTROM, N. *Deep utopia, life and meaning in a solved world.* Washington DC. Ideapress Publishing. 2024.

BOSTROM, Nick. *Superintelligence: Paths, Dangers, Strategies.* Oxford. Oxford University Press. 2014.

BRODERICK, Damien. *The Spike: how our lives are being transformed by rapidly advancing technologies.* New York, A Forge Book - Tom Doherty Associates, LLC. 2021 CAMÕES, Luís, *Os Lusíadas,* Editora Melhoramentos, 1572/2014.

CHALMERS, David, *The Concious Mind.* New York. Oxford University Press. 1996. CHALMERS, David, The Singularity: A Philosophical Analysis. in "Could artificial intelligence really out-think us (and would we want it to)? In Uziel Awret, David Chalmers, eds., *The Singularity Could artificial intelligence really out-think us (and would we want it to*? Imprint Academic Ltd., 3-32. 2016.

CHOMSKY, Noam. The False Promise of ChatGPT. In The New York times. 2023, march 8).

DAHLIN, Emma. Think Differently We Must! An AI Manifesto for the Future. AI & Society. 2023.

DOMINGOS, Pedro. *The master algorithm: how the quest for the ultimate learning machine will remake our world.* Filadelfia. Perseus Group. 2015.

DREYFUS, Hubert., *What Computers still can't do: a critique of Artificial Reason.*, Massachusetts, The MIT Press Cambridge, 1972.

EDEN, Amnon; MOOR, James Johnny; SORAKER, Johnny, STEINHART, Eric.

Singularity Hypotheses, A Scientific and Philosophical Assessment. Springer. 2012. FJELLAND, Ragnar. Why general artificial intelligence will not be realized. Humanities and Social Sciences Communications. 7 (1): 1–9. 2020.

FLORIDI, Luciano. Devemos ter medo da inteligência artificial? https://aeon.co/essays/true-ai-is-both-logically-possible-and-utterly-implausible. 2015. FLORIDI, Luciano. *The Fourth Revolution: How the Infosphere is Reshaping Human Reality* Oxford, Oxford University Press, 2014.

FRANA, Philip L; KLEIN, Michael J. *Encyclopedia of Artificial Intelligence: the Past, Present, and Future of AI.* ABC-CLIO021. 2021. ISBN: 978-1-4408-5327-2. FRANKISH, Keith., RAMSEY, William (editors). *The Cambridge handbook of artificial intelligence.* Cambridge. Cambridge University Press. 2014.

FRISCHMANN. Brett; SELINGER, Evan. *Re-engenering Humanity.* Cambridge, Cambridge university Press. 2018.

GOOD, John. Speculations concerning the first ultraintelligent machine. In F. Alt & M.

Ruminoff (eds.), Advances in Computers, volume 6. Academic Press. 31-88. 1966. https://edoras.sdsu.edu/~vinge/misc/singularity.html. 1993.

https://publications.jrc.ec.europa.eu/repository/handle/JRC118163.10.2760/382730 (online).

HUME. David. *An Enquiry concerning Human Understanding,* Oxford. Oxford university Press. 2007. ISBN 978-0-19-921158-6.

KAPLAN, Jerry. *Artificial intelligence: What everybody needs to know.* Oxford, Oxford University Press, 2016.

Kalagatos
Revista de Filosofia

EDIÇÃO ESPECIAL
2024

V.21, N.2.
e-ISSN: 1984-9206

15

KURZWEIL., Raymond. *The Age of Spiritual Machines*. New York: Viking Press, 1999.

KURZWEIL., Raymond. *The Singularity is Near*. New York: Viking Press, 2005.

LANDGREBE, Jobst; SMITH, Barry. Why machines will never rule the world: artificial intelligence without fear. New York. Routledge, 2023.

LEIBNIZ, Gottfried. *Discurso de metafísica* (e outros textos), Martin Fontes. 2004 MAYER-SCHÖNBERGER, Viktor; CUKIER, Kenneth. *Big Data: A Revolution That*

Will Transform How We Live, Work, and Think. Boston, Eamon Dolan Book Houghtn Milffin Harcourt, 2013.

MINSKY, Marvin. Logical versus Analogical or Symbolic versus Connectionist or Neat versus scruffy. Magazine Volume 1991. 1 Number 2.

MINSKY, Marvin. Will Robots Inherit the Earth? Scientific American, Oct, 1994. http://www.inf.ufsc.br/~mauro.roisenberg/ine6102/leituras/Will%20Robots%20Inherit%20the%20Earth.htm. Transcrição consultada em 20/4/2024.

MORAVEC, Hans. *Mind Children, the future of robot and human intelligence*, USA, Harvard University Press. 1988.

NOAH, Yuval. *21 lições para o século 21*. Tradução Paulo Geiger. São Paulo. Companhia das letras, 2018.

PENROSE, Roger. *Shadows of the Mind, a Search for the Missing Science of Consciousness*, Oxford, Oxford University Press, 1994.

RAINER, Born (Ed). *Artificial Intelligence, the case against*. New York, Routledge, 2018 online (1987, primeira edição em suporte físico). ISBN: 978-1-351-14152- 9.

Russell, Stuart., Norvig, Peter. *Artificial intelligence, a modern approach*. New Jersey. Prentice Hall. 2010. Third Edition.

SAMOILI Sofia; LOPEZ COBO Montserrat; GOMEZ GUTIERREZ Emilia; DE PRATO Giuditta; MARTINEZ-PLUMED Fernando; DELIPETREV Blagoj, *Defining Artificial Intelligence*, EUR 30117 EN, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-17045-7, doi:10.2760/382730, JRC118163 (consultado) 22 Maio 2024.

SAYGIN, Ayse; CICEKLI; Ilyas & AKMAN, Varol. Turing test: 50 years later. Minds and Machines. 10 (4):463-518. 2000.

SEARLE, John. 1992. *The rediscovery of the mind*. Cambridge, MA: MIT Press.

SEARLE, John. Minds, Brains and Programs. Behavioral and Brain Sciences. 1980. 3: 417–424.

SEARLE, John. What Your Computer Can't Know. New York. Review of Books, October 9. 2014.

SWINBURNE, Richard. *The Existence of God*. New York. Clarendon Press. 2004 (Second Edition).

TEGMARK, Max. *LIFE 3.0: being human in the age of artificial intelligence*. New York. Alfred A. Knopf. 2017.

TURING, Alan. Computing machinery and intelligence. Mind 59. 433-60. 1950.

Kalagatos
Revista de Filosofia

EDIÇÃO ESPECIAL
2024

V.21, N.2.
e-ISSN: 1984-9206

16

VILAÇA, Murilo; KARANDISKY, Murilo; CANDIOTTO, Kleber. Reflexões sobre o futuro da inteligência artificial. Entrevista com Luciano Floridi. Unisinos Journal od philosophy, 2024.

VINGE, Vernor. The coming technological singularity. 1993. https://edoras.sdsu.edu/~vinge/misc/singularity.html.

WATANABE, Masataka. *From Biological to Artificial Consciousness, Neuroscientific Insights and Progress.* Springer, 2022. (electronic). ISBN 978-3-030-91138-6 (eBook). WEIZENBAUN, Joseph. *Computer Power and Human Reason.* Massachusetts. The MIT Press. 1976.