

How to think with Fernando Pessoa: Heteronymism as a model for critical thinking and multi-agent systems

[Como pensar com Fernando Pessoa:
Heteronimismo como modelo para o pensamento crítico
e sistemas multiagente]

Ana Ilievska*

Keywords

Fernando Pessoa, Heteronymism, Critical Thinking, Generative Artificial Intelligence, Multi-Agent Systems.

Abstract

This paper explores the relevance of Fernando Pessoa's heteronymism for thinking in the age of generative Artificial Intelligence, focusing on systems like ChatGPT. Drawing on Julian Jaynes's theory of consciousness and Hannah Arendt's reflections on thinking, I argue that AI, while offering cognitive enhancement, risks eroding critical reflection by automating judgment—echoing Jaynes's "bicameral mind" and Arendt's "great evil" of unthought. Pessoa's heteronymism provides a model for cultivating robust inner dialogue, where multiple perspectives generate meaning. Such plurality sustains human critical capacities and offers a framework for envisioning dialogical, multi-agent AI systems, resisting cognitive entropy and passive engagement.

Palavras-chave

Fernando Pessoa, Heteronimismo, Pensamento crítico, Inteligência artificial generativa, Sistemas multiagente.

Resumo

Este artigo explora a relevância do heteronimismo de Fernando Pessoa para o pensamento na era da Inteligência Artificial generativa, com foco em sistemas como o ChatGPT. Com base na teoria da consciência de Julian Jaynes e nas reflexões de Hannah Arendt sobre o pensar, argumento que a IA, embora ofereça formas de ampliação cognitiva, corre o risco de erodir a reflexão crítica ao automatizar o juízo – ecoando a "mente bicameral" de Jaynes e o "grande mal" da ausência de pensamento em Arendt. O heteronimismo pessoano surge, assim, como modelo para cultivar um diálogo interior robusto, sustentando a pluralidade crítica humana e inspirando sistemas de IA dialógicos e multiagentes.

* American University of Beirut – Mediterraneo.

The question that I want to ask is this: what can we learn in the twenty-first century from Fernando Pessoa's so-called "heteronymic machine" (MORRIS, 2014: 127)? My hypothesis is that we can learn how to think. Pessoa's invention of heteronyms—complete literary personae endowed with distinct voices, ideas, and styles—constitutes one of the most audacious experiments in modern literature, a dramatization of consciousness as dialogue with itself. His heteronymism¹ not only unsettles the boundaries between author and speaker but also stages thought as an inexhaustible plurality, a chorus of perspectives that contend, contradict, and converse. I suggest that Pessoa's heteronymism provides a model for cultivating internal dialogue and critical plurality at a time when generative Artificial Intelligence—most prominently represented by OpenAI's ChatGPT—threatens to automate and homogenize thought. In what follows, I place Pessoa's experiment in heteronymism in conversation with Julian Jaynes's account of consciousness as the breakdown of the bicameral mind and Hannah Arendt's understanding of thinking as an inner dialogue, to argue that Pessoa's model can guide us both in resisting cognitive entropy and in reimagining multi-agent AI systems that engage dialogically.

According to Jaynes, whose findings were based both on clinical studies and an original reading of Homer's epics, humans developed a robust sense of consciousness when they accepted that the motivation for their actions and thoughts did not come from the gods but originated in their own minds. Once the realization took place that what we hear in our minds is our own internal dialogue and not the orders of intrusive deities, we became autonomous thinking beings who act, make decisions, and can be held accountable. Arendt's position, developed in her essay "Thinking and Moral Considerations," just a few years before the publication of Jaynes's book, involves a similar concept of thinking as a "soundless dialogue (*eme emautô*) between me and myself, the two-in-one" which is the 'I' (ARENDT, 1971: 442). The constant practice of this internal dialogue, according to Arendt, is the first crucial step in preventing "great evil" from happening, or obversely, it is the suspension of the internal dialogue that can lead to great evil, as she had already concluded in her provocative article, and later book, on Eichmann's trial (ARENDT, 1963).

Now, I want to take this constellation of Pessoa's heteronymism,² Jaynes's theory of consciousness formation, and Arendt's concept of thinking to weigh upon

¹ Jerónimo Pizarro has recently made an excellent case for using the term "heteronymism" instead of "heteronymy," the former being indeed the original term used by Pessoa himself to describe the genesis of the heteronyms (PIZARRO, 2021: 34).

² Not to be confused with "heteronomy," a philosophical concept elaborated by Immanuel Kant denoting the opposite of autonomy—that is, a condition of dependence on the will or authority of something or someone external to oneself. Many thanks to the careful reviewer of this essay for suggesting a clarification and for the excellent recommendations about how to further deepen the argument, perhaps in a future essay.

some problems of human-machine cognitive distribution, i.e., on the division of cognitive labor between humans and machines in our times. Why is this combination of similar theories about the workings of consciousness and thought relevant at this moment? One less pressing reason is the rise of Artificial Intelligence and the related problem of Artificial Consciousness, which our lack of technical competence in the field prevents us from discussing in a legitimate way. However, if we want to create so-called thinking machines, then it is both the technologist's and the philosopher's task to understand whether and how they can become conscious. The more important reason behind the present reflections is that we want to understand what thinking means in the context of our engagement with AIs and whether insights into existing theories about human thinking practices can be useful in undertaking this task. It is important to raise the topic of consciousness and thought in artificial systems not because they are already capable of thought and have consciousness as many (wrongly) believe, but because humans increasingly relinquish their cognitive tasks—planning, decision making, thinking—to AIs. What is the impact of such human-machine cognitive distribution? In what way will it affect fundamental human skills such as reading, writing, listening, and speaking, not to mention our ability to interpret and critically engage with reality, other people, and the plethora of communications on the web and beyond?³

Thinking in our age (and here I follow in the footsteps of philosophers such as Kant, Heidegger, and Gadamer) is coextensive with language—that is, with reading, writing, listening, and speaking. As Heidegger wrote, “[l]anguage is the house of Being. In its home man dwells” (HEIDEGGER, 1993: 217). With the rise of language-based generative models, these practices—and consequently our capacity for thought—are undergoing a profound transformation. The house of language now seems to have acquired new tenants, ones that speak without dwelling.

Such systems assist us in structuring ideas, drafting and revising text, making choices and planning, thereby assuming part of our cognitive labor. However, rather than enhancing our well-being, as might be expected, studies and public sentiment suggest otherwise (see ROOSE, 2021; LEE *et al.*, 2025). The difficulty, in part, lies in our present inability to properly and critically harness its potential for human flourishing. What currently appears to many as a wondrous, even “magical” collaboration between human and machine is, in practice, edging us toward Kant's heteronomy if not towards an entropy of thought—a condition reminiscent of the “bicameral mind” described by Jaynes. It is as if the gods have returned, telling us what to do, what to choose, whom to vote, what clothes to buy, what people to date. “It is one god who makes Achilles promise not to go into battle,” Jaynes writes, “another

³ Since I first presented this paper in 2023, researchers at Carnegie Mellon University and Microsoft Research have shown that the use of GenAI in the realm of knowledge does, in fact, lead to a reduced capacity for critical thinking (LEE *et al.*, 2025).

who urges him to go, and another who then clothes him in a golden fire” so that the actions of Iliadic men “are not in conscious plans, reasons, and motives; they are in the actions and speeches of gods” (JAYNES, 1976: 72). Except that today, as Neil Gaiman had presciently observed in his *American Gods* (2001), our divinities are not called Athena, Zeus, Thetis, or Aphrodite, but Social Media, Algorithm, Google, and ChatGPT.

If we are to redirect human-machine cognitive distribution toward more beneficial ends, we must let the humanities furnish the compass by which this course is set.⁴ They offer methods for cultivating reflection and dialogue that can enrich both human thinking and machine performance. Some initiatives already gesture in this direction—for instance, the “Socratic Artificial Mind” (SAM Chatbot), designed to stimulate critical thought through Socratic questioning (CANEVA *et al.*, 2023; see also VAN HULLE, 2019; HOEBEKE, 2023; YANG and NARASIMHAN, 2023).⁵ My proposal, however, places particular emphasis on fostering internal dialogue—within humans and, by analogy, within generative models such as GPTs. Establishing such dialogic structures may nurture greater autonomy and accountability in thought, while also expanding our imagination of what forms of machine cognition might be possible, without presupposing the existence of consciousness in AI.

In this sense, I suggest that Pessoa’s unique heteronymic model can play an important role in two complementary ways:

- 1) as a training ground for humans, helping us resist the automatization of thought and cultivate habits of reflective, dialogic thinking;
- 2) analogously to the Socratic dialogue, as an inspiration for developers, offering a literary model for how multi-agent systems may be trained to engage dialogical forms of “thinking”.

Some efforts have already been made in this direction, primarily on a technical level, through multi-agent systems in which different LLMs (Large Language Models) dialogue or collaborate with one another to improve performance on complex tasks (cf. CHEN *et al.*, 2023; WU *et al.*, 2023; WANG *et al.*, 2024; JIANG *et al.*, 2023; ZHANG *et al.*, 2024). These projects, however, are not inspired by preexisting humanistic models, nor do they draw on literary or philosophical traditions of dialogue. By neglecting the humanities, these projects risk reducing dialogue to mere information exchange, overlooking the imaginative, ethical, and critical dimensions that make dialogue transformative rather than purely instrumental. What I propose here is, then, an interdisciplinary collaboration that harnesses our human heritage—literary, philosophical, and historical—to cultivate growth, imagination, and shared forms of thinking between humans and machines.

⁴ I began developing this line of argument in “*Frankenstein* is about ChatGPT: Thinking and the Future of Literary Study in the Age of Generative AI” (ILIEVSKA, 2024).

⁵ I am grateful to Prof. Giuseppe Palazzolo (University of Catania) for drawing my attention to the SAM project. I also wish to thank the anonymous reviewer who recommended the work of Dirk Van Hulle and Jonathan Hoebeke.

The danger posed by the rise of generative AI is not, I suggest, the apocalyptic scenario of machines annihilating humanity or enslaving us — visions entertained by Elon Musk or by the philosopher Nick Bostrom — but rather the quieter, more insidious risk that we ourselves relinquish the labor of thinking to a machine that does not think. And by “we” I do not mean scholars, politicians, or business leaders alone, but any human being with access to the World Wide Web. The abdication of thought has more than one precedent in history, and always with catastrophic consequences. Extremism, conformity, and bigotry flourish in the soil of un-thinking — that is, when we allow others to think for us, submitting to “conventional, standardized codes of expression and conduct” (ARENDT, 1971: 418). Heidegger named this force “*das Man*” in *Being and Time* (1927), transforming the impersonal German pronoun *man* — roughly “one” in English — into a noun that captures the anonymous collective shaping everyday life: “what one does,” “what people say.” Such codes offer protection from cognitive noise, overstimulation, even from reality itself. As D. N. Rodowick observes, “to fall back on opinion and conventional beliefs is a socially sanctioned way of protecting oneself against the variety and intensity of occasions that lay claim to critical thought and attention” (RODOWICK, 2021: 3). Yet overreliance on such codes suspends the very activity of thinking. This suspension is not limited to societies under the sway of ideologies or dictators; it thrives equally in liberal, individualistic contexts, wherever we consent to let someone — or something — do our thinking for us.

We are not, however, without resources to resist this tendency, quite the contrary. Pessoa’s heteronymism, Jaynes’s account of the bicameral mind, and Arendt’s conception of thinking as an “I-confronting-the-I” in the privacy of the mind⁶ all offer powerful blueprints for cultivating dialogicity, or polyphony, in the contemporary world, to borrow from Bakhtin. One implication is stark: thinking cannot take place on platforms such as Instagram, TikTok, or Facebook, nor in the division of cognitive labor between myself and ChatGPT, or myself and any technology, however prosthetically useful or embodied it may appear. For thinking is, first and foremost, a solitary act, only afterwards shared and public. Therein lies perhaps the deeper existential problem of our contemporary times that becomes sharper the more it is ignored: in the privacy of our minds, we have only ourselves to confront. This confrontation is often uncomfortable, even unbearable, which is why we so readily outsource it. Yet it is precisely here, in this silent and inescapable dialogue with oneself, that the very possibility of judgment, responsibility, and indeed freedom is grounded.

⁶ This is not to suggest that thinking is an exclusively private matter for a political thinker such as Arendt. Rather, as she elaborated in her *Lectures on Kant’s Political Philosophy*, private reflection is the first step, which then moves to a public confrontation with others in the collective practice of judgment — a loop that constitutes the art of critical thinking (see RODOWICK, 2021).

Among twentieth-century poets, Pessoa stands out as the one who seems to have taken these imperatives most literally, making the cultivation of dialogue and a radical polyphony the very ground of his poetic practice. He not only spent his entire life away from the marketplace, working on a deep ontology of the self, moderating his views, and avoiding certainty by ensuring the plurality of perspectives in his oeuvre; but he also mastered the art of the split self. There isn't just "another fellow" with whom to be in dialogue in his work, but a whole series of other fellows, each with his own *Weltanschauung*, style and character. The solitary mansion of Pessoa's thinking was, to say the least, extremely crowded. And these interlocutors were not only in an amicable relationship with one another, but they also challenged, inspired, and critiqued each other to the point of constituting, in effect, autonomous beings in their own right—as Pessoa himself maintained. This is why he spoke of heteronymism rather than mere pseudonym in the famous letter to Adolfo Casais Monteiro from January 13, 1935: "the mental origin of my heteronyms lies in my relentless, organic tendency to depersonalization and simulation. Fortunately for me and for others, these phenomena have been mentally internalized, such that they don't show up in my outer, everyday life among people; they erupt inside me, where only I experience them" ["a origem mental dos meus heterónimos está na minha tendência orgânica e constante para a despersonalização e para a simulação. Estes fenómenos – felizmente para mim e para os outros – mentalizaram-se em mim; quero dizer, não se manifestam na minha vida prática, exterior e de contacto com outros; fazem explosão para dentro e vivo – os eu a sós comigo"] (PESSOA, 2001: 254; PESSOA, 1986: 226). As Pessoa's remark makes clear, the heteronyms embody an endless internal dialogue much like the one that Arendt, via Socrates, had in mind—an interior multiplicity that resists the entropy of thought—offering a model for cultivating robust inner discourse at a time when the overreliance on generative AI threatens to erode our own capacities for thinking and writing. By adopting Pessoa's model, we can learn to inhabit plurality within ourselves, preparing us to engage with generative AI critically rather than passively, and thereby creating the conditions for critical thinking to truly flourish.

Much has been written about the origins and nature of Pessoa's heteronymism—variously interpreted as a modernist technique, a philosophical system, a symptom of personality disorder such as schizophrenia, or even a precursor of Deleuze and Guattari's rhizomatic thought. For Eduardo Lourenço, Pessoa's heteronyms were the manifestation of an "exploded consciousness" ["consciência explodida"] (LOURENÇO, 2002: 153); for Jorge de Sena, they were proof that "Pessoa never existed" ["Pessoa nunca existiu"] (SENA, 2000: 224);⁷ Adam Morris, in his enlightening essay "Fernando Pessoa's Heteronymic Machine", sees in Pessoa a philosophical thinker

⁷ Unless otherwise noted, all translations into English are mine.

whose “literary machine” was an exercise in endless “becoming-heteronym” (MORRIS, 2014: 142), anticipating theories of becoming and the nomadic subject. For Agostinho da Silva, writing in the 1950s, heteronymism acquired a political dimension: he read in Pessoa’s work a critique of European capitalism: through his playfulness and modesty, the Portuguese poet offered an alternative to Europe’s exploitative, profit-driven mentality (SILVA, 1959). Today, heteronymism might likewise serve as an antidote to the profit (and optimization) oriented logics of contemporary tech, offering instead humanistic alternatives for cultivating slow thought and critical thinking in the age of AI.

Various, predominantly U.S.-based, academics have referred to Pessoa’s writing as “automatic writing” (MULDOON, 2002: 40) that allows for an “endless round and round” within a complex dissociative process (MAHR, 1998) that generates endless discourse: “we are stories telling stories” [“somos contos contando contos”] the heteronym Ricardo Reis affirmed (PESSOA, 2015: 113). One group of scholars deny the heteronyms any individuality or empirical existence of their own apart from their creator, the empirical Pessoa; another group understands them as independent literary figures. Others yet accept the ‘plurality as unity’ interpretation (see Thurston’s “multiple singularity,” 2008)⁸ with Jerónimo Pizarro seeing in Pessoa “a refracting prism” [“um prisma que refracta”] (PIZARRO, 2018: 14). All these takes are correct in their own way and surely it is understandable from such readings that, whatever the true origins and nature of the heteronymism, Pessoa is “one of the decisive poets of the century, particularly if we try to think of him as a possible condition for philosophy” (BADIOU, 2005: 36). Indeed, I would even go so far as to agree with Badiou that it is “philosophy [that] is not—at least not *yet*—under the condition of Pessoa. Its thought is not yet *worthy of Pessoa*” (BADIOU, 2005: 36).

What some have called a “literary machine” and “automatic writing” and what Badiou sees as a philosophical scandal, I see as the ultimate antidote to unthinking and as the fulfillment of philosophy: the practice of thought par excellence. I believe that this is what António Mora, the “philosophical” heteronym, meant when he wrote that “philosophy is an anthropomorphism in all systems” [“a filosofia é um antropomorfismo em todos os sistemas”] (PESSOA, 1993c: I, 7). Philosophy, in other words, is a series of contrasting but dialoguing subjective takes on whatever given problem or phenomenon, eventually forming a necessary albeit fictional whole. At the same time, the heteronymism can be seen as a model of consciousness or, perhaps, several potential models, as Burghard Baltrusch has shown.

Baltrusch gives four possible ways of interpreting the heteronyms: 1) as one consciousness generating various “narratives” (heteronyms) that in turn influence and manipulate it; 2) as several forms of consciousness generating several “narra-

⁸ For a more detailed summary of the various positions, see BALTRUSCH (1996: 339f).

tives" (heteronyms); 3) as being three variants of two types of consciousness involved in the reconstruction of paganism; and 4) as one synoptic consciousness that "from a fundamental dialogic disposition generates various heteronymic narratives and eventually even enters in dialogue with its own fiction" (BALTRUSCH, 1996: 347). In this sense, the heteronyms appear "as an autistic model of communication" (BALTRUSCH, 1996: 347). Putting aside the neurological explanation of the heteronymism, the fourth and last model appears most in line with Jaynes's and Arendt's view on how consciousness and thinking work. Consciousness in the world of Pessoa is not a monologic narrative, but a synoptic dialogic constellation consisting of several narratives in perpetual movement; it is inherently plural and thrives on this plurality. Without consciousness, there cannot be thought; and without dialogicity, there cannot be critical thought. In this sense, Pessoa's heteronymism may also be read as a precursor to contemporary conceptions of such multi-agent systems as mentioned above, where autonomous yet interdependent agents interact to generate richer, more complex forms of knowledge and perspective.

If Pessoa, as Lourenço saw it, is first and foremost a "provocateur," then he has realized Socrates's dream of the never-ending dynamism of thought. This is "an indefinite continuous posing of undefinable problems" "ir pondo indefinido de problemas indefiníveis"] of which Alfredo Antunes spoke in "Fernando Pessoa e o problema do ser" (ANTUNES, 1962: 123). Mar Talegre calls it "a vast interrogation that jumps out at every moment, from walking in the world, from encountering things, real events" ["uma vasta interrogação que salta a cada instante, dos seus passos pelo mundo, do seu encontro com as coisas, com os acontecimentos reais"] (TALEGRE, 1947 in ANTUNES, 1962: 124).

Thinking is, first and foremost, a constant questioning that, however, leaves the one who questions empty-handed: "The need to think," Arendt writes, "can be satisfied only through thinking" (ARENDT, 1971: 422). No clear outcomes can be expected, no definite conclusions about the good life, but, as in the Socratic elenchus, a dialectical back-and-forth which "guarantees that the goals pursued will be seen clearly in relation to one another, and crucial issues will not be missed by haste and inadvertence" (NUSSBAUM, 2012: 49). As Bernardo Soares writes in the *Livro do Desassossego*, "to think is to live and to feel is nothing more than food for thought" ["pensar é viver e sentir não é mais que o alimento de pensar"] (PESSOA, 1982: I, 237). Pessoa's philosophy of Sensacionismo ("Sensationism") merits a future in-depth analysis, but for now let us note that, for Soares, actions and feelings are the stuff of thought, but not vice versa, so that whoever practices the art of thought is always variously engaged while remaining empty-handed, like an adroit juggler with several balls in the air and yet none fully in her fist: "Each time that I think about something, I am thinking about something else" ["E sempre que estou pensando numa coisa estou pensando noutra"], Álvaro de Campos writes (PESSOA, 1993: 26d).

There is an echo of this proposition in Arendt when she affirms that “an object of thought is always a re-presentation” for “thinking always deals with objects that are absent, removed from direct sense perception” (ARENDT, 1971: 423). Thinking about the representation of something in a poem written by a heteronym is the absolute maelstrom of thought. Pessoa is never certain about who is looking, who is thinking, and what is being thought about. The thinking gaze wanders from the real object to the represented object and back and does not dwell at any point except at the question itself: “I look at the Tagus in such a way | that I forget that I am looking | and suddenly this hits me | in my reverie — | What is it to be a river and to run? | What is me looking at it?” [“Olho o Tejo e de tal arte | Que me esquece olhar olhando | E súbito isto me bate | De encontro ao devaneando — | O que é ser-río, e correr? | O que é está-lo eu a ver?”] (“Abismo”, PESSOA, 1984: 36). Furthermore, there is always a question of who the conscious other is who does the seeing and observing. Who is “the other fellow” participating in the internal dialogue? Who is “das fremde Ich” (“the foreign ‘I’”) to borrow Georges Güntert’s words? “Whose gaze | Peeks through my eyes? | While I think that I am looking, | Who keeps looking | While I think?” [“De quem é o olhar | Que espreita por meus olhos? | Quando penso que vejo, | Quem continua vendo | Enquanto estou pensando?”] (“De Quem é o Olhar”; PESSOA, 1995: 64).

“Being is tiring, feeling hurts, thinking destroys”, Pessoa affirms in a poem from 1921, “Cansa ser, sentir dói, pensar destrói” (PESSOA, 1990: 31), as if stung by the Socratic torpedo fish. For Pessoa, as for Socrates, to think is to dismantle, to destabilize, to introduce aporia and confusion where previously certainty reigned. But just as we want to put into question the certainty of this epistemic uncertainty, Alberto Caeiro’s presence announces itself with his infantile gaze and the refusal to think in an eternal interplay of changing perspectives between himself, Pessoa-Ipse, Ricardo Reis and Álvaro de Campos: “I feel multiple,” we read in the *Páginas Íntimas*, “I am like a room with countless fantastic mirrors that twist into false reflections one single prior reality that resides in none and yet in all” [“Sinto-me múltiplo. Sou como um quarto com inúmeros espelhos fantásticos que torcem para reflexões falsas uma única anterior realidade que não está em nenhuma e está em todas”] (PESSOA, 1966: 93). In this sense, Pessoa’s universe is not so much a rhizome or a machine as it is a tesseract cube, a perpetually moving yet coherent distortion of space and time, of the “I” and its thoughts. This is the dream of Socrates, who preferred death to the tyranny of stasis. It is the realization of what Bakhtin had called the “dialogic imagination.”

But the dream of Socrates and Pessoa’s reality are the nightmare of Artificial Intelligence. For Artificial Intelligence, in its more concrete and currently hyped form as generative AI, is based on prediction and statistics; it is derivative and aims at probability based on pattern recognition. This for a very simple reason: AI is not alive. It has no consciousness nor hidden or repressed desires and metabolic needs.

Unlike Pessoa, it does not answer to its own metabolism. There are no internal mirrors doubling its “self” because it simply does not have one; it does not know fear, desire nor finitude. Its existence is a series of on/off/on/off states. As Italian philosopher Maurizio Ferraris writes, this is the crucial distinction between a mechanism and an organism or between an automaton and a soul (or between ChatGPT and Fernando Pessoa): for an organism “life is characterized by the irreversibility of a metabolism, an absolute on/off that differs radically from the serial on/off of a mechanism” (FERRARIS, 2023: 71; see also FERRARIS, 2024 and 2025). Organisms respond to the needs of the metabolism and aim at sustaining themselves and deferring death. For an AI to be truly intelligent and think, it would need to double itself internally of its own free will and accord and create “the other fellow” with whom to converse as a way of making decisions and coming to terms with its mortality. However, currently, ChatGPT and generative AIs in general are still what Julian Jaynes had called “bicameral minds,” that is, they do not have an internal interlocutor and thus no accountability but rather receive their orders “from above”, from external agents. These external agents are humans, the real gods in the machine.

This is where Pessoa’s work can be of use if we want to train generative AIs in the art of thinking. We read again in *Páginas Íntimas*:

For some temperamental reason that I do not intend to analyze, nor does it matter if I do, I have built within myself various characters who are different from each other and from me, characters to whom I have assigned various poems that are not how I, in my feelings and ideas, would write them.

[Por qualquer motivo temperamental que me não proponho analisar, nem importa que analise, construí dentro de mim várias personagens distintas entre si e de mim, personagens essas a que atribuí poemas vários que não são como eu, nos meus sentimentos e ideias, os escreveria.]
(PESSOA, 1966: 106)

This is the essence of thought (and perhaps the essence of what it means to be human) as practiced by Socrates and celebrated by Arendt.

Pessoa accepts that living with oneself is always living with another, with many others:

I do not know who I am, what soul I have. When I speak sincerely, I do not know with what sincerity. I am variously other than an ‘I’ that I do not know whether it exists (if it is these others). I feel beliefs that I do not hold. I am filled with yearnings that I repudiate. My perpetual attention to myself perpetually points out mental treacheries to a character that perhaps I do not have, nor does it think I do. I feel multiple.

[Não sei quem sou, que alma tenho. Quando falo com sinceridade não sei com que sinceridade falo. Sou variamente outro do que um eu que não sei se existe (se é esses outros). Sinto crenças que não tenho. Enlevam-me ânsias que repudio. A minha perpétua atenção sobre mim perpetuamente me aponta traições de alma a um carácter que talvez eu não tenha, nem ela julga que eu tenho. Sinto-me múltiplo.]

(PESSOA, 1966: 93)

In a similar vein, Pessoa-Ipse writes: “I don’t know, but I feel | Estranged from my being” [“Não sei, mas meu ser | Tornou-se-me estranho”] (“Ao longe, ao luar;” PESSOA, 1995: 105) and even the usually consistent Caeiro confirms this internal disidentification or pluralization: “Nor am I always the same in what I say and write” [“Nem sempre sou igual no que digo e escrevo”] (PESSOA, 1993a: 54).

This art of doubling, mastered by Pessoa, is the distinguishing feature of the human mind according to Jaynes, for whom the mind is a:

[...] secret theater of speechless monologue and convenient counsel, an invisible mansion of all moods, musings and mysteries, an infinite resort of disappointments and discoveries. A whole kingdom where each of us reigns reclusively alone, questioning what we will, commanding what we can. A hidden hermitage where we may study out the troubled book of what we have done and yet may do.

(JAYNES, 1976: 1)

But ChatGPT is far from having such an internal dialogue or let alone a mind. It does not need to “study out the troubled book of what [it] has done and yet may do”. It has no personal accountability. As Noam Chomsky writes:

The crux of machine learning is description and prediction; it does not posit any causal mechanisms or physical laws. Of course, any human-style explanation is not necessarily correct; we are fallible. But this is part of what it means to think: To be right, it must be possible to be wrong. Intelligence consists not only of creative conjectures but also of creative criticism.

(CHOMSKY *et al.*, 2023)

While human thought and, in its most extreme form, Pessoa’s form of thinking, is a continuous course correction without a final destination—fallible, dynamic, renewable—generative AI is an attempt at designing a definite map of the journey or, in other words, at stabilizing the human, channeling it into a repeatable, teachable, transferable form through statistics and prediction.

At this historical moment, generative AI is aimed at the automatization of human thought. Why? For the same reason Pessoa continually described thinking as laborious: “who has a soul has no calm” [“quem tem alma não tem calma”] (PESSOA, 1993: 48). Thinking (which involves discernment, decision-making, and planning) just like labor (which involves the exertion of physical force needed to make something) is strenuous. Álvaro de Campos also acknowledged this when he wrote of thinking as a prison house: “Prison of thought, is there no escape from you?” [“Cárcere do pensar não há libertação de ti?”] (PESSOA, 1993: 94). Unless done for leisure, thinking, just as physical labor, can be outright painful. It is then not surprising that one of the main goals of human civilization until the twentieth century has been to reduce the physical strain on the body through the creation of technologies

that will decrease or obviate the need for physical exertion. Since the twentieth century and the accompanying advancements in digital technology, we have shifted our attention to the mind. Now, our foremost interest lies in extending our mental capacities or, more correctly, in reducing the strain on our minds and attention. We want to create technologies that will make it unnecessary or easier for us to think, plan, and make decisions.

In technical terms, the dual effort to simplify our mental and physical lives is known as automation. Marshall McLuhan calls this a strategy of equilibrium whereby “with the arrival of electric technology, man extended, or set outside himself, a live model of the central nervous system itself” (MCLUHAN, 1964: 43). Automation offers some respite, yet despite its promise to alleviate mental and physical burdens, it does not solely assist in repetitive tasks and heavy industrial labor. “We are not getting happier at work,” Kevin Roose writes, “despite the fact that our jobs are safer and less grueling than ever” (ROOSE, 2021: 16; see also KORNBLUH, 2023). If automation does not increase happiness, if machines can supplant or augment our physical abilities, and AIs can supplement or replace our cognitive functions, surpassing us in efficiency and speed, then what remains of our human essence as thinking and laboring beings? Here lies the pressing question for Arendt, tailored to our contemporary concerns and spirit of the times: Can significant harm arise from cognitive automation and its associated technologies? To put it more plainly, can ordinary individuals wielding these technologies unleash great harm? Without resorting to the nationalist/patriotic interpretation, Agostinho da Silva’s reading of Pessoa’s work as anti-capitalist feels timely in this context. Can we think of Pessoa’s work as the guardian of human thought in the age of speed-thinking and optimization? Pessoa, after all, accepts the “doença”, the malady of thought, the prison-house of thought. Can one man’s prison-house become the twenty-first century’s liberating force? I will conclude by offering some specific proposals.

First, we need to cultivate spaces of solitude where the internal dialogue can germinate and become robust. But when we speak of solitude, we speak of one of the most insidious problems of Western society today. Norwegian philosopher Lars Svendsen writes that “the main problem we are facing today is not that loneliness is on the rise, but rather that solitude is too scarce” (SVENDSEN, 2017: 17). According to Svendsen, rather than suffering from loneliness, we suffer from “hypersociality” (SVENDSEN, 2017: 106) because of the omnipresence of communication devices and social media in our lives. Even when we return home in the evening after work or school, and even if we live alone, we are online, perpetually entangled in other people’s stories, words and images.

I believe that literature plays a crucial role in the activation of thinking because literature is the only human artifact that, to repeat Arendt’s words, “deals with objects that are absent, removed from direct sense perception” (ARENDT, 1971: 423) while being perfectly intelligible to all because of its reliance on everyday language.

Álvaro de Campos is of the same mind when he writes that: “E sempre que estou pensando numa coisa estou pensando noutra.” Furthermore, unlike philosophy, literature is also entertaining, which offers the first “hook”, but it is infinitely more stratified and subjective, as Pessoa’s work demonstrates, than any other form of entertainment. Works of literature, as literary critic Amir Eschel writes, are “an attentive reservoir of the human experience” (ESCHEL, 2020: 44), which is inherently plural and contradictory.

Second, we need to create public spaces—a modern Socratic marketplace, an “ágora”—where the internal dialogue can be externalized and confronted with that of others. Thinking might be timeless and untimely, but it is logistically bound to time and space just as the individual human beings who engage in thinking are. The university and the classroom already have in place the spatiotemporal infrastructure necessary for the practice of thought. However, currently the activity within these institutional spaces is determined by the drive for professionalization and the output of visible products in the manner of the “hard” sciences. The humanities, in particular, suffer greatly from this emphasis on what John Guillory calls “the professional profile” (GUILLORY, 2022: 42). Rather than a training camp for the practice of form and method, the humanities classroom should be our Socratic marketplace for the cultivation of moral judgment and Pessoa-style critical thinking.

This is not to say that students should be taught to distinguish between “good” and “bad” novels or poems. There is no such mandate in Pessoa’s work. Rather, the type of moral judgment that we have in mind is a process or praxis that allows us to investigate time and again what it means to be human.⁹ And this work can be excellently cultivated through the study of Pessoa’s works because it is in his heteronymic machine that complex, diverging perspectives, not bound by a single system of thought, are constantly negotiated. And why not? In collaboration with generative AIs, we can even try to create our own heteronymic chatbots, following Pessoa’s model.

At the same time, generative AIs and similar future systems can be trained on existing humanistic models of dialogue and consciousness such as the Socratic dialogue, the Bakhtinian dialogic imagination, and Pessoa’s heteronymism as proposed in this paper. Perhaps in this way not only can humans practice thinking in the privacy of their minds and judgment in dialogue with AIs, but AIs themselves can be trained through the creation of heteronyms, each in dialogue with one another, each subjecting the other to a Socratic elenchus, so that the internal doubling necessary for consciousness and critical thought may happen. As Maria Irene Ramalho writes, Pessoa is “the most remarkable example of the art of rumination in Western modernism” [“o mais notável exemplo da arte da ruminação no modernismo ocidental”]

⁹ See Markus Gabriel, *I Am Not a Brain* (2017: 2): To be human is to “constantly work out a historically shifting and culturally varying account of what exactly it takes to be the kind of minded animal we are.”

(RAMALHO, 2021: 159]. If anything, Pessoa's work is the greatest archive of internal dialogues ever put to paper. The time is now ripe to put them back into the mind, whether human or artificial.

Bibliography

- ANTUNES, Alfredo (1962). "Fernando Pessoa e o Problema do Ser." *Revista Portuguesa de Filosofia*, vol. 18, n.º 2, pp.123-154.
- ARENDT, Hannah (1971). "Thinking and moral considerations." *Social Research*, vol. 38, n.º 3, pp. 417-446.
- (1963). *Eichmann in Jerusalem: A Report on the Banality of Evil*. New York: Viking Press.
- BADIOU, Alain (2005). *Handbook of Inaesthetics*. Translated by Alberto Toscano. Stanford: Stanford University Press.
- BALTRUSCH, Burghard (1997). *Bewusstsein und Erzählungen der Moderne im Werk Fernando Pessoa's*. Frankfurt am Main: Peter Lang.
- CANEVA, Gianfranco; CALCAGNO, Salvatore; GIORDANO, Daniela (2023). "Socratic artificial mind (SAM): Lessons from the evaluation of a GPT-3 based chatbot for Socratic dialogue." *Sistemi intelligenti*, n.º 2, August, pp. 413-434. DOI: 10.1422/108138.
- CHEN, Justin; SAHA, Swarnadeep; BANSAL, Mohit (2024). "ReConcile: Round-Table Conversation among LLMs Improves Reasoning." *arXiv preprint*. <https://arxiv.org/abs/2309.13007>
- CHOMSKY, Noam; ROBERTS, Ian; WATUMULL, Jeffrey (2023). "Noam Chomsky: The false promise of ChatGPT." *The New York Times*, March 8. Cf. <https://www.nytimes.com>
- CLARK, Andy (2008). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford: Oxford University Press.
- DAMASIO, Antonio (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Harcourt.
- ESCHEL, Amir (2020). *Poetic Thinking Today: An Essay*. Stanford: Stanford University Press.
- FERRARIS, Maurizio (2025). *La pelle. Che cosa significa pensare nell'epoca dell'intelligenza artificiale*. Roma-Bari: Laterza.
- (2024). *Webfare: A Manifesto for Digital Well-Being*. Translated by Ana Ilievska. Bielefeld: Transcript.
- (2023). "Archeologia dello Spirito." *Filosofia Morale / Moral Philosophy*, vol. 3, pp. 55-80. DOI: 10.7413/fmmp0024.
- GABRIEL, Markus (2017). *I Am Not a Brain: Philosophy of Mind for the 21st Century*. Translated by Christopher Turner. Medford, MA: Polity.
- GAIMAN, Neil (2001). *American Gods*. William Morrow: New York.
- GUILLORY, John (2022). *Professing Criticism: Essays on the Organization of Literary Study*. Chicago: The University of Chicago Press.
- GÜNTERT, Georges (1971). *Das fremde Ich: Fernando Pessoa*. Berlin and New York: de Gruyter.
- HEIDEGGER, Martin (1993). *Letter on Humanism*. In: *Basic Writings*, edited by David Farrell Krell. San Francisco: Harper San Francisco.
- (1976) [1927]. *Sein und Zeit*. Frankfurt am Main: Klostermann.
- HOEBEKE, Jonathan. (2023) "En écrivant comme Beckett: A dialogue between generative AI and genetic criticism." *Short Paper to DHB2023*. Online at: <https://zenodo.org/records/8003714>
- ILIEVSKA, Ana (2024). "Frankenstein is about ChatGPT: Thinking and the Future of Literary Study in the Age of Generative AI." *Status Quaestionis*, n.º 26 (special issue on *Theory / Teoria: Perspectives on Literary Theory and Comparative Literature*, edited by Mimmo Cangiano et al.), pp. 179-206. DOI: 10.13133/2239-1983/18773.
- JAYNES, Julian (1976). *The Origins of Consciousness in the Break-Down of the Bicameral Mind*. New York and Boston: Houghton Mifflin.
- KORNBLUH, Anna (2023). *Immediacy or, The Style of Too-Late Capitalism*. London and Brooklyn: Verso.
- LEE, Hao-Ping (Hank); SARKAR, Advait; TANKELEVITCH, Lev; DROSOS, Ian; RINTEL, Sean; BANKS, Richard; WILSON, Nicholas (2025). "The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects from a Survey of Knowledge Workers." *Proceedings of the 2025 CHI Conference on Human Factors in Computing*

- Systems (CHI '25), Association for Computing Machinery. April 26-May 1, 2025, Yokohama, Japan. DOI: 10.1145/3706598.3713778.
- LOURENÇO, Eduardo (2002). *Poesia e Metafísica: Camões, Antero, Pessoa*. Lisboa: Gradiva.
- MAHR, Greg (1998) "Pessoa, Life Narrative, and the Dissociative Process." *Biography*, vol. 21, n.º 1, Winter, pp. 24-35.
- MCLUHAN, Marshall (1964). *Understanding Media: The Extensions of Man*. Cambridge, MA: MIT Press.
- MORRIS, Adam (2014). "Fernando Pessoa's Heteronymic Machine." *Luso-Brazilian Review*, vol. 51, n.º 2, pp. 126-149.
- MULDOON, Paul (2002). "In the Hall of Mirrors: 'Autopsychography' by Fernando Pessoa." *New England Review* (1990-), vol. 23, n.º 4, Fall, pp. 38-52. <https://www.jstor.org/stable/40244175>
- NUSSBAUM, Martha C. (2012). *Not for Profit: Why Democracy Needs the Humanities*. New Jersey: Princeton University Press.
- PESSOA, Fernando (2015). *Poemas de Ricardo Reis*. Edição de Luiz Fagundes Duarte. Lisboa: Imprensa Nacional-Casa da Moeda. Based on te critical edition (1994). <https://impresanacional.pt/wp-content/uploads/2022/03/PoemasRicardoReis.pdf>
- _____ (2001). *The Selected Prose*. Edited and translated by Richard Zenith. New York: Grove Press.
- _____ (1995) [1942]. *Poesias*. Nota Explicativa de João Gaspar Simões e Luiz de Montalvor. Lisboa: Ática.
- _____ (1993). *Álvaro de Campos: Livro de Versos*. Edição crítica. Introdução, transcrição, organização e notas de Teresa Rita Lopes. Lisboa: Estampa.
- _____ (1993a) [1946]. *Poemas de Alberto Caeiro*. Nota explicativa e notas de João Gaspar Simões e Luiz de Montalvor. Lisboa: Ática.
- _____ (1993b) [1973]. *Novas Poesias Inéditas*. Direção, recolha e notas de Maria do Rosário Marques Sabino e Adelaide Maria Monteiro Sereno. Lisboa: Ática.
- _____ (1993c) [1968]. *Textos Filosóficos*. Estabelecidos e prefaciados por António de Pina Coelho. Lisboa: Ática. 2 vols.
- _____ (1990) [1956]. *Poesias Inéditas (1919-1930)*. Nota prévia de Vitorino Nemésio e notas de Jorge Nemésio. Lisboa: Ática.
- _____ (1986). *Escritos Íntimos, Cartas e Páginas Autobiográficas*. Introdução, organização e notas de António Quadros. Lisboa: Publicações Europa-América.
- _____ (1984). "Além-Deus". *Orpheu* 3. Preparação do texto, introdução e cronologia de Arnaldo Saraiva. Lisboa: Ática.
- _____ (1982). *Livro do Desassossego*. Recolha e transcrição dos textos de Maria Aliete Galhoz e Teresa Sobral Cunha. Prefácio e organização de Jacinto do Prado Coelho. Lisboa: Ática. 2 vols.
- _____ (1966). *Páginas Íntimas e de Auto-Interpretação*. Textos estabelecidos e prefaciados por Georg Rudolf Lind e Jacinto do Prado Coelho. Lisboa: Ática.
- PIZARRO, Jerónimo (2021). *Fernando Pessoa: A Critical Introduction*. Brighton, Chicago, and Toronto: Sussex Academic Press.
- _____ (2018). *Ler Pessoa*. Lisboa: Tinta-da-china.
- _____ (2011). "Fernando Pessoa: Not one but multiple isms." *Portuguese Modernisms: Multiple Perspectives on Literature and the Visual Arts*. Edited by Steffen Dix and Jerónimo Pizarro. New York: Legenda, pp. 24-41. Cf. DOI: 10.4324/9781315089607.
- RAMALHO, Maria Irene (2021). *Fernando Pessoa e Outros Fingidores*. Lisboa: Tinta-da-china.
- RODOWICK, David N. (2021). *An Education in Judgment: Hannah Arendt and the Humanities*. Chicago: The University of Chicago Press.
- ROOSE, Kevin (2021). *Futureproof: 9 Rules for Humans in the Age of Automation*. New York: Random House.
- SENA, Jorge de (2000) [1982]. *Fernando Pessoa & C.ª Heterónima*. Lisboa: Edições 70.
- SILVA, Agostinho da (1959). *Um Fernando Pessoa*. Porto Alegre: Instituto Estadual do Livro.
- SVENDSEN, Lars (2017). *A Philosophy of Loneliness*. Translated by Kerri Pierce. Chicago: The University of Chicago Press.

- TALEGRE, Mar [Armando Martins Janeira] (1947). *Três Poetas Europeus: Camões, Bocage, Pessoa*. Lisboa: Livraria Sá da Costa.
- THURSTON, Luke (2008). "Splinters of Being: Fernando Pessoa as Multiple Singularity." *Qui parle*, vol. 17, n.º 1, Fall/Winter, pp. 175-192.
- VAN HULLE, Dirk (2019). "Artificial imagination, imagine: New developments in digital scholarly editing". *International Journal of Digital Humanities*, vol. 1, pp. 137-140. DOI: 10.1007/s42803-019-00020-w.
- VILA MAIOR, Dionísio (1994). *Fernando Pessoa: Heteronímia e Dialogismo. O Contributo de Mikhaïl Bakhtine*. Coimbra: Livraria Almedina.
- WANG, Junlin; WANG, Jue; ATHIWARATKUN, Ben; ZHANG, Ce; ZOU, James (2024). "Mixture-of-Agents: A Modular Approach to Multi-LLM Collaboration." *arXiv preprint*. <https://arxiv.org/abs/2406.04692>
- WU, Qingyun; BANSAL, Gagan; ZHANG, Jieyu; WU, Yiran; LI, Beibin; ERKANG, Zhu; JIANG, Li; *et al.* (2023). "AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation. *arXiv preprint*. <https://arxiv.org/abs/2308.08155>
- YANG, Runzhe; NARASIMHAN, Karthik (2023). "The Socratic Method for Self-Discovery in Large Language Models". Princeton Group Blog. <https://princeton-nlp.github.io/SocraticAI/>
- ZHANG, Yusen; SUN, Ruoxi; CHENM, Yanfei; PFISTER, Tomas; ZHANG, Rui; ARIK, Sercan Ö; *et al.* (2024). "Chain-of-Agents: A Modular Multi-Agent Framework for Long-Context Reasoning". *arXiv preprint*. <https://arxiv.org/html/2406.02818v1>

ANA ILIEVSKA holds a PhD in Comparative Literature from the University of Chicago (2020) and has studied and held positions at Yale, Brown, Lisbon, Tübingen, Bonn, and the University of Catania. Before joining AUB-Mediterraneo, she was Senior Research Fellow on a joint project between the Universities of Bonn and Cambridge on “Desirable Digitalisation: Rethinking AI for Just and Sustainable Futures.” Between 2021-2023 she taught at Stanford University and was Andrew W. Mellon Postdoctoral Fellow at the Stanford Humanities Center. Specializing in late 19th- and early 20th-century Southern European/Mediterranean literature and thought, her work examines non-hegemonic humanistic perspectives on technology and AI and the role of the humanities in the digital age. A Fulbright alumna (2018-2019), she has published widely and has been invited speaker at academic and industry conferences at Cambridge, the Sorbonne, Amsterdam, Dubrovnik, Lisbon, Valencia, Catania, Turin, Aix-en-Provence, Hamburg, and Tokyo, on topics spanning literature, philosophy, AI, critical thinking, and translation.

ANA ILIEVSKA é doutora em Literatura Comparada pela Universidade de Chicago (2020) e estudou e ocupou cargos em Yale, Brown, Lisboa, Tübingen, Bonn e na Universidade de Catânia. Antes de ingressar na AUB-Mediterraneo, foi pesquisadora sênior em um projeto conjunto entre as Universidades de Bonn e Cambridge sobre “Digitalização Desejável: Repensando a IA para Futuros Justos e Sustentáveis”. Entre 2021 e 2023, lecionou na Universidade de Stanford e foi bolsista de pós-doutorado Andrew W. Mellon no Stanford Humanities Center. Especializada em literatura e pensamento do sul da Europa/Mediterrâneo do final do século XIX e início do século XX, seu trabalho examina perspectivas humanísticas não hegemônicas sobre tecnologia e IA e o papel das humanidades na era digital. Ex-aluna da Fulbright (2018-2019), ela publicou amplamente e foi palestrante convidada em conferências acadêmicas e industriais em Cambridge, Sorbonne, Amsterdã, Dubrovnik, Lisboa, Valência, Catânia, Turim, Aix-en-Provence, Hamburgo e Tóquio, tratando sobre tópicos que abrangem literatura, filosofia, IA, pensamento crítico e tradução.