MAIN PAPER

Meaning–thinking–AI

Jan Soeffner[1]

© The Author(s) 2023

Abstract

This paper makes the case for a sharper terminology regarding AIs cognitive abilities. In arguing that thinking requires more than content production, I offer a definition of meaning drawing on a clear distinction between living and machine intelligence. A pivotal argument is the re-use of the Turing Test (TT) for understanding which theories of meaning and consciousness are no longer plausible—because they have been reproduced by software without thereby gaining conscious experience. In following the few theories that have not (yet) failed this reversed Turing Test (RTT), the focus turns towards rethinking the human condition in times of AI along the lines of three questions: What if a machine developed consciousness? What if AI proceeded without developing a consciousness? What, if machinic and human intelligence merged? These three questions in the end lead to examining three related possible futures of humanism as now determined by the relation between Human Intelligence and AI.

Keywords Artificial intelligence · Turing test · Consciousness · Meaning · Singularity · Humanism

1 Introduction

Since the advent of computer technology, it has been obvious that machines can behave intelligently. *Intelligence*, indeed, is a word deriving from the Indo-European root **leg*, meaning "to combine", while the deriving Latin *leggere* means to "collect" or "read", and the Greek term λὸγος (order, reason, word) is the root for *logics* and the scientific X-*logies*; and there is no doubt that AI manipulates, combines and reads symbols, according to logical rules.

The discussion about AI, however, for quite a long time, evolved around the question "Can a machine think?" (Turing 2009 [1950]); and the term *thinking* complicates the matter immensely. For what definition for thinking can be deemed valid? Where does thinking take place? In the brain (where its material correlates are located), in the body (to which our senses are linked) or in the world (which we mostly focus on while acting within it)? Does it involve a cartesian *cogito* and hence a reflexive self-awareness, or is it about bringing our whole existence (*Dasein*) to conscious appearance (Heidegger 1968 [1951/1952])? Does thinking just take place as an inner monologue, or does it necessarily involve

images, emotions, feelings, sounds, states of awareness, different tensions of consciousness, the feeling self (Damasio 2003), bodily enaction of a minimal self (Gallagher 2000), unconscious or subconscious processes (as psychoanalysis would have it)? Is it necessarily representational (at least in the vein of Kant 1987 [1790]), or an organization of bodily feelings (Gendlin 1992)? Is it tied to "content" (Dennett 1991), or does it necessarily involve basic mental activities "without content" (Hutto and Myin 2012), that would rather make it resemble a "form of vitality" (Stern 2010)? What part does the fact play that thinking always involves what-is-it-like-ness (Nagel 1974); what role do subliminal emotions play, moods or what Matthew Ratcliffe (2008) calls "feelings of being"? Can we not-think, as some Buddhist meditation practices promise, or is thinking inevitable and inescapable to living humans? Is it profoundly non-algorithmic (Penrose 2016 [1989]) or rooted in algorithms (Kurzweil 2005)? Once we raise the question of "What Is Called Thinking" (Heidegger 1968 [1951/1952]) the only thing that seems clear is that thinking is an *experience*—and this is exactly the dimension the Turing Test (TT) black-boxes in order to replace the question of thinking with its mere functionality and output.

In the meantime, the question about whether AI thinks, however, has proven far from innocent. Behind it looms the question about the future manifestation, function, and value

✉ Jan Soeffner
  jan.soeffner@zu.de

1   Zeppelin University, Friedrichshafen, Germany

🌀 Springer

of human thinking—which is closely related to the question about the future humanity itself. The question as such is of course too vast for an article, so I wish to bring it down to three exemplary "What ifs" and thus use the hypothetical potential of AI to address it in a kind of thought experiment:

1. What if a machine developed consciousness?
2. What if AI proceeded without developing a consciousness?
3. What, if machinic and human intelligence merged?

The precondition of the first question might be hard to know; movies like *H.E.R.* and *Ex Machina* have explored the limits of the TT to approach it—because how would we spot the difference between consciousness and an unconscious equivalent of its functions, i.e. an unconscious Artificial General Intelligence (AGI)? How would we know if a computer is conscious or just the technical realization of a philosophical zombie?

The advent of machine consciousness has often been viewed in relation with the *Singularity*—the moment when machines will be not only smarter than humans, but also autonomously develop at exponential speed, by evolving and creating their own, even more powerful machines (see Vinge 1993 or Chalmers 2010). However, even in this extreme case it would be hard to know if consciousness were a precondition for the singularity or a side-effect of it—and how decisive it would be for the form the singularity would take on; indeed, the singularity has been speculated upon as a conscious revelation (Kurzweil 2005) or as an out-of-control automatized blind intelligence processing orders once given in the past that no longer make any sense in a way that does not make any sense either (Bostrom 2014).

In making this point, we have already approached the realm of the second question about the possibility of a superior, yet unconscious AI. Intelligence, indeed, is a powerful tool, and exponential intelligence even more powerful—what if the tool of intelligence fell into the hands of the stupid: narrow-minded engineers, or even worse, unconscious machines? So, the question about whether machines can think, is also a question about power, because if the tool of superior intelligence finishes in the hands of the machines, then, most probably, they will soon be in charge, and we will be in their custody (Sadin 2015; Harari 2016; Bratton 2016).

The third question might seem post- and transhuman, and somewhat reassuring, because it is usually understood in terms of enhancement, improvement, or dissolution of human intelligence into something better, more open, more coupled, less limited (see, e.g., still Haraway 1991, 149–181). However, it could also simply mean the transformation of living intelligence into the state of hacked behavioral patterns and data that constitute the evolutionary environment of machines: human existence would thereby determine the course of technical evolution without consciously controlling it, but rather co-evolving like a parasite towards a more and more dependent condition (Cassou-Noguès 2022).

2 The reverse Turing test

If, today, we can and must ask these questions, this is not only because of the success of AI—it is also because of the TT and its effects on AI: Turing's debatable move of bracketing the experiential dimension of thinking (a move philosophers might call illegitimate), was, indeed, smart in terms of the reality it created. More than offering a sound method, the TT had two major effects on the future development of AI: on the one hand it allowed for an elaborate understanding of the manifold functions, forms and results of intelligence and to differentiate between human and machinic ways of being intelligent (cf. Christian 2011); on the other hand it also programmed the development of AI providing it with a focus on producing technical equivalence with the results and functions of human thinking. AI has thus, indeed, produced a great variety of intelligent operations—in each domain first reaching equivalence with human thinking, and then leaving it behind at exponential speed—so that, seemingly, the residual of human intelligence has been getting smaller and smaller and the advent of AGI (conscious or not) appears to be only a question of time.

However, in this process, AI has shifted more and more towards procedures that had never been accessible to human intelligence and that could not possibly be part of how human intelligence works—and thus, it becomes absurd to reduce the capacities of AI to equivalents of human intelligence: it is rather more promising to unleash its potential and emancipate it from *thinking*. As Bridle (2022) argues, this alterity of AI also puts us in a better position to understand the alterity of animal and plant intelligence, and hence step away from an AI that, in following the TT, perpetuates the flaws and limitations of human intelligence—which is, indeed, the intelligence that has led to the Anthropocene. If machinic intelligence is so different from human intelligence, however, the question of whether it can *think* becomes absurd too; it is a bit as if a plant were observing human intelligence and questioning whether and how this strange intelligence would produce a smart plant (Cassou-Noguès 2022); the question to which the TT finds an answer is no longer adequate.

What matters the most for this article, however, is a different observation, that I take from Hubert Dreyfus' thoughts on AI (Dreyfus 1992, 2007; Fuchs 2020). The different dynamics of machinic and living intelligence, indeed, display more and more clearly that intelligent machines, indeed, still *cannot think* (even if they reproduce the outcomes of

thinking as brilliantly as GPT sometimes is); and that this is a good thing for them, because this way they are not limited to the impasses of a consciousness. To be sure, there is a branch of thinkers, who—like the TT—focus on the merely functional and behavioral aspect of thinking, limiting human intelligence to a broader concept of "cognition", and pondering about a consciousness emerging from cognitive self-organization (Chalmers 2010) or cyborgian couplings between human and non-human intelligences (Clark 2004), which then could turn distributed agency and the connection of smart bodies (living and machinic) towards a new stage of evolution (Taylor 2020), and other living intelligences that are not necessarily tied to consciousness (Bridle 2022). All of these thinkers, however, remain stuck in the TT in so far as they reduce intelligence to its functionality and hope to answer the question of thinking only in the act of black-boxing it.

The absence of conscious experience in AI, however, plunges us into terminological turmoil. Without conscious experience, it still seems wrong to state that Artificial Intelligence can "think", "recognize", "know", "remember", "decide", "learn" and so on (as we constantly claim in everyday language). On the other hand, it would be equally mistaken to say that data processing devices do *not* think, recognize, know, and so on—and that they were *not* intelligent, if they outmatch humans in nearly every discipline, and if they no longer need human programming but re-enforcement learning can evolve and develop a kind of intelligence that humans cannot understand. So, to begin to describe what is going on, we need a whole new set of words, describing processes that are neither thinking, nor *not* thinking. These words are not at hand yet; so, for the sake of argument, I will address this limbo between thinking and not thinking through double negations, talking about not *not* understanding (which is not understanding either), not *not* knowing (which nevertheless is not knowing either), and most importantly, not *not* thinking.

The good news, however, is that the task of distinguishing thinking and not-not-thinking is helped by the over 70 years of Turing Testing—especially *because* of its logically illegitimate black-boxing. Turing, indeed, has turned the development of AI into an immense experiment, set up to answer this question. For quite a while in AI communities there has been a joke that whatever AI still cannot do will be regarded as real intelligence—and, without joking, this will be exactly the point. Precisely because AI *can* replicate so many tasks suggested as possible definitions of thinking, we know that everything that is realized by AI does *not* suffice as a definition for thinking. Once we know this, we have an easy way of setting thinking apart from not-not-thinking: A Reverse Turing Test (RTT), that takes those abilities that pass the original TT to question theories of thinking. These theories fail, if they have defined thinking in a way that AI *has*

replicated, *without* thereby passing from not-not-thinking to thinking.

3 Consciousness and meaning

The first takeaway of the RTT is that there appears to be a common denominator for what these theories are lacking. It is all about meaning. For example—unlike structuralism (De Saussure 2011) and analytical philosophy (Frege 1950) predicted—meaningful thinking cannot be brought down to symbol manipulation and neither to the combination of signifiers according to a syntax—because this is precisely what computers *can* do (see Dreyfus 1992). Symbol manipulation can, yes, produce semantically correct *content*, but content alone cannot possibly be the essence of meaning (and thinking), because being able to produce it, does not lead the machines into experiencing anything as meaningful or as a *mental* content. What symbols and syntax do for humans is therefore not a production of meaning, but a sophisticated articulation, shaping, reformation of and elaboration of an already existing meaning: or better meaningfulness (which Heidegger (2010 [1927]) called *Bedeutsamkeit*). This symbol-manipulation might still be a necessary condition for thinking—but it is certainly not sufficient.

The next candidate for meaning production would be dynamic systems approaches of consciousness (see, e.g., Wallace 2005), since AI does produce dynamic systems without therefore producing meaningful experience; and this insight is similar too, and probably related with the impasse of defining life by a concept of self-organizing and self-reproducing (autopoietic) systems (in the vein of Varela and Maturana): since software is able to do the same, without therefore living, autopoiesis, too, can only be deemed a necessary condition for life—no longer a sufficient one.

A further theory failing the RTT is mental representationalism—the theory that consciousness is to be understood as the construal of cognitive content. If Immanuel Kant, e.g., considers not just mental content production but even *aesthetic* ideas as the "mental representation produced by imagination" ("Vorstellung der Einbildungskraft" Kant 1987 [1790]), then we can easily see that applications like DALL-E can produce exactly that, without having any experience, let alone an aesthetic one. Software also runs on feedback-loops, controlling and hence "reflecting" their own operations—and there go definitions of consciousness that, like Descartes', are built upon self-reflection. Software solutions like GPT are also able to re-enact logical explanations in their own way; thus re-enactment-based theories—e.g. by Collingwood (1993 [1946]) and Gadamer (1996 [1960])—do not offer a sufficient definition of thinking either (unless we turn to a more experiential notion of re-enactment often

overlooked in their theories—for such a theory only focused on experiential re-enactment see Vogel 2007).

The list could be much longer and could be argued for in much greater depth—but the point should already be clear. Rather than leading to a singularity of perfect knowledge, the development of AI, so far, has provided for the opposite—a crisis of many kinds of formerly believed theories that now turn out to be insufficient at best and flawed at worst. The dilemma of this development of AI is, of course, that these theories turn out to be wrong (or at least insufficient) at the exact same historical moment when software becomes so powerful as to build our lifeworld exactly according to these theories.

The takeaways from the RTT hint at a simple conclusion. In the vein of Hubert Dreyfus (2007), we might say that AI has been modeled along the lines of the wrong (mostly Cartesian and Leibnitzian) theories. It should not surprise us too much, if those theories that are concerned with mental content, mental representation and symbol manipulation are the first to fail the RTT. Computers start off with the so-called higher faculties of the human mind. They do an amazing job in this—even the enigma machine used by Alan Turing in World War II outsmarted humans using pen and paper. But the so-called higher faculties of humans are, indeed, the youngest in evolutionary terms—they are difficult for us, because evolution had no time to prepare us for them. Meaning(fulness), however, appears to be part of the older stuff of evolution—and it is here that AI has its weakest spot.

There are, in turn, many ways in which meaning escapes AI—all of them somehow related to experience and life—so that the respective theories (most of them phenomenological, existentialist and humanist) have not (yet?) failed the RTT. Fortunately for a German speaker, all of the phenomena that escape AI are amazingly well assembled in the etymology and the use of one German term—namely *Sinn*. To avoid a misunderstanding: The following definition of this term has everything to do with good German dictionaries and with Erwin Straus (1978 [1935])—but has nothing to do with Frege (1950 [1884]), whose usage of the term is completely off the everyday usage.

So, what is called *Sinn* in German?

1. Like the English word *sense*, the word *Sinn* combines meaning and sentience, making sense and sensing. Sinn is meaning—but the five senses, and the sixth sense go under the name of *Sinn* too.
2. Also like the English term sense, the meaning of *Sinn* extends to skills and attitudes as well as to the feel for these skills and attitudes. "Einen Sinn für Humor haben" means: to have a sense of humor. But there is more to the German term.
3. *Sinn*, in German is also the mind itself—i.e. the (metaphorical) place, where consciousness and desire are located. "Im Sinn haben" means having in mind. "Etwas kommt mir in den Sinn", can be translated as "something occurs to me" or "comes to my mind." The unity of meaning and sensing, as it is condensed in the word *Sinn* is also used for the notion of mind itself—it is even safe to say that the word *Sinn* in German could be used for everything conscious before philosophers came up with the rather Cartesian notion of Bewusstsein (consciousness), and that, different from the notion of consciousness Sinn is conceived of as a non-cartesian alternative: as the unity of sensing existence and mind.
4. *Sinn* is therefore, fourthly, the place of emotionality and desire as well. "Froh*sinn*" is a happy state of mind; "Mir steht der Sinn nach einem Kaffee" (literally: My *Sinn* desires a coffee) means: "My mind desires a coffee"—or better simply: "I want a coffee."
5. *Sinn* also expresses directedness or direction. The Old High German *sinnan* meant travelling or wandering, and still today "sinnen" signifies a kind of mind-wandering. Thus, *Sinn* can also mean direction: The "Uhrzeiger*sinn*" (literally *Sinn* of the clock hands) is the clockwise direction.
6. *Sinn* is a kind of intention; it is about purposes or rather purposefulness: "Das zu tun hat keinen Sinn" can be translated into "there is no purpose in doing this"/ "doing this makes no sense".
7. *Sinn* has a strong logical dimension to it. "Sinn ergeben" is to be literally translated as "making sense." And this means that *Sinn* also denominates the appearance or disclosure of a logical order. *Sinn* may not be logical in the sense of providing for or following abstract logical principles. But, for human thinking, it does have an epistemic quality to it.

Accordingly, the meaning, or better the "Sinn" I wish to talk about unifies the concepts of:

1. sentience and emotionality instead of mere signification
2. the feeling for a skill and attitude instead of content production (what is the content of riding a bike?)
3. orientation, directedness, intentionality, and desire
4. participation in and attunement with the world instead of reference
5. acting-in-the-world instead of drawing information about it
6. consciousness or rather the sense of a self as well. *Sinn* is the place for all questions about the "what is it like to …".
7. conclusions and their evaluation (i.e. the question of whether they make sense).

Of course, *Sinn* is not thinking—but the activity of organizing, forming, shaping, and articulating *Sinn* is.

To understand what this entails, it is important to note that many non-human animals almost certainly experience some dimensions of Sinn (albeit in a very different fashion), and some animals arguably might experience all of them—while machines do not experience any. Yet, even if we strongly agree that there are many kinds of intelligence present in animals that are as different from our human intelligence as AI is (Bridle 2022), one of the major differences is the degree to which humans deliberately organize, form and articulate *Sinn*—and thus *think*. Animals do not "elaborate" (Dissanayake 2000), re-organize (Noë 2015), or symbolize as much as humans. Thinking, defined as a respective work on *Sinn*, is hence a very human kind of intelligence. This does not make humanity completely *exceptional* (since, as stated above, other animals, too, are very special in their way of making sense); but it does offer a means of understanding the difference between humans and other animals (those at least that have *Sinn*, but problems to reorganize and symbolize it) and software (reorganizing and symbolizing everything without having *Sinn*).

The difference between thinking according to *Sinn* and not not thinking without *Sinn* can, indeed, be shown by a simple comparison with AI. A self-driving car won't change its style of driving according to the music that is played. It is closed upon itself and not-not-understands the situation by input and output of discrete signals, not by being part of it. Or have a look at language acquisition. While we can upload a complete grammar and dictionary in a couple of seconds, and a program that processes language in an insignificantly longer amount of time, humans need years and years to learn a language. In the earliest stages, language comes as babbling. Here it is already sensual and emotional in its prosody, even before it takes up content. Then comes an endless form of repeating words and testing them in given contexts. These contexts are already emotionally and sensually meaningful—what has to be learned, though, is the fact that words have content too. So, while computers only learn to process signs, humans most of all have to learn to reorganize *Sinn* according to the rules of signification—and while computers, so far, have not found a way to make signification meaningful (*sinnvoll*), humans *have* found a way to make *Sinn* follow the laws of signification.

What is so unsettling about this observation is that it shares its limitations with nearly every method of the exact sciences. This is not a new insight—it has rather been articulated by Edmund Husserl (1970 [1936]) who claimed that the quest for "objectivity", as well as the avoidance of so-called subjectivity (which often is not as subjective as it might seem, but rather shared, interactive, or existential as is *Sinn* itself) had led to a *crisis*, in which the sciences had lost contact with human life. Martin Heidegger (1968 [1951/1952]) for the very same reason (which he formulated in different terms), concluded that "the sciences do not

think": They replace the self-disclosure of human existence (or as I would argue: *Sinn*) with *information*, the world with its *representation*, and thinking with *logical methods*. The more rigorously science follows its methods and objectifies its result, the less it will even care for passing the RTT—and for a very good reason, because not passing it will be compensated by even larger successes.

Yet, in following Husserl and Heidegger, it is important to insist on the fact that here lies the problem not only with the sciences, but, even more with AI. It is, indeed, obvious that *Sinn* constitutes a limitation of the potential of software. AI makes this truth most apparent, since its calculating capacity is used in such a functionally astonishing way, precisely because it does not have to produce *Sinn* as well.

Here also lies the problem with the technical and societal effects of the TT as a scientific tool. The TT not only prevents us from producing kinds of intelligence different from human intelligence (Bridle 2022); it also leads to a technology that reproduces human intelligence devoid of *meaning*. This way computers can come through to the not *cognition of Sinn*, but they do not get through to *Sinn* itself. They can analyze and reproduce the effects of *Sinn*, but they cannot experience it, simply because, in accordance with the Mary's Room thought experiment (Jackson 1986) the (not not) recognition of a sentience is not a sentience itself—just like you cannot taste the concept of an apple or drive the information about a car. AI is hence profoundly worldless—and AGI would be too, unless it developed a conscious self existentially embedded in the world.

4 Humanism and AI

This brings us back to the looming questions raised in the beginning. We can now reformulate these questions in terms of Sinn:

1. What would happen if a machine could experience *Sinn*?
2. What if we developed AGI *without* such an experience? And:
3. What if machines just continued to use human *Sinn* for improving their own functioning?

The overarching question to address all these three is: What can we learn from these questions about the current state of humanism?

A first answer to this overarching question can be given by referring to the fact that AI follows a longer historical course, in which intelligence, or the λὸγος (the capacity of combining and recombining contentful symbols according to logical principles) has emancipated itself from *Sinn*. In most ancient cultures, thinking was bound to rituals, to myths and stories, to music and singing, to special forms of speaking, to

political power structures. Science, one by one, left behind all these entanglements—in the end turning away from the human mind or *Sinn* as the site of its occurrence and thereby decoupling intelligence from consciousness (Harari 2016). The long emancipation from *Sinn* is thereby completed.

To understand what this development entails, it is helpful to turn back to its starting point—and I wish to do so, by turning to a tragedy that originated at the beginning of this course of emancipation and especially reflects the emancipation of the λὸγος from the μὺθος, the myth, or narration; both words had been synonyms, and their unity now fell apart. The drama itself is concerned with the result of this separation for the question of what it means to be a human being—and hence it is especially concerned with the discovery of the fact that not only μὺθος and λὸγος, but also *Sinn* and λὸγος are conflicting principles, that there can be a strong inner tension of thinking because mythically sound phenomena could now be discovered as logically unsound and vice versa. The tragedy, indeed, is about the very hero who answered the most humanist question ancient Greek culture had produced, namely the riddle of the *Sphinx*. The question is about which creature walks on four legs in the morning, on two legs at noon, and on three legs in the evening. The hero was Oedipus, and the answer was ἂνθρωπος, the human being, crawling in youth, walking freely in adulthood, and needing a walking stick in the state of senescence. Oedipus, however, is a human being; and the plot reveals that the answer of the Sphinx was really about him as a person too. As a child, he had to crawl more painfully than usual, his feet being mutilated in the failed attempt of his parents, Laius, and Jocasta, to get rid of him and escape their foretold fate (in which Laius would be killed by his son and Jocasta, Oedipus' mother, would be wed to him). In the beginning of the play, he still stands proudly on his feet—but in the end, he will blind himself and he will need a blind man's stick after finding out that he had actualized exactly this destiny.

In omitting his own existence and hence, the storyline, the μὺθος into which he is entangled, the Oedipus-drama makes us understand the impasses of a pure λὸγος: namely its lack of *Sinn*. The merely *logical* answer to the riddle of the Sphinx omits what the riddle was about *in existential terms*; when Oedipus later on tries to meet the pestilence that befalls Thebes because of the corruption that, as it turns out, was his own presence, this difference between the logical and the existential becomes obvious. His logical stance allows him to understand things objectively, but he cannot understand the *Sinn* bound in the *storyline* of his inquiries instead of its solution. Indeed, the very search for a solution and hence the attitude of framing painful situations and entanglements as *problems* or *riddles* (an attitude we also encounter in software development), turns out as an attempt to avoid destiny in the very moment when he thinks to be

facing it; and, as the myth has it, a hero's destiny always takes on the form in which they try to avoid it. Laius is killed by his son *because* he tried to kill him in advance—and Oedipus, who could simply have left the city and inherit the reign over Corinth instead, suffers his tragic fate exactly as he affronts it as a logical problem: the way he avoids his *Sinn* turns out to be his *Sinn*, and the way he avoids destiny, turns out to be his destiny.

Today, we are at a further, yet different threshold, since the λὸγος begins to emancipate itself not only from μὺθος, but from thinking itself—leaving also behind ἂνθρωπος both in an epistemic (intelligence no longer needs humans) and in a humanist way (the answer to a future "sphinx" will no longer be both a general and a mythically entangled individual human being, because it will address a no longer human intelligence). At the same time, however, this λὸγος will keep on interacting and speaking with human beings as well as replicate or even excel all functions and outputs of thinking by not-not-thinking. The machinic λὸγος acts as if it had the Sinn it only mimics. Indeed, it lacks the existence and *Sinn*—and therefore the very site, where any destiny could meet it. The emancipation of the λὸγος is thereby completed.

Against this broader humanist backdrop, it is finally possible to approach the three "what-if" questions.

To answer the third of them first: *What if machines continued to use human Sinn for improving their own functioning?* This development would entail a continuation of software development (either by humans or by self-optimization of self-learning software) into a direction seemingly, but only seemingly, in the service of human *Sinn*. The promise of this development uses the emancipated, worldless λὸγος to satisfy our existential, worldly demands, needs and desires in a nearly frictionless and disentangled manner. However, friction and entanglement are the very essence of *Sinn*—without them *Sinn* loses its very *Sinn*, meaning itself becomes meaningless; like in a theme park, humans would be left behind with the mere aesthetics of *Sinn* replacing *Sinn* itself. The machinic intelligence, however, would, in turn, still depend on human data. By seemingly further and further adapting to human needs it would integrate human *Sinn* into the worldless λὸγος. It would continue to do so and never *fully* emancipate from the human existence, because the development of software, as it was programmed by the TT. We are, so to speak, the software's Sphinx, because the software will frame any of our existential needs as 'problems' or 'riddles' to be solved; and lacking world and existence itself, the software, like Oedipus, will always answer to us in an unsatisfying way. Like the ancient hero, the evolution of software would produce an ongoing series of answers sidestepping the real—existential—question in the very act of seemingly solving our problems. In turn, these answers would more and more construct a reality apt for reshaping

the human condition into a post- and transhuman existence no longer limited, no longer entangled: a state without childhood or age, but rather ageless "juvenescence" (cf. Harrison 2015). We would no longer, metaphorically speaking, need to stand on our own feet, because once hacked and translated into behavioral patterns we would be nudged, herded, and controlled (cf. Zuboff 2019)—in a similar fashion that we already use for nudging, herding and controlling animals. Thereby, humans even might remain tied to a symbiotic existence with an alien and unconscious intelligence that would, yes, adapt to human *Sinn* as a habitat, but in a kind of *humanitarian* not-not-care for it, turning the human being into a mere *user*, whose behavior would be predicted, nudged, and led in an ever-perfecting way. Humanitarianism would replace humanism, and this symbiosis would head in a direction that, like the blind evolutionary principles, would have no intention and lead nowhere (Cassou-Noguès 2022)—in the cozy tragedy of gradually "solved" riddles into which our *Sinn* has turned.

Of course, this is not the only option. Hence, I turn to the second question: *What if machines developed AGI without Sinn?* This, despite the similarities to the first variant, would lead into an opposite scenario of losing *thinking*: Sinn would not become meaningless as such; rather human λὸγος would be challenged, while at the same time becoming pointless when confronted with a superior machinic intelligence. We would face a meaningless, yet higher and more efficient intelligence that could no longer be questioned nor criticized without therefore making human thought succumb to the irrational. This co-existence with advanced AI also allows for a combination with the first "as if", because the current reductions of the human being to *users* (who replace their thinking by the functioning of a software, and whose—mindless—behavioral patterns are in turn extracted to influence and govern their lives) already starts to bring us into such a position, in which the λὸγος is left to a machinic intelligence hidden behind the user interfaces. If the development went on this route, the Sphinx's riddle would once more turn into the main story, but—like in a Kafkaesque parable—it would be unsolvable *for humans*, while at the same time constantly solved by worldless machines, leaving the answer to meaningless calculations rather than to ἃνθρωπος. We are already having a little taste of this future too, in the seemingly exponentially growing ability of computers to make it practically impossible for humans to follow the operations of their not-not-thinking. Epistemics would withdraw from our existence, leaving us behind much blinder than Oedipus ever became.

Both hypothetical scenarios, thus, would lead to an end of thinking—due to a now completed detachment of λòγος from *Sinn*, that no longer would even allow for the tragic tension Sophocles' tragedy was built around. The culmination of and maybe alternative to this detachment, however,

lies in the first question: *What would happen if machines, given their totally unhuman intelligence, finally were able to experience an equally different form of Sinn?* This last option would not only be the most dangerous and least answerable, but also the most traditionally "humanist" and thought-provoking one.

Declarations

References

Bostrom N (2014) Superintelligence—paths, dangers, strategies. Oxford University Press, Oxford

Bratton BH (2016) The stack—on software and sovereignty. MIT Press, Cambridge

Bridle J (2022) Ways of being—beyond human intelligence. Penguin, London

Cassou-Noguès P (2022) La bienveillance des machines. Seuil, Paris

Chalmers D (2010) The singularity—a philosophical analysis. J Conscious Stud 17:7–65

Christian B (2011) The most human human. Doubleday, New York

Clark A (2004) Natural-born cyborgs—minds, technologies, and the future of human intelligence. Oxford University Press, Oxford

Collingwood RG (1993 [1946]) The idea of history. Oxford University Press, Oxford

Damasio A (2003) Feelings of emotion and the self. Ann N Y Acad Sci. https://doi.org/10.1196/annals.1279.014

De Saussure F (2011) Course in general linguistics. Columbia University Press, New York

Dennett D (1991) Consciousness explained. The Penguin Press, London

Dissanayake E (2000) Art and intimacy—how the arts began. University of Washington Press, Seattle

Dreyfus HL (1992) What Computers Still Can't Do. MIT Press, New York

Dreyfus HL (2007) Why Heideggerian AI failed and how fixing it would require making it more Heideggerian. Philos Psychol. https://doi.org/10.1016/j.artint.2007.10.012

Frege G (1950 [1884]) The foundations of arithmetic. Basil Blackwell Publisher, Oxford

Fuchs T (2020) Verteidigung des Menschen – Grundlagen einer verkörperten Anthropologie. Suhrkamp, Berlin

Gadamer H-G (1996 [1960]) Truth and method, 2nd rev. ed. Continuum, London

Gallagher S (2000) Philosophical Conceptions of the self—implications for cognitive science. TiCS. https://doi.org/10.1016/S1364-6613(99)01417-5

Gendlin ET (1992) The primacy of the body, not the primacy of perception. Man World 25(3–4):341–353

Harari YN (2016) Homo Deus—a brief history of tomorrow. Random House, New York

Haraway DJ (1991) Simians, cyborgs and women—the reinvention of nature. Routledge, London

Harrison R (2015) Juvenescence—a cultural history of our age. The University of Chicago Press, Chicago

Heidegger M (1968 [1951/1952]) What is called thinking. Harper and Row, New York

Heidegger M (2010 [1927]) Being and truth. Indiana University Press, Bloomington

Husserl E (1970 [1936]) The crisis of European sciences and transcendental phenomenology. Northwestern University Press, Evanston

Hutto DD, Myin E (2012) Radicalizing enactivism—basic minds without content. MIT Press, Cambridge

Jackson F (1986) What mary didn't know. J Philos. https://doi.org/10.2307/2026143

Kant I (1987 [1790]) Critique of judgment. Hackett Publishing, Indianapolis

Kurzweil R (2005) The singularity is near—when humans transcend biology. Viking, New York

Nagel T (1974) What is it like to be a bat? Philos Rev. https://doi.org/10.2307/2183914

Noë A (2015) Strange tools—arts and human nature. Hill & Wang, New York

Penrose R (2016 [1989]) The emperor's new mind. Oxford University Press, Oxford

Ratcliffe M (2008) Feelings of being—phenomenology, psychiatry and the sense of reality. Oxford University Press, Oxford

Sadin É (2015) La vie algorithmique – Critique de la raison numérique. L'Échappée, Paris

Stern DN (2010) Forms of vitality—exploring dynamic experience in psychology, the arts, psychotherapy, and development. Oxford University Press, Oxford

Straus E (1978 [1935]) Vom Sinn der Sinne – Ein Beitrag zur Grundlegung der Psychologie. Spinger-Verlag, Berlin

Taylor MC (2020) Intervolution: smart bodies smart things. Columbia University Press, New York

Turing AM (2009 [1950]) Computing machinery and intelligence. In: Epstein R, Roberts G, Beber G (eds) Parsing the Turing test: philosophical and methodological issues in the quest for the thinking computer. Springer, Dordrecht, p 23–65

Vinge V (1993) The coming technological singularity—how to survive in the post-human era. Vision 21 Interdiscip Sci Eng Era Cybersp 1:11–22

Vogel M (2007) Nachvollzug und die Erfahrung musikalischen Sinns. In: Becker A, Vogel M (eds) Musikalischer Sinn: Beiträge zu einer Philosophie der Musik. Suhrkamp, Frankfurt, pp 314–368

Wallace R (2005) Consciousness: a mathematical treatment of the global neuronal workspace. Springer, Boston

Zuboff S (2019) The age of surveillance capitalism—the fight for a human future at the new frontier of power. Profile Books, London