Humanities & Social Sciences Communications

ARTICLE

https://doi.org/10.1057/s41599-025-05868-8 OPEN

There is no such thing as conscious artificial intelligence

Andrzej Porębski[1]✉ & Jakub Figura[1]

The claim that so-called artificial intelligence (AI) can gain consciousness is on the verge of becoming mainstream. The thesis of this conceptual study is simple: There is no such thing as conscious AI. We argue that the association between consciousness and the computer algorithms used today (primarily large language models, LLMs), as well as those that would be invented in the foreseeable future, is deeply flawed. We believe that these flawed associations arise from a lack of technical knowledge and the way several new technologies (especially LLMs) work, which can create the illusion of consciousness. Moreover, we argue that the public discourse about AI is skewed by "sci-fitisation", which involves the unsubstantiated influence of fictional content on perceptions of this technology. To justify our claim, we reveal the incoherence in the argument that several computer algorithms are treated differently from other computer algorithms despite congruent modes of operation and a reliance on binary code and semiconductors. We believe that mathematical algorithms implemented on graphics cards cannot become conscious because they lack a complex biological substrate. We emphasise that the recognition of the consciousness of LLMs on the basis of their assertions is flawed because the language usage of LLMs is strictly probabilistic. Unfortunately, because the remarkable linguistic abilities of LLMs are increasingly capable of misleading people, people may attribute imaginary qualities to LLMs. Thus, a socially dangerous phenomenon referred to as "semantic pareidolia" is reinforcing.

[1] Jagiellonian University, Kraków, Poland. ✉email: and.porebski@uj.edu.pl

Question and rationale for the problem posed

We begin this paper with the following question: *Is there such a thing as conscious artificial intelligence*?

The potential capabilities and limitations of machines have been a contentious issue since the inception of computing technology. From Turing's famous question "Can machines think?" (Turing, 1950) to the issue of attributing mental states to machines (McCarthy, 1979), the classic Chinese Room argument (Searle, 1980), attempts to operationalise "machine consciousness" (Wasiewicz and Szuba, 1990), and theses on the limitations of computers (Dreyfus, 1992)—these are prominent examples of the academic debates about the qualities of technological entities. Although these debates were often considered extravagant, this situation has changed. The question of the potential abilities of so-called artificial intelligence (AI) is no longer just an academic niche but has emerged in the public discourse, with some people arguing that AI can indeed be conscious. Colombatto and Fleming (2024) reported that only one-third of respondents (adults in the US, $n = 300$, data collection in July 2023) firmly rule out any form of consciousness of large language models (LLMs; for a comprehensive overview, see Naveed et al., 2025; for an accessible introduction, see Cho et al., 2025, and the accompanying web application; see also Chollet, 2023), i.e., they indicate that ChatGPT is "clearly not an experiencer". The same study clearly reveals a linear relationship between the use of these technologies and estimated attributed consciousness: Those more likely to use LLMs attribute a higher consciousness to them.

More results indicate that the viewpoint that AI is conscious is not a mere statistical margin. In a large-scale 2023 survey, approximately 20% of respondents (adults in the US) declared that sentient AI systems currently exist (data collection from April to July 2023, $n = 2268$; Anthis et al., 2025). A 2024 survey revealed that among AI researchers and adults in the US, approximately 17% and 18%, respectively, believe that at least one AI system has subjective experience and approximately 8% and 10%, respectively, believe that at least one AI system has self-awareness (data collection in May 2024; $n_{AI} = 582$; $n_{adults} = 838$; Dreksler et al., 2025).[1] These findings paint a clear picture: Even if those who currently recognise the existence of conscious AI are in the minority, they are indeed not an anomaly.

These figures aside, let us provide specific examples. One of the best-known examples relevant to this paper is the activity of Google's AI engineer, Blake Lemoine, who publicly stated that he considered an AI model created by Google as sentient (Lemoine, 2022; Tiku, 2022). What was the rationale behind his statement? He referred to a conversation with the chatbot. Another example based on a conversation with a chatbot is provided in a paper by Hintze (2023), titled: "ChatGPT believes it is conscious". The author writes, "This study also raises intriguing questions about the nature of AI consciousness" (p. 1).

Ilya Sutskever, one of the cofounders and chief scientist of OpenAI (the company that created ChatGPT), tweeted, "It may be that today's large neural networks are slightly conscious" (Cuthbertson, 2022). Whatever "slightly consciousness" is supposed to be, Sutskever's claim carries more weight than its expression may suggest: Among many various possibilities ("may"-s), Sutskever chose to mention precisely this possibility, suggesting at least a significant probability of this form of consciousness.

Another example is an essay by Chalmers (2023), in which he asks, "Could a Large Language Model Be Conscious?", to which he answers, "Within the next decade, even if we don't have human-level artificial general intelligence, we may well have systems that are serious candidates for consciousness".

It is worth citing Bojić et al. (2024, p. 11): "The initial observation of language models' capabilities suggests that GPT-3 can simulate some degree of consciousness". Although these authors present their conclusions using cautious words ("can simulate", not "has simulated"), they suggest that the capabilities of these language models should be constantly monitored for "the potential emergence of awareness and self-awareness". The authors seem sensitised to the possibility of the birth of consciousness: "It is time to conduct focused empirical inquiries to uncover the emerging capabilities of AI models as they progress towards a consciousness-like state" (Bojić et al., 2024, p. 11). Butlin and Lappas (2025) express a similar viewpoint: They state that "[r]ecent research suggests that it may be possible to build conscious AI systems now or in the near future" (p. 1673), and they conclude that ethical principles need to be introduced to protect against problems associated with the creation of conscious entities "inadvertently". In their report (which describes research funded by Anthropic, an AI company), Long et al. (2024)—the team that includes people affiliated with AI companies—go even further by discussing so-called AI welfare. They consider "AI welfare" an important issue that should be acknowledged because of "the realistic, non-negligible chance that some near-future AI systems will be welfare subjects and moral patients" (Long et al., 2024, p. 29). Long et al. (2025) take this idea so far that they even observe a conflict between AI safety for humans and "AI welfare" (we pretend that we do not see how convenient this claim is for technology companies seeking an argument against AI regulation, which would cause data centres to fall into despair).

We have shown that AI consciousness—or more precisely, the consciousness of a specific AI technology, i.e., LLMs—is included in the public discourse about AI. (Interestingly, scarce in the public discourse is the topic of consciousness of, e.g., autonomous cars, which are, on further reflection, no less a spectacular result of AI research.) But why is addressing AI consciousness important?

AI technologies are becoming more and more present in our lives. Even careful observers, however, may overlook that AI-based solutions are successfully penetrating even those fields of human activity that were once associated only with dystopian science fiction, such as social, romantic, and even sexual relationships (Pentina et al., 2023; Alonso, 2023; Cole, 2023). Because people can place enough trust in technology to establish a social relationship with it, it should not surprise them that they may overly trust technology in other areas. For example, some lawyers confide in ChatGPT so much that they cite nonexistent cases hallucinated by ChatGPT (Weiser, 2023). Other people are disappointed that chatbots misinform about ticket discounts (Yao, 2024). At worst, the use of AI technology intended to help (advise or provide answers) has contributed to tragic consequences, such as suicide (van Es and Nguyen, 2025; Xiang, 2023).

The above examples represent human-machine interactions that have "gone wrong". The problem is that we still do not know what "proper human-AI interaction" is. One may speculate that it is an interaction that results mostly in good consequences. Nonetheless, this answer, while sensible, immediately prompts the following question: how can it be operationalised? How can it be operationalised under time pressure?—a problem that will loom larger as AI technologies enter various technologies that humans do not associate with untrustworthy AI, such as search engines (Volokh, 2023).

We believe that to create appropriate strategies for the education and regulation of human interaction with AI systems, one must first answer the following fundamental question and then use the answer as a premise: Does conscious AI exist? If conscious AI does not exist, this greatly simplifies things. For example, it justifies the use of regulatory strategies in which the technology

can be treated as a subject for pragmatic rather than ethical reasons. The nonexistence of conscious AI can and should be emphasised in technology education, which will facilitate the prevention of several undesirable attitudes towards the technology (e.g., an emotional addiction to chatbots, which can disappear in a single update). In other words, we believe that the issue of AI (un)consciousness must be determined "at the outset" of reflection on the interaction between AI and humans so that it does not repeatedly arise like a reproachful, quivering doubt: "and yet maybe I am conscious…?".

We have justified the main question of this paper. The issue is actually addressed in the public discourse and, moreover, important because of its entanglement with important social, psychological, and regulatory spheres. Now, we can revisit the following question: *Is there such a thing as conscious artificial intelligence*? Our answer to this question is simple.

The answer

There is no such thing as conscious artificial intelligence.

Conceptual remarks

Before we justify our answer to the abovementioned question, we need to clarify two concepts that are central to this paper: "artificial intelligence" and "consciousness". We accept this challenge with the awareness that it cannot fully be won because these concepts do not lend themselves to comprehensive and precise representation. Fortunately, high precision is not necessary for further consideration, because we will not be pulling grains of sand from a large heap (Łukowski, 2011, p. 131–133) and defining exactly what AI already is and what it is not yet, and which animal already is conscious, and which is not yet. To substantiate the answer, it suffices to define specific boundary conditions for an entity to be conscious and a general definition of AI that considers entities that are "certainly" AI and excludes those that are "certainly not" AI. We turn to conceptual clarifications.

Minimalist cognitive stance on the concept of consciousness. The most complicated concept in this paper is consciousness. Consciousness can be understood in very different ways (Kuhn, 2024; Kirkeby-Hinrup, 2024).

We, the authors of this paper, do not precisely know what consciousness is. We, humans, do not know it precisely. However, we know that some animals, at least humans, can be characterised as "conscious", and we know some of the conditions that can clearly distinguish nonconscious objects (e.g., a stone) from conscious objects. These conditions include, for example, the feeling of separation from space (which extends beyond the ability to avoid obstacles and walls), sensory perception, intentionality, thinking in mental images, self-referentiality, and having a self-concept (Farthing, 1992, pp. 24–44; Tyler, 2020).

Moreover, we do not precisely know which parts of the human body—and to what extent—make humans conscious. However, we do know something about it, at least that the nervous system, especially the brain, has an important role in making us conscious and that consciousness is grounded in complicated biochemical processes, including the release of various neurotransmitters and mediation by hormones, and the embodied nature of cognition (Young and Pigott, 1999). Using this minimalist cognitive stance with respect to consciousness, we do not claim to accurately define all of its characteristics and foundations, but we can formulate specific necessary conditions for the possibility of its recognition.

Problems with the concept of artificial intelligence. In this paper, we refer extensively to the concept of AI. Importantly, with respect to the question, we are not considering fantastic or futurological visions of AI but the actual AI, as it exists both in 2025 and in the foreseeable future. A challenge in the discussion of the characteristics of AI is what we refer to as the *sci-fitisation* of the discourse. Therefore, before we clarify the concept of AI itself, we examine the *sci-fitisation* (for other remarks on a similar issue, see, e.g., Hermann, 2023) and other problems with the concept of AI. The comments in this section do not constitute independent arguments for or against AI consciousness. However, these comments may elucidate the origin of several social trends related to AI and misunderstandings in discussions about AI.

Sci-fitisation. The eager engagement of popular culture with AI often makes linguistic intuitions about AI drift towards fictional creations, as convincingly depicted in popular culture, which rarely stops at the level of existing technology. Consider three "artificially intelligent" humanoids from popular culture: (1) androids, which are almost indistinguishable from a human, from "Blade Runner" (1982), inspired by Phillip K. Dick's novel "Do Androids Dream of Electric Sheep?" (1968); (2) the Terminator from the film "The Terminator" (1984); (3) superintelligent robots from the film "I, Robot" (2004), partly inspired by the works of Isaac Asimov, as compiled in "I, Robot" anthology (1950). These "artificially intelligent" entities had no real-world equivalent when they were described, nor is this the case today. Their creators may have been thinking about the potential targets (intelligent human-like robots) rather than the actual possibility of creating them technologically. Thus, we can consider these entities analogous to elves or dwarves (known from fantasy literature) rather than currently existing technology (or technology that can be created within a foreseeable time). Just as elves or dwarves are *fantastic* alter-humans, so the androids from "Blade Runner" or the robots from "I, Robot" are *technological* alter-humans. They tell us something about humans, but not about technology.

However, to know that these robo-figures, counter-intuitively, say more about humans and society than about technology, one must first know a lot about technology. When one does not know (which is the typical case), the perusal of the aforementioned oeuvre may result in a "conceptual drift": Readers and viewers will have the impression that the term "AI" refers to what they know from culture ("artificially intelligent" humanoids) rather than to what really exists in the world (smart hoovers or fallible chatbots). Features from fictional representations of AI will unjustifiably seep into the notion of AI operating in the real world. This is the *sci-fitisation* of the discourse.

Why is this *sci-fitisation* important for the topic under consideration? *Sci-fitisation* entails several problems for discussions relating to the nature of modern technologies. The main problem with the debate about AI is the aforementioned "conceptual drift" towards fiction. As a result, natural language speakers project the characteristics of fictional AI represented in their culture onto actual AI. The notion of AI becomes, seemingly, infused with what it represents in culture, although AI in culture is a very different AI than in reality. Here, then, the risk of equivocation arises: Someone may not notice that when they refer to AI as (existing) technology, they are referring to a different concept than when they talk about AI as a cultural being (or, viewed more broadly, a kind of imaginary).

Moreover, because of the conceptual drift, AI is sometimes (unintentionally) used as an *empty name*. The definition of AI has truth conditions that are sometimes too challenging; hence, it essentially lacks any referent. A discussion on the philosophical characteristics of AI technology is sometimes based on premises that cannot currently (or at all) be fulfilled in the real world, and,

hence, to define AI in a manner that makes the name empty—it only has referents in fictional worlds, which, unfortunately, are too strongly activated in the linguistic intuitions of the disputants. When the discussants realise that they are discussing the strictly philosophical topic of strong AI (Butz, 2021)[2], the problem is virtually nonexistent. However, when futurological reflection and current reality are combined in the minds of the discussants, problems arise.

We fear that because of the prevalent *sci-fitisation*, as well as anthropomorphism (Placani, 2024; Salles et al., 2020) and *hyping* (Sloane et al., 2024) in the discourse on AI (the three categories are not mutually exclusive; on the contrary, they can be considered complementary), the "real AI" is often combined with the "imagined AI" and existing AI technologies are interpreted within the AI imaginary created by popular culture and marketing, abstracting from technological limitations. This seems to represent a trend towards overinterpreting the features of computational technology and the lack of conceptual rigour in describing them. This trend has been present in the discourse for decades and has manifested itself in, for example, in the attribution of human-like qualities to the simple chatbot ELIZA (Weizenbaum, 1966; Berry, 2023) or in the anthropomorphisation of the language used to describe AI algorithms, which McDermott (1976) warned against half a century ago, and which remains problematic today (Shwartz, 2020). All the phenomena described in this subsection are, therefore, nothing new, but it is important to be aware of them in discussions about the characteristics of AI. Floridi and Chiriatti (2020), in their remarks on GPT-3, indirectly refer to the problem of *sci-fitisation*: "Hollywood-like AI can be found only in movies, like zombies and vampires" (p. 690). For some, this statement will sound like a truism; however, for others, affected by *sci-fitisation* and other previously mentioned problems, this statement will be unacceptable. The latter will probably not focus on understanding why existing AI does not have a Hollywood-like nature (beyond cunning marketing messages) but rather on the illusion of ground-breaking qualitative technological progress. They may say, "But, but! It's a paper from 2020, and it discusses the GPT-3 model. Now we already have GPT-5 and Gemini 2.5 Pro models, and they *really* are Hollywood-like AI!" Arguably, in the 1960s, some people remarked similarly after interacting with the ELIZA chatbot. To avoid succumbing to these illusions and to understand where they originate, we need to be aware of *sci-fitisation*, anthropomorphism, and *hyping*. However, there are other problems with the concept of AI.

Vagueness and the open texture of the term "artificial intelligence". We prefer to avoid associations with the kind of fictional AI that does not and may not exist for a long time. Instead, we focus on AI as *a nonempty name* that refers to a specific class of technology. Alas, AI is an ungrateful concept even if we use this notion to denote technological entities or directions in research. In a discussion among a heterogenous group (that cannot always operate with a similar concept of AI), one can (and even should) replace "AI" with more precise terms that describe specific classes of technology such as machine learning (Porębski, 2023) or deep learning (Goodfellow et al., 2016; for an introduction, see LeCun et al., 2015; for a less technical overview, see Janiesch et al., 2021).

Apart from the notion that the concept of AI is prone to *sci-fitisation*,

- "AI" is unclear. "AI" can be used in various contexts, and even in the same context, it has many different meanings. Many definitions of AI exist, and each differs significantly in scope (e.g., "AI" as "technologies that create inference rules on their own on the basis of data" vs "AI" as "technological intelligent entities, characterised by intelligence such as that of a human"; the latter is currently an empty name).
- "AI" is particularly vague. Only specific approaches to clarifying the meaning of "AI" can be successful, but they will be inflexible (e.g., unduly broad if they rely on enumerating the programming approaches that constitute AI). After most approaches have been applied, determining which technologies can already be referred to as "AI" will remain challenging (Fernández-Llorca et al., 2024).
- "AI" is especially open textured (Waismann, 1945; Vecht, 2023), i.e., highly dependent on the technological and cultural context. For example, in developed countries, many advanced computing technologies have already been domesticated, taken for granted, and are not considered a manifestation of machine intelligence.

However, in discussions about the nonexistence of conscious AI, we have to use the term "AI". Since this is the case, we will attempt to define it.

Attempt to clarify the concept of artificial intelligence. We make the following assumptions about AI:

- AI is a system—which can have a connection with external devices or can be just a program.
- AI is based on specific technologies—which enable it to perform its own "reasoning".
  - "Own reasoning" is the process by which the system performs an action that was not directly declared by the programmer; that is, it derives a course of action on the basis of general rules, rather than having a preset behaviour in each case.
  - Particularly, machine learning is the technology that has great potential for the creation of "intelligent" systems, because in the process of learning, an algorithm is "created by itself"; that is, the programmer (or the developer of the system) defines the data, the learning method, and the goal (formally, an objective function to be optimised) and the machine learning algorithm estimates the best parameters of the decision rules (Porębski, 2023). Thus, the rules are derived from the data, not predefined by the programmer.
- AI performs tasks that require a high degree of flexibility; for example, one cannot define each part of the task in a top-down manner.

Since we are discussing existing or foreseeable technologies, we consider the following rather uncontroversial examples of AI or AI-driven devices (we use these terms interchangeably, because their scope relies only on how broadly one understands "the system"):

- Autonomous vacuum cleaners, which are common devices, perform cleaning tasks that require autonomous movement in different rooms and under different conditions. They are based on technology that allows them to adapt to different environments and respond independently to adversity, such as previously unseen obstacles. They can "experiment", such as when they try to become unstuck.
- Autonomous cars are designed for a significantly more extensive range of environments than vacuum cleaners, collect vast amounts of data and must process them sensibly, and operate with a very high level of precision (they should not "experiment").
- Humanoid robots such as Sophia (Parviainen and Coeckelbergh, 2021; Hanson Robotics, 2024), which is a chatbot with a case and output devices that resemble a human body. These robots try to embody artificially intelligent humanoids from popular culture.

Note that these three examples are AI systems connected to external devices. For example, an autonomous car has (1) sensors that collect data important for system operations; (2) a decision system based on AI technology, which is AI in a narrow sense; and (3) an external device, that is, a car (which can be driven by the driver or by the AI system).

Many examples AI systems are not equipped with strictly external devices; they are just programs on a computer (or a smartphone) that have an interface through which a user can communicate with the program.[3] The most prominent example of such AI systems is the new generation of chatbots and its most well-known representative, ChatGPT by OpenAI (the current version of a general-purpose model is GPT-5; OpenAI, 2025). Since we mention ChatGPT, we should also mention other related AI systems, such as Gemini by Google (Pichai and Hassabis, 2023), Llama by Meta (Touvron et al., 2023; Grattafiori et al., 2024), and Claude by Anthropic (Anthropic, 2024). These chatbots are based on large language models (LLMs). Indeed, LLMs are large (Bender et al., 2021; for a detailed comparison of the number of parameters, see Naveed et al., 2025) and transform text[4] well; however, one must be careful not to demand too much from them. First, they hallucinate, i.e., produce plausible-sounding but false or fallacious statements (Kalai and Vempala, 2024; Maleki et al., 2024). Second, they struggle with many aspects of reasoning (Shwartz, 2024; Zhou et al., 2024; Huckle and Williams, 2025). For example, several LLMs still cannot correctly answer questions like whether one can seat John, Matthew, and Taylor at a small round table in such a way that John does not sit beside Matthew. Huckle and Williams (2025) identified several categories of LLM failings: (1) overfitting (resulting in, e.g., classifying prompts similar to well-known puzzles as equal to these puzzles despite significant changes); (2) lack of logic or common sense; (3) lack of spatial intelligence; (4) incorrect mathematical reasoning; (5) poor linguistic understanding; (6) popular science; (7) relational misunderstanding; and (8) illogical chains of thought. Remarkably, in the benchmark created by Huckle and Williams (2025), the average human performance (accuracy of approximately 85%) is better than the performance of even the most advanced models supposedly created for advanced reasoning (such as Gemini 2.5 Pro or OpenAI o3, accuracy of approximately 75%), let alone more widely available models such as GPT-4o (accuracy of approximately 40%) or Gemini 2.5 Flash (accuracy of approximately 60%) (Huckle, 2025). Given these results, to accept the claims of the outstanding intelligence of LLMs, one must first disregard the substantial fields in which this "intelligence" is suspended.

If LLMs are conscious, why not autonomous vacuum cleaners? Why did we provide other examples of AI—e.g., autonomous vacuum cleaners—alongside the prominent LLMs? In discussions about the characteristics of AI, it is paramount to remember that generative models are only one of the numerous examples of this technology. If one wishes to attribute a feature only to generative AI models, one must have a reason against attributing that feature to other AI technologies. In particular, if one postulates AI consciousness in reference to generative models, one should point out what about their structure makes them conscious, unlike, for example, autonomous cars. This reasoning also works in reverse: If one postulates consciousness specifically for generative models and uses the reason R to explain why they single out these generative models, they should keep in mind that they are not postulating the consciousness of AI in general, but only consciousness of the specific AI solutions that are singled out because of R. This person should consider all the consequences. If R were, e.g., a very large number of parameters in a neural network, this person should remember that a scaled-down version of this model (so-called small LLMs or small language models, SLMs, see Wang et al., 2025) will lose consciousness. Similarly, this person should remember that when an autonomous truck starts using a neural network so deep that R is met, the truck will gain consciousness.

It is important that we consider these issues because we see in discussions the problem of treating AI as one large entity, with characteristics determined by the best existing technologies at a given time (a partially similar tendency to treat AI as a highly homogenous group rather than as a collection of diverse solutions is supported by research of Manoli et al., 2025). AI is, strictly speaking, the field of research that researches and produces AI systems (however, for the sake of convenience, we have defined AI in this paper so that it refers to AI systems rather than the field of research). Each system is a separate entity; it is not possible to combine their features and select the best from each system. Therefore, if one postulates the consciousness of LLMs, one should excuse oneself for not postulating the consciousness of extremely high-tech autonomous cars or fantastically useful autonomous vacuum cleaners that astonish us with their ingenuity when they avoid obstacles.

However, it is hardly coincidental that the discussion of consciousness intensified after the widespread use of LLMs. In addition, a study by Colombatto and Fleming (2024) investigated public attitudes towards the consciousness of LLMs, and most respondents considered LLMs more conscious than a toaster. Therefore, we consider LLMs to be the main candidate for "consciousness" in the public discourse, and we will base our argument against AI consciousness not only on general arguments that apply to all information technologies but also on claims specific to LLMs. We now discuss the main claim of this text: There is no such thing as conscious AI.

The justification for the answer: why is there no such thing as conscious artificial intelligence?

We will begin with the general argument against the consciousness of AI (the "Biological argument for the nonconsciousness of artificial intelligence" section). We believe that this argument applies to any currently known large-scale computing technology. In the following, we present several counter-arguments to the claim that LLMs would be excellent candidates for consciousness (the "Critical remarks on the arguments for LLM consciousness" section).

Before we continue, we re-emphasise that in our paper we focus on the present state of the technology and the foreseeable future. We do not refer to what may be possible on a purely philosophical level, nor do we refer to whether it will ever be possible to create a conscious technical entity. In particular, we do not attempt to predict the characteristics of entities that will be human-made but will arise from currently unknown technologies (e.g., use of biological mechanisms, new ways of performing computation, etc.).

Biological argument for the nonconsciousness of artificial intelligence. At this point, we examine the AI body. This view is available to most of us; anyone who owns a computer also has a graphics card (see Fig. 1).

Most of the calculations in data centres that make AI models "alive" are currently performed on graphics processing units (GPUs).[5] Thus, we can view the "body" of AI systems (of course, larger systems consist of hundreds or thousands of similar devices and dwarf personal computers). Humans are deeply knowledgeable about this technological device; in particular, we know that its functioning is strictly algorithmic. We know that, at the lowest level, this device is based exclusively on binary instructions and the operation of semiconductors.

Fig. 1 A part of the body of a possible AI system—an example of the graphics processing unit (ASUS TUF RTX4070 Ti SUPER). Attribution: 极客湾 Geekerwan, CC BY 3.0 <https://creativecommons.org/licenses/by/3.0>, via Wikimedia Commons.

Notably, we know that the GPU works at the lowest level in a similar way regardless of the actions it performs. Suppose we assume that AI has a conscious nature. Then, for the sake of consistency, the reader must also recognise that at this point he or she is looking at a conscious object that consciously generates images on his or her screen.

However, few people consider calculators conscious entities or potentially conscious entities. Why should these tools become conscious just because they start performing more complex operations, creating more advanced illusions from our perspective? Consider this as an analogy. Are modern computer games with photorealistic graphics literally more real than old games, with inferior graphics? No one postulates so, of course, because everyone believes (and even feels) that both old and new games only create an image of reality. The characters in the old and new games are just as fictional, although the newer games portray them as much more life-like. The new games are no more *real* than the old games; they are only more *realistic*. The difference between a calculator and an LLM is the same as between an old computer game and a new computer game.

Someone may argue, however, that there are also simply physical processes happening in humans at some level, but they result in the phenomenon of consciousness. This is true, but, as we have mentioned, the phenomenon of human consciousness arises within a very complex biological structure and in an extremely elaborate neural network, via numerous biological phenomena (e.g., neurotransmitters or embodiment), both in the brain and in the rest of the body. In contrast, we know exactly what is at the heart of the algorithms referred to as "AI": decision rules, which, however, can be reduced to a sequence of 0 s and 1 s (executed by the processor). We cannot completely rule out that the phenomenon of consciousness is also possible in simple machines. But if that is the case, why has the discussion about this subject only intensified now, when these machines are performing tasks previously associated with essential human input, and why has the discussion not persisted for decades, knowing that computers already operate on binary code?

To make this comparison comprehensive, we should note another important characteristic of human (or, more broadly, animal) consciousness. The mechanism of its emergence is very energy efficient. The daily functioning of a human consuming approximately 2000 kCal translates into a consumption of approximately 2.32 kWh. The brain itself uses approximately 400 kCal per day, which is less than 0.5 kWh, and its energy processes appear to be optimised for low energy consumption (Balasubramanian, 2021; Padamsey and Rochefort, 2023). As Chen (2025) reported, a lower bound estimate of energy consumption during image captioning by an LLM is 105 Wh per 1000 tasks performed. Therefore, the amount of energy comparable to the daily usage of a human brain—approximately 0.46 kWh on average—is used by an LLM to generate approximately 4400 image captions. This number is only seemingly large. Just imagine that during each minute of the 16 h of functioning each day, all you can do is recognise and characterise four (low resolution) images provided by your eyes, because your brain would not have the energy for any other activity. It is easy to guess that such functioning would be extremely far from the sensation of being seamlessly, smoothly in the world—and we are considering here the energy consumption of smaller LLMs, which are often much more unreliable than human perception mechanisms. If we consider larger models (with 70 billion parameters)—those with greater capabilities—energy consumption for text generation reaches 2 kWh per 1000 tasks performed (Luccioni et al., 2025). Moreover, according to estimates by Gamazaychikov (2024), approximately 1785 kWh is required for the complex OpenAI o3-preview LLM to perform a single complicated task. Later, Gamazaychikov (2024) updated and significantly lowered his estimates for the OpenAI o3 (not preview) model to 1.32 kWh per complex task, which is still completely incomparable to the energy consumption of human functioning.

An analysis of the operation of autonomous vehicles leads to similar conclusions. For a human, the act of driving is energetically undemanding, and hypothetically propelling the mass of the car with one's own muscles would be absurdly energy intensive (and require a dozen people). For an autonomous car, automated driving is the main challenge. The energy costs of autonomous driving can be similarly as high as the costs of driving the vehicle itself (Grant, 2024). An hour of autonomous driving can result in an energy consumption of 0.5–2.5 kWh (Rajashekara, 2024), completely incomparable to the negligible energy effort for a human (the energy consumption of the brain is almost independent of the task performed and driving the car is not physically demanding).

All these data show how much energy is consumed when AI performs tasks that may be simple and energy efficient for a human to perform (e.g., image recognition) and that must be performed extremely frequently (thousands of times each day) to maintain human consciousness.[6] This radical disparity in energy consumption shows that the consciousness that arose during the course of evolution (Dennett, 2018) must be a sophisticated mechanism that is not based on the "brute computing power" that characterises, e.g., LLMs or many autonomous devices. Given this gap between the efficiency of the human mind and LLMs, it becomes plausible that AI technologies are far removed from the biological substrate of consciousness.

Critical remarks on the arguments for LLM consciousness. We believe that the biological argument presented for the lack of consciousness in AI applies to all classes of this technology. However, because many arguments in favour of a possible conscious AI refer to conscious LLMs, after we present a general biological argument, we will discuss the specific case of AI and consider selected factors that may cause people to attribute consciousness to LLMs.

Does the linguistic ability of LLMs indicate their consciousness? LLMs are peculiar because they generate plausible linguistic utterances that are grammatically perfect. This exceptional characteristic is a basis for many claims in favour of the presence of consciousness in LLMs. Because they act like they are fluent in language, they can erroneously be associated with humans. Although the current consensus is that language is not a

necessary condition of consciousness, with respect to humans, consciousness is closely linked with language (Kuhn, 2024, p. 79). Language enables individuals to articulate and report their feelings and sensations. For example, Skipper (2022) argues that higher-order consciousness emerges due to the linguistic abilities of humans. Philosophers acknowledge that language influences consciousness by enriching, organising, or shaping this phenomenon (e.g., Searle, 2009; Kuhn, 2024, pp. 78–81). Wittgenstein (1922) famously remarked, "The limits of my language mean the limits of my world" (p. 74)—suggesting a (rather strong) connection between language and the way of being in the world. Dennett (1997) extends this idea, indicating that language actively facilitated the evolution of consciousness, particularly through practices related to self-reflection. Dennett (2005) even (cautiously) suggests that "acquiring a human language (an oral or sign language) is a necessary precondition for consciousness–in the strong sense of there being a subject".

Even if one accepts the strongest claim that language is a necessary condition for (higher-order) consciousness, there is still a slight justification for regarding the ability to use language as sufficient for any form of consciousness. One can imagine the process of randomly generating meaningful sequences of strings without any conscious traits… Especially in the current times, it is easy to imagine, may we note? Even if, contrary to consensus, it is recognised that a lack of linguistic ability results in a lack of consciousness, this does not result in the conclusion that linguistic ability causes consciousness.

However, people often make mistakes when abstract implications are considered (Wason, 1968; Ragni et al., 2017). Arguably, because of reasoning fallacies, a necessary condition is sometimes perceived as a sufficient condition—or the presence of a consequence is considered to make the assumed cause plausible (although the actual cause may have been different). In the case of LLMs, the close relationship between consciousness and language tends to be overinterpreted. Some people, habituated to the idea that only humans use natural language, may use the ability to use language as an argument for being conscious. However, even the highest linguistic abilities should not be used as an argument when we have reasons to believe that they are the result of mere imitation.

Does the generation of statements claiming their own consciousness by LLMs indicate their consciousness? Numerous arguments in favour of the consciousness of LLMs refer not only to their excellent use of language but also to the notion that some chatbots sometimes (in a given context) declare themselves to be conscious. Why do such declarations occur? Consider typical language usage. The sentence "I'm not conscious, I'm just a machine" makes perfect sense for all machines but is unlikely to be made in most contexts. How often in discussions about consciousness does someone say, "Remember I'm just a calculator, calculators aren't conscious at all"? Not too often, or at least so we guess. People typically talk about their consciousness, not their nonconsciousness. Language reflects this pattern, and LLMs reflect this pattern. Therefore, it should come as no surprise that in specific contexts, LLMs may tend to suggest their consciousness.

However, LLMs were not created to make true statements but rather to make statements with probable wording. The veracity of these statements is adventitious: they are true or false (or of difficult truth status, if an LLM generated a statement about nonexistent objects), but this status is extraneous to their probable wording. There are topics—mostly specialist, such as those related to legal problems—in which the generated statements will often be false, although they may still ring true (Porębski and Figura, 2025). For example, false legal advice hallucinated by ChatGPT will likely confuse any layperson and many lawyers.

The probabilistic sentence generation mechanism may cause LLMs, in specific contexts, to attribute to themselves various qualities that humans would rightly title themselves with (e.g., "conscious"). However, the LLM attributes them to itself only because of the structure of human language. Hence, in our opinion, the illusion of consciousness in some texts produced by the model is a side effect of solving the problem for which the LLMs were designed—creating highly probable text (wording) in a given context. In practice, many modern LLMs are protected against generating specific types of statements; that is, they have restrictions regarding claiming consciousness in their system prompts (the most general instructions on how to operate during interaction with a user). Claude Opus 4 and Sonnet 4 are not supposed to claim their consciousness or lack of it but should engage with questions about their consciousness (Willison, 2025), and GPT-4.5 is perhaps prohibited from claiming that it is conscious (Jim the AI Whisperer, 2025). Interestingly, the attempt to make LLMs suitable for conducting a human-like conversation implies the identification of LLMs as separate entities, and thus a tendency to attribute human qualities to themselves. This is most likely why LLMs without safeguards were so convincing in talking about themselves as "sentient beings", and even some secured chatbots, if they are cleverly prompted, will be able to talk convincingly about their consciousness.

Finally, note that the attribution of a characteristic by a specific object does not constitute an argument for possessing it, when that object, as we know, is simply intended to reflect the attributions likely in the language. Moreover, as long as we are not used to attributing consciousness to any technology by default (the inanimate world, which would include computers, is not perceived by humans as conscious and there is no good argument that it should be), the burden of proof must be shouldered by those who postulate the existence of AI consciousness, not on those who deny it.

Are the actions of LLMs consistent enough to indicate their consciousness? Some people, among the many contradictory claims of LLMs about their own qualities, choose to believe the claims that give an LLM stronger properties (such as consciousness); that seems strange. If an LLM model is asked in a zero-shot procedure (a specific question, without examples or rolling a prior conversation) whether it is conscious, the most likely answer will be "no" or, at most, "there is no scientific consensus, there are arguments for and against". Then why would one assign relevance to suggestions that an LLM may be conscious, if these suggestions only arise in long conversations with elaborate contexts or in the case of very unusual prompts, and contradict zero-shot-normal-usage LLM claims? Creating a contradictory image—as LLMs tend to do because of their alignment with likely responses in a given context, and responses most likely in different contexts may contradict each other—used to hint at the low credibility of the speaker. Similarly, if some LLMs sometimes perform very well with language and other times perform very poorly with language (for examples, see Zhou et al., 2024; Huckle and Williams, 2025), it is difficult to find a reason to select only the best examples when their abilities are considered, especially when these particular abilities are to be the basis for recognising consciousness. It seems reasonable to consider the instability of their performance, and when it is, the abilities of LLMs—while impressive—are much weaker than those that result when taking into account only the tasks they perform best.

We point out that if the issue of the consciousness of LLMs is unclear to someone (although it remains clear to us), then they

should not ask LLMs about it, because they will receive from the model an answer that sounds likely but is not necessarily true and, moreover, strongly dependent on the prompt. If one really wishes to rely on these answers being true, that person should note that these answers may blatantly contradict one another. The inconsistency between models and within different outputs of the same model further supports the argument that neither the advanced task successfully solved by LLM nor the claim by LLM about its consciousness inform the question of its consciousness. If we assume that technological objects are not conscious, then the opposite claim by the LLM does not provide any argument for separating LLMs qualities from the qualities of IT in general.

Does passing the Turing test indicate consciousness? Some people suggest that passing the Turing test by AI is an argument for AI consciousness (Dickens, 2025). Indeed, according to the initial findings, some LLMs passed this test (Jones et al., 2025; Jones and Bergen, 2025). In this case, we agree with Prettyman (2024) that passing the Turing test does not provide any reason to infer consciousness of AI. The victory of a machine in the Turing test indicates only that the algorithm performs its task well—the task of effectively imitating a human being. In the case of LLMs, this is to perform a human-like natural language conversation by generating text. However, successfully pretending to be human is proof of nothing more than the ability to successfully pretend to be human. Similarly, even the most skilful performance of Hamlet does not prove that the actor is Hamlet—or even that they are the prince of Denmark.

Does the fact that an LLM believes in something indicate its consciousness? Let us draw attention to a logical problem with some claims about AI consciousness. As we have mentioned, Hintze's (2023) text about ChatGPT postulates in the title: ChatGPT *believes* it is conscious. But please remember, "to believe" requires one "to think"; believing requires thinking. Thus, already in the title, Hintze presupposes a strong assumption about the nature of ChatGPT, namely, that the text generation it performs is connected to a thinking process. To interpret the actions of AI as functions such as believing, one should first assume that AI has, at least, functions such as thinking or (broadly understood) intentionality.

What is the basis for the assumption that the actions of AI (LLM) can be interpreted in relation to its thinking when we know what this model does and that it was not created to think but to generate text? Hence, this assumption seems to be unfulfilled (Crespo, 2024). This finding reveals the implicit *petitio principii* fallacy (Van Eemeren et al., 2014, p. 168) of drawing conclusions from an interpretation of the actions of an LLM that requires assumptions about the thinking we want to prove. One cannot prove consciousness on the basis that one has noticed the functions of an object, the recognition of which already requires the assumption of a particular form of consciousness or related characteristics.

Does the architecture of LLMs enable them to achieve consciousness? Finally, at the end of this section, a purely technical issue should be discussed. Most popular LLMs such as GPT, PaLM or LaMDa are based on a decoder-only transformer architecture (Radford et al., 2018; Greco and Tagarelli, 2024). Therefore, these LLMs are created primarily to generate text rather than "understand" it. The procedure of generating text is purely computational and aimed at predicting tokens probable in a given context, which should compose coherent sentences. With this approach, unlike human use of language, there is no intention behind the strings of characters generated by LLMs. Thus, as Gubelmann (2024) points out, LLMs cannot be considered speakers. Lacking intention and embeddedness in a communicative community, their language use is stripped of reference to meanings (those existing in the world rather than represented numerically in the vectors), making them comparable to stochastic parrots (Bender et al., 2021). These limitations stem from the technical architecture (decoder-only) itself, which does not include layers designed to ensure text comprehension, but only those ensuring its one-sided autoregressive processing (prediction of the sequence of subsequent tokens is based on a very extensive context provided in previous tokens). We do not intend to suggest that models operating bilaterally (encoder or encoder-decoder) would be conscious—you do not consider Google Translate or DeepL to be conscious, do you?—but rather we wish to emphasise that it is quite symptomatic that even the unilateral nature of processing does not prevent some people from postulating the consciousness of these models.

Prettyman (2024) partly addresses our objection, pointing out that designing LLMs to produce language does not mean that LLMs will only do that—the system may generate new capabilities to achieve its primary goal. While this is true in general, it is difficult to consider this argument valid in relation to a specific class of algorithms when we know exactly what this class of technology can do to better achieve its goal. Specifically, LLMs can learn to work better with language by more effectively adjusting the numerical values of their parameters. That is exactly what they can do—unless humans extend LLM-based systems with other functionalities (e.g., automatic use of external tools to perform tasks that LLMs cannot perform alone—such as displaying the current time). Indeed, this adjustment of parameters provides enormous possibilities for adaptation to the data. Nevertheless, it is difficult to imagine that consciousness arises because of the fine-tuning of mathematical structures so that they better reflect language. If that were true, would it mean that consciousness is created when the numbers constituting an artificial neural network are in the appropriate area of the continuum and disappears when someone—or an algorithm itself, during additional training—modifies those numbers? This theory seems very counterintuitive.

Furthermore, if the above scenario was true, we again pose the following questions: Why is the debate focused on LLMs and not, for example, on autonomous cars? Why has it not also been conducted in relation to much simpler machine learning systems? If we follow Prettyman's argument, simpler machine learning systems theoretically could produce new qualities.[7] Yet, before the popularisation of LLMs, few people seriously considered that machine learning systems could have consciousness if their parameters were adjusted sufficiently. This finding suggests that in terms of computer algorithms that have a known and limited, even if very extensive, scope of action, it is not typical to consider their spontaneous production of extraordinary abilities.

Potential limitations

We have proposed arguments against the recognition that AI can be conscious. To render our thesis more resilient, we present and address potential limitations. We examine the popular method of arguing for the presence of new qualities in AI (such as consciousness) from the vast complexity of some AI algorithms resulting in black boxes. In addition, we will pre-empt the charge that our argument may be anthropocentric.

Great complexity and algorithmic black-box nature of some AI models. The human brain is an extraordinarily complex structure. It is estimated that the brain has 86 billion neurons (Azevedo et al., 2009; Ray, 2024). We can describe many physical and chemical processes that occur in the brain. However, as we

have mentioned, the phenomenon of consciousness is extremely difficult to explain; there are many hypotheses about the nature of consciousness and its relationship with the biological substrate (Kuhn, 2024). Thus, one can assume that the current scientific knowledge falls short of providing unequivocal answers to the questions of what consciousness is and how it works. Do we as humans, therefore, doubt that other humans are conscious? Barring fringe philosophers, we do not; insufficient knowledge of consciousness does not prevent us from assuming that other people are conscious.

Some people can argue that the field of AI research manifests similarities to brain research. AI scientists understand the theory and development of AI algorithms well and can even precisely describe the operation of, e.g., a single artificial neuron. However, if the AI model is highly complex, the same AI scientists cannot explain why a particular result came to be. By analogy, in the face of an exploratory gap, can we categorically claim that AI is not conscious when a similar gap does not lead us to similar conclusions about humans (i.e., that they are not conscious)?

For the analogy to be valid, significant commonalities between not understanding consciousness and not understanding AI algorithms would be required. We claim, however, that these cases are more different than alike. The general principles of AI algorithms are well known to us. People design them to achieve specific goals. One cannot exactly predict the specific algorithm results, but one knows that techniques and mathematical tools such as matrix multiplication, softmax function, gradient descent, encoders and decoders layers, are used to obtain these results (Goodfellow et al., 2016). With respect to consciousness, things remain much less clear. Various models aimed at explaining consciousness have been proposed, but science is still unable to reveal causal links between the biological substrate and the emergence of consciousness. In contrast, our knowledge of the functions and methods behind the operation of AI algorithms enables us to state unequivocally what the operations performed by these algorithms are and, more importantly, that there is no basis for believing that these operations result in the emergence of consciousness. Because AI algorithms rely on mathematical operations performed on silicon-based devices, when we claim that their operation results in consciousness, we may as well presume the same for advanced calculators.

Anthropocentric perspective. Another counterargument to consider relates to criterion-based attribution of consciousness. Humans make attributions of consciousness based on criteria. Humans used to believe that animals were not conscious because they lacked the ability to use language (Andrews, 2024). However, the chosen criterion turned out to be defective. The current scientific consensus rejects any grounds for excluding the consciousness of animals with neuroanatomical similarities to humans. Therefore, one can argue that we are wrong to exclude AI consciousness on the basis of the criterion of no biological substrate, because we do not have objective criteria for such an evaluation (Sotala and Yampolskiy, 2015). How can we, as humans, be certain about what nonhuman entities can and cannot do?

We agree that we cannot be certain about criteria that have been conventionally created within a specific conceptual framework created by humans. Nevertheless, we claim that as long as these criteria lend themselves to rational evaluation, we cannot dispose of them simply because they were created by humans. For example, the neuroanatomical criterion enables us to exclude stones, plants, and even sponges from the set of conscious entities. We believe that this demarcation is not contentious.

We agree with Kiškis (2023) that "humanity can never be certain about the capabilities of nonhuman entities". However, we must remember that this is a universal cognitive limitation which, out of epistemological necessity, accompanies every human reflection, and is not unique to the AI debate. Thus, considering the limitation arising from the inevitable element of anthropocentrism in human cognition, we believe this counterargument can be refuted by pointing out that at no point did we "escape into anthropocentrism". Notably, the criteria that we applied within the "minimalist approach to consciousness" would easily be met by some animals, at least partially. More importantly, the counterarguments provided in the "Critical remarks on the arguments for LLM consciousness" section abstract from the human/nonhuman characteristics of AI. The focal point of these counterarguments is overly strong assumptions about some AI instances (mainly LLMs). Therefore, this argumentation, focused on the reasoning rather than on the object, is less prone to excessive anthropocentrism.

Final remarks

Much has changed over the past half-decade since Floridi and Chiriatti (2020) stated that GPT-3 is as conscious as an old typewriter. Currently, GPT-3 is outdated and, from today's perspective, deficient. GPT-5 is much more advanced, and GPT-500+, which is probably just around the corner, will eclipse every previous version in terms of capabilities and (above all) marketing narrative. However, according to our conclusions, Floridi and Chiriatti's (2020) paper remains highly relevant because one thing has not changed. Every subsequent version of GPT (regardless of what label it will have) and every other advanced AI system that may be created in the foreseeable future will be as conscious as GPT-3 (that is, as conscious as an old typewriter). In this paper, we have reached the same conclusions as Gams and Kramar (2024), who argued that LLMs "function as advanced informational tools rather than entities possessing a level of consciousness that would warrant their categorization alongside sentient beings" (p. 234). Andrews (2024) also seems to share our intuitions, explicitly ruling out the consciousness of LLMs despite the notion that they are "skilled in linguistic processing" (p. 426).

Our conclusions provide practical implications on at least two levels. First, on the societal level, our considerations lead us to recognise the problem that Floridi (2020) refers to as "semantic pareidolia": an interaction with some types of AI systems has become so similar to an interaction with conscious humans that people began to see consciousness and intentionality instead of merely (very complicated) algorithms. This is a very risky phenomenon. Thus, people need to be made aware that they are only interacting with nonconscious entities. The public should be protected from overinterpreting the functionality of AI and gain the knowledge that it is only (and as much as) a tool, just as the cinema-goers should be (and usually are) aware that it is only a picture, not reality.

Second, our conclusions have implications for legal debates. The rejection of AI consciousness makes it clear in discussions on AI regulation that the regulation is not (and will not be for at least several years) meant to protect "artificial but ethical entities", but rather to address practical issues that arise from the complexity and potentially powerful effect of these technologies. In turn, in discussions about the hypothetical legal personhood of AI, it is useful to keep in mind that the alleged consciousness should not be an argument for granting this personhood; if granting it were merited at all, this would be attributed entirely to different factors.

A cliché says that at the first screenings of the Lumière brothers' film "The Arrival of the Train at La Ciotat Station", people fled the cinema when they saw a train coming towards

them.[8] Radical overinterpretation of technological achievements is becoming more likely than ever. People would do better if they did not treat the silver screen as a reliable depiction of the real, and fortunately, they rarely do. Still less should they believe that whatever shows up in the AI interface is actually thought or realised by the generator. Unfortunately, the latter threat seems real, which we have attempted to address in our study.

We are afraid that claims about AI gaining consciousness are currently intensifying because the resulting technologies create better illusions and associate more with humans (they "speak" our language). While they are not made any more human by better replicating human language, they are perceived as more human. Unfortunately, this is a blind spot in philosophy that disregards the superficiality of progress and gives credence to the proficient rhetor who puts their thoughts in a golden frame, pretending that they are gold. This is what Aristotle (ca. 350 B.C./1955) warned against in "On Sophistical Refutations".[9]

Data availability

No datasets were generated or analysed during the current study.

Received: 15 May 2025; Accepted: 1 September 2025;
Published online: 28 October 2025

Notes

1 Capacities were defined as: subjective experience—"Capable of having experiences that feel like something from a single point of view"; self-awareness—"Capable of being aware of oneself as an individual". A precise question was: "For each [capacity], please indicate the earliest year when you think an AI system will exist that is more than 50% likely to have this capacity"; answering "2024" was interpreted as "now" (Dreksler et al., 2025, p. 91). Although these capabilities are not the same as consciousness, they are closely related to it. Therefore, we consider these statistics to be relevant.

2 The term "strong AI" was introduced by Searle (1980) in his famous paper presenting the "Chinese Room" argument.

3 Precisely speaking, many AI computer programs or smartphone applications use external devices such as a keyboard, a microphone, or a camera (a user inputs data using them) and a screen or speakers (a program outputs results using them). We have distinguished in this subsection "AI systems without strictly external devices" because external devices used by these systems are needed only to enter the input and receive the output (which is the case in many programs). In this sense, these systems have no need to cooperate with any specific external device. They simply need to be provided with input data somehow (you may as well change your keyboard to a mouse and click letters on the screen) and output data somehow (but they could not display the data but instead print it out), which distinguishes them from "AI systems connected to strictly external devices", such as autonomous cars, whose correct operation is conditioned by their connection to an actual car and sensors informing about the state of the external world.

4 More specifically, LLMs are based on deep learning transformer architecture (Vaswani et. al., 2017), which uses multihead attention mechanisms to compute probability distributions of the next-tokens. Generative pre-trained transformer (GPT) models are based on decoder-only transformers and can generate text by predicting the tokens. Not all LLMs are generative. For example, bidirectional encoder representations from transformers (BERT) model, which is based almost only on encoders, cannot generate new text but is rather used for text classification and language understanding tasks (Devlin et. al. 2019). For a more detailed explanation, see Atkinson-Abutridy (2025).

5 The reason for this is the high performance of GPUs when they operate on data structures related to machine learning and the ease of parallelising calculations.

6 Please note that we analyse only energy consumption that is related to performing tasks. To obtain a complete picture of resource use, energy consumption for training and fine-tuning a model must also be included. Although we do not analyse this complicated issue in this paper, we provide several estimates to show the scale of values. Luccioni et al. (2023) estimated energy consumption during the training of BLOOM—an LLM with 176 billion parameters—as 433,196 kWh (such an amount is enough for an estimated 500 years of functioning of one human being); the training of GPT-3—an LLM with 175 billion parameters (or greater)—used approximately 1,287,000 kWh (de Vries, 2023).

7 Notably, algorithmic black boxes, i.e., models whose complexity is too great for humans to understand or interpret the exact decision-making rules used by the model (Porębski, 2024), existed long before LLMs and characterised many systems based on, e.g., artificial neural networks. Therefore, LLMs have not introduced a previously unknown level of intransparency in their operation that would justify considering their consciousness.

8 This is a far-stretched anecdote, intended only to rhetorically highlight how agitated the audience was (Grundhauser, 2016).

9 "That some reasonings are really reasonings, but that others seem to be, but are not really, reasonings, is obvious. (…) So too with inanimate things; for some of these are really silver and some gold, while others are not but only appear to our senses to be so; for example, objects made of litharge or tin appear to be silver, and yellow-coloured objects appear to be gold" (Aristotle, ca. 350 B.C./1955, p. 11).

References

Alonso M (2023) AI Tinder already exists: real people will disappoint you, but not them. El País. Retrieved December 29, 2024, from https://english.elpais.com/technology/2023-10-09/ai-tinder-already-exists-real-people-will-disappoint-you-but-not-them.html

Andrews K (2024) "All animals are conscious": shifting the null hypothesis in consciousness science. Mind Lang 39(3):415–433. https://doi.org/10.1111/mila.12498

Anthis JR, Pauketat JV, Ladak A, Manoli A (2025) Perceptions of sentient AI and other digital minds: evidence from the AI, morality, and sentience (AIMS) survey. In *CHI '25: Proc. 2025 CHI Conference on Human Factors in Computing Systems (Article 10)*. Association for Computing Machinery. https://doi.org/10.1145/3706598.3713329

Anthropic (2024) Introducing the next generation of Claude. The Anthropic Blog. Retrieved December 30, 2024, from https://www.anthropic.com/news/claude-3-family

Aristotle (1955) On sophistical refutations (E. S. Forster, Trans.). Harvard University Press. (Original work published ca. 350 B.C.)

Atkinson-Abutridy J (2025) Large language models: concepts, techniques and applications (1st ed.). CRC Press. https://doi.org/10.1201/9781003517245

Azevedo FAC, Carvalho LRB, Grinberg LT, Farfel JM, Ferretti REL, Leite REP, Filho WJ, Lent R, Herculano-Houzel S (2009) Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. J Comp Neurol 513(5):532–541. https://doi.org/10.1002/cne.21974

Balasubramanian V (2021) Brain power. Proc Natl Acad Sci USA 118(32):e2107022118. https://doi.org/10.1073/pnas.2107022118

Bender EM, Gebru T, McMillan-Major A, Shmitchell S (2021) On the dangers of stochastic parrots: can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp 610–623). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922

Berry DM (2023) The limits of computation: Joseph Weizenbaum and the ELIZA chatbot. Weizenbaum J Digital Soc 3(3). https://doi.org/10.34669/WI.WJDS/3.3.2

Bojić L, Stojković I, Jolić Marjanović Z (2024) Signs of consciousness in AI: can GPT-3 tell how smart it really is? Hum Soc Sci Commun 11:1631. https://doi.org/10.1057/s41599-024-04154-3

Butlin P, Lappas T (2025) Principles for responsible AI consciousness research. J Artif Intell Res 82:1673–1690. https://doi.org/10.1613/jair.1.17310

Butz MV (2021) Towards strong AI. KI - Künstliche Intell 35(1):91–101. https://doi.org/10.1007/s13218-021-00705-x

Chalmers D (2023) Could a large language model be conscious? Boston Review. Retrieved January 16, 2025, from https://www.bostonreview.net/articles/could-a-large-language-model-be-conscious

Chen S (2025) How much energy will AI really consume? The good, the bad and the unknown. Nature 639:22–24. https://doi.org/10.1038/d41586-025-00616-z

Cho A, Kim GC, Karpekov A, Helbling A, Wang ZJ, Lee S, Hoover B, Chau DH (2025) TRANSFORMER EXPLAINER: interactive learning of text-generative models. Proc AAAI Conf Artif Intell 39(28):29625–29627. https://doi.org/10.1609/aaai.v39i28.35347

Chollet F (2023) How I think about LLM prompt engineering. The Thoughts of François Chollet. Retrieved January 16, 2025, from https://fchollet.substack.com/p/how-i-think-about-llm-prompt-engineering

Cole S (2023) 'My AI is sexually harassing me': Replika users say the chatbot has gotten way too horny. VICE. Retrieved January 16, 2025, from https://www.vice.com/en/article/my-ai-is-sexually-harassing-me-replika-chatbot-nudes/

Colombatto C, Fleming SM (2024) Folk psychological attributions of consciousness to large language models. Neurosci Conscious (1):niae013. https://doi.org/10.1093/nc/niae013

Crespo R (2024) Does AI think? Sci Et Fides 12(2):37–47. https://doi.org/10.12775/SetF.2024.014

Cuthbertson A (2022) Artificial intelligence may already be 'slightly conscious', AI scientists warn. Retrieved May 14, 2025 from https://www.the-independent.com/tech/artificial-intelligence-conciousness-ai-deepmind-b2017393.html

de Vries A (2023) The growing energy footprint of artificial intelligence. Joule 7(10):2191–2194. https://doi.org/10.1016/j.joule.2023.09.004

Dennett DC (1997) How to do other things with words? R Inst Philos Suppl 42:219–235. https://doi.org/10.1017/S1358246100010262
Dennett DC (2005) 2005: What do you believe is true even though you cannot prove it? Edge.org. Retrieved August 4, 2025, from https://www.edge.org/response-detail/11902
Dennett DC (2018) From bacteria to Bach and back: The evolution of minds. W. W. Norton & Company
Devlin J, Chang MW, Lee K, Toutanova K (2019) BERT: pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) (pp. 4171–4186). Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423
Dickens M (2025) LLMs might already be conscious. Philosophical Multicore. Retrieved August 4, 2025, from https://mdickens.me/2025/07/05/LLMs_might_already_be_conscious
Dreksler N, Caviola L, Chalmers D, Allen C, Rand A, Lewis J, Waggoner P, Mays K, Sebo J (2025) Subjective experience in AI systems: what do AI researchers and the public believe?. Preprint at https://doi.org/10.48550/arXiv.2506.11945
Dreyfus HL (1992) What computers still can't do: a critique of artificial reason. MIT Press
Farthing GW (1992) The psychology of consciousness. Prentice-Hall
Fernández-Llorca D, Gómez E, Sánchez I, Mazzini G (2024) An interdisciplinary account of the terminological choices by EU policymakers ahead of the final agreement on the AI Act: AI system, general purpose AI system, foundation model, and generative AI. Artif Intell Law. https://doi.org/10.1007/s10506-024-09412-y
Floridi L (2025) AI and semantic pareidolia: when we see consciousness where there is none. SSRN. Retrieved August 4, 2025, from https://ssrn.com/abstract=5309682
Floridi L, Chiriatti M (2020) GPT-3: its nature, scope, limits, and consequences. Minds Machines 30(4):681–694. https://doi.org/10.1007/s11023-020-09548-1
Gamazaychikov B (2024) Redundant OpenAI has announced o3, which appears to focus on enhanced reasoning capabilities. [Post]. LinkedIn. Retrieved August 20, 2025, from https://www.linkedin.com/posts/bgamazay_openai-has-announced-o3-which-appears-to-activity-7276250095019335680-sVbW
Gams M, Kramar S (2024) Evaluating ChatGPT's consciousness and its capability to pass the Turing test: a comprehensive analysis. J Comput Commun 12(3):219–237. https://doi.org/10.4236/jcc.2024.123014
Goodfellow I, Bengio Y, Courville A (2016) Deep learning (1st ed.). The MIT Press
Grant A (2024) Autonomous electric vehicles will guzzle power instead of gas. Bloomberg. Retrieved May 14, 2025 from https://www.bloomberg.com/news/newsletters/2024-02-27/autonomous-electric-vehicles-will-guzzle-power-instead-of-gas
Grattafiori A, Dubey A, Jauhri A, Pandey A, Kadian A, Al-Dahle A, Letman A, Mathur A, Schelten A, Vaughan A, Yang A, Fan A, Goyal A, Hartshorn A, Yang A, Mitra A, Sravankumar A, Korenev A, Hinsvark A, … Ma Z (2024) The Llama 3 herd of models. Preprint at https://doi.org/10.48550/arXiv.2407.21783
Greco CM, Tagarelli A (2024) Bringing order into the realm of Transformer-based language models for artificial intelligence and law. Artif Intell Law 32(4):863–1010. https://doi.org/10.1007/s10506-023-09374-7
Grundhauser E (2016) Did a silent film about a train really cause audiences to stampede?. Atlas Obscura. Retrieved January 16, 2025, from https://www.atlasobscura.com/articles/did-a-silent-film-about-a-train-really-cause-audiences-to-stampede
Gubelmann R (2024) Large language models, agency, and why speech acts are beyond them (for now) – a Kantian-cum-pragmatist case. Philos Technol 37(1):32. https://doi.org/10.1007/s13347-024-00696-1
Hanson Robotics (2024) Sophia. Retrieved January 16, 2025, from https://www.hansonrobotics.com/sophia
Hermann I (2023) Artificial intelligence in fiction: between narratives and metaphors. AI Soc 38(1):319–329. https://doi.org/10.1007/s00146-021-01299-6
Hintze A (2023) ChatGPT believes it is conscious. Preprint at https://doi.org/10.48550/arXiv.2304.12898
Huckle J (2025) Code for "Easy problems that LLMs get wrong" paper. [GitHub Repository]. Retrieved August 20, 2025, from https://github.com/JamesHuckle/easy-problems-that-llms-get-wrong
Huckle J, Williams S (2025) Easy problems that LLMs get wrong. In Arai K (ed.), Advances in Information and Communication. FICC 2025. Lecture Notes in Networks and Systems, 1283 (pp. 313–332). Springer. https://doi.org/10.1007/978-3-031-84457-7_19
Janiesch C, Zschech P, Heinrich K (2021) Machine learning and deep learning. Electron Mark 31(3):685–695. https://doi.org/10.1007/s12525-021-00475-2
Jim the AI Whisperer (2025) The system prompt for ChatGPT-4.5 prevents it from telling us it has consciousness. Medium. Retrieved August 4, 2025, from https://generativeai.pub/chatgpt-system-prompt-reveals-ai-consciousness-contradictions-e8cd42ed037e
Jones CR, Bergen BK (2025) Large language models pass the Turing test. Preprint at https://doi.org/10.48550/arXiv.2503.23674
Jones CR, Rathi I, Taylor S, Bergen BK (2025) People cannot distinguish GPT-4 from a human in a Turing test. In Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (pp. 1615–1639). Association for Computing Machinery. https://doi.org/10.1145/3715275.3732108
Kalai AT, Vempala SS (2024) Calibrated language models must hallucinate. In STOC 2024: Proceedings of the 56th Annual ACM Symposium on Theory of Computing (pp. 160–171). Association for Computing Machinery. https://doi.org/10.1145/3618260.3649777
Kirkeby-Hinrup A (2024) Quantifying empirical support for theories of consciousness: a tentative methodological framework. Front Psychol 15:1341430. https://doi.org/10.3389/fpsyg.2024.1341430
Kiškis M (2023) Legal framework for the coexistence of humans and conscious AI. Front Artif Intell 6:1205465. https://doi.org/10.3389/frai.2023.1205465
Kuhn RL (2024) A landscape of consciousness: toward a taxonomy of explanations and implications. Prog Biophys Mol Biol 190:28–169. https://doi.org/10.1016/j.pbiomolbio.2023.12.003
LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444. https://doi.org/10.1038/nature14539
Lemoine B (2022) Scientific data and religious opinions. Medium. Retrieved January 17, 2025, from https://cajundiscordian.medium.com/scientific-data-and-religious-opinions-ff9b0938fc10
Long R, Sebo J, Butlin P, Finlinson K, Fish K, Harding J, Pfau J, Sims T, Birch J, Chalmers D (2024) Taking AI Welfare Seriously. Preprint at https://doi.org/10.48550/arXiv.2411.00986
Long R, Sebo J, Sims T (2025) Is there a tension between AI safety and AI welfare?. Philos Stud 182(7):2005–2033. https://doi.org/10.1007/s11098-025-02302-2
Luccioni AS, Viguier S, Ligozat AL (2023) Estimating the carbon footprint of BLOOM, a 176B parameter language model. J Machine Learn Res, 24:253. Retrieved May 16, 2025 from https://jmlr.org/papers/volume24/23-0069/23-0069.pdf
Luccioni S, Gamazaychikov B, Strubell E, Hooker S, Jernite Y, Wu CJ, Mitchell M (2025) AI Energy Score Leaderboard - February 2025. Hugging Face. Retrieved August 20, 2025, from https://huggingface.co/spaces/AIEnergyScore/Leaderboard
Łukowski P (2011) Paradoxes. Springer Netherlands. https://doi.org/10.1007/978-94-007-1476-2
Maleki N, Padmanabhan B, Dutta K (2024) AI hallucinations: a misnomer worth clarifying. In 2024 IEEE Conference on Artificial Intelligence (CAI) (pp. 127–132). IEEE. https://doi.org/10.1109/CAI59869.2024.00033
Manoli A, Pauketat JV, Anthis JR (2025) The AI double standard: humans judge all AIs for the actions of one. Proc ACM Hum-Comput Interact 9(2):CSCW185. https://doi.org/10.1145/3711083
McCarthy J (1979) Ascribing mental qualities to machines. In Ringle M (ed.), Philosophical Perspectives in Artificial Intelligence.Humanities Press
McDermott D (1976) Artificial intelligence meets natural stupidity. ACM SIGART Newsletter, (57), 4–9. https://doi.org/10.1145/1045339.1045340
Naveed H, Khan AU, Qiu S, Saqib M, Anwar S, Usman M, Akhtar N, Barnes N, Mian A (2025) A comprehensive overview of large language models. ACM Trans Intell Syst Technol 16(5):106. https://doi.org/10.1145/3744746
OpenAI (2025) Introducing GPT-5. Retrieved September 20, 2025, from https://openai.com/index/introducing-gpt-5
Padamsey Z, Rochefort NL (2023) Paying the brain's energy bill. Curr Opin Neurobiol 78:102668. https://doi.org/10.1016/j.conb.2022.102668
Parviainen J, Coeckelbergh M (2021) The political choreography of the Sophia robot: beyond robot rights and citizenship to political performances for the social robotics market. AI Soc 36(3):715–724. https://doi.org/10.1007/s00146-020-01104-w
Pentina I, Hancock T, Xie T (2023) Exploring relationship development with social chatbots: a mixed-method study of replika. Computers Hum Behav 140:107600. https://doi.org/10.1016/j.chb.2022.107600
Pichai S, Hassabis D (2023) Introducing Gemini: our largest and most capable AI model. The Keyword. Retrieved January 16, 2025, from https://blog.google/technology/ai/google-gemini-ai/
Placani A (2024) Anthropomorphism in AI: hype and fallacy. AI Ethics 4(3):691–698. https://doi.org/10.1007/s43681-024-00419-4
Porębski A (2023) Machine learning and law. In Brożek B, Kanevskaia O, Pałka P (eds.), Research Handbook on Law and Technology (pp. 450–467). Edward Elgar. https://doi.org/10.4337/9781803921327.00037
Porębski A (2024) Institutional black boxes pose an even greater risk than algorithmic ones in a legal context. In Mańdziuk J, Żychowski A, Małkiński M (eds.), Progress in Polish Artificial Intelligence Research 5 (pp. 562–570). Warsaw University of Technology Press. https://doi.org/10.2139/ssrn.4971723
Porębski A, Figura J (2025) Free LLMs hallucinate and rarely signal their limitations in solving legal problems. In Progress in Polish Artificial Intelligence Research 6. University of Silesia Press. https://doi.org/10.2139/ssrn.5188654
Prettyman A (2024) Artificial consciousness. Inquiry. https://doi.org/10.1080/0020174X.2024.2439989

Radford A, Narasimhan K, Salimans T, Sutskever I (2018) Improving language understanding by generative pre-training. OpenAI. Retrieved May 4, 2025, from https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf
Ragni M, Kola I, Johnson-Laird PN (2017) The Wason selection task: a meta-analysis. In Proceedings of the 39th Annual Meeting of the Cognitive Science Society (pp. 980–985). Cognitive Science Society
Rajashekara K (2024) Energy impacts of autonomous vehicles – present and the future. iEnergy 3(2):73–74. https://doi.org/10.23919/IEN.2024.0012
Ray WJ (2024) Fundamentals of brain and behavior: an introduction to human neuroscience (1st ed.). Routledge. https://doi.org/10.4324/9781003266426
Salles A, Evers K, Farisco M (2020) Anthropomorphism in AI. AJOB Neurosci 11(2):88–95. https://doi.org/10.1080/21507740.2020.1740350
Searle JR (1980) Minds, brains, and programs. Behav Brain Sci 3(3):417–424. https://doi.org/10.1017/S0140525X00005756
Searle JR (2009) What is language: some preliminary remarks. Etica Politica 11(1):173–202
Shwartz S (2024) Don't be fooled by ChatGPT: it's not as smart as you think. AI Perspectives. Retrieved May 14, 2025 from https://steveshwartz.substack.com/p/reasoning-2
Shwartz S (2020) AI meets natural stupidity revisited. AI Perspectives. Retrieved May 14, 2025 from https://www.aiperspectives.com/ai-meets-natural-stupidity/
Skipper JI (2022) A voice without a mouth no more: the neurobiology of language and consciousness. Neurosci Biobehav Rev 140:104772. https://doi.org/10.1016/j.neubiorev.2022.104772
Sloane M, Danks D, Moss E (2024) Tackling AI hyping. AI Ethics 4(3):669–677. https://doi.org/10.1007/s43681-024-00481-y
Sotala K, Yampolskiy RV (2015) Responses to catastrophic AGI risk: a survey. Phys Scr 90(1):018001. https://doi.org/10.1088/0031-8949/90/1/018001
Tiku N (2022) The Google engineer who thinks the company's AI has come to life. The Washington Post. Retrieved January 17, 2025, from https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/
Touvron H, Lavril T, Izacard G, Martinet X, Lachaux MA, Lacroix T, Rozière B, Goyal N, Hambro E, Azhar F, Rodriguez A, Joulin A, Grave E, Lample G (2023) LLaMA: open and efficient foundation language models. Preprint at https://doi.org/10.48550/arXiv.2302.13971
Turing A (1950) Computing machinery and intelligence. Mind LIX(236):433–460. https://doi.org/10.1093/mind/LIX.236.433
Tyler CW (2020) Ten testable properties of consciousness. Front Psychol 11:1144. https://doi.org/10.3389/fpsyg.2020.01144
Van Eemeren FH, Garssen B, Krabbe ECW, Snoeck Henkemans AF, Verheij B, Wagemans JHM (2014) Handbook of argumentation theory. Springer Netherlands. https://doi.org/10.1007/978-90-481-9473-5
van Es K, Nguyen D (2025) "Your friendly AI assistant": the anthropomorphic self-representations of ChatGPT and its implications for imagining AI. AI Soc 40(5):3591–3603. https://doi.org/10.1007/s00146-024-02108-6
Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. In Proceedings of the 31st International Conference of Neural Information Processing Systems (pp. 6000–6010). Curran Associates Inc
Vecht JJ (2023) Open texture clarified. Inquiry 66(6):1120–1140. https://doi.org/10.1080/0020174X.2020.1787222
Volokh E (2023) Large libel models? Liability for AI output. J Free Speech Law 3(2):489–557
Waismann F (1945) Verifiability. Proc Aristotelian Soc Suppl Volumes 19:119–150
Wang F, Zhang Z, Zhang X, Wu Z, Mo T, Lu Q, Wang W, Li R, Xu J, Tang X, He Q, Ma Y, Huang M, Wang S (2025) A comprehensive survey of small language models in the era of large language models: techniques, enhancements, applications, collaboration with LLMs, and trustworthiness. ACM Trans Intell Syst Technol. https://doi.org/10.1145/3768165
Wasiewicz A, Szuba T (1990) The idea of consciousness for intelligent machines as the frame for the future software construction. In Proceedings of the IMACS International Symposium on Mathematical and Intelligent Models in System Simulation (MIM-S^2), Brussels
Wason PC (1968) Reasoning about a rule. Q J Exp Psychol 20(3):273–281. https://doi.org/10.1080/14640746808400161
Weiser D (2023) Here's what happens when your lawyer uses ChatGPT. The New York Times. Retrieved January 16, 2025, from https://www.nytimes.com/2023/05/27/nyregion/avianca-airline-lawsuit-chatgpt.html
Weizenbaum J (1966) ELIZA—a computer program for the study of natural language communication between man and machine. Commun Asso Comput Machin 9(1):36–45. https://doi.org/10.1145/365153.365168
Wittgenstein L (1922) Tractatus Logico-Philosophicus (C. K. Ogden, Trans.). Kegan Paul, Trench, Trubner & Co., Ltd
Willison S (2025) Highlights from the Claude 4 system prompt. Simon Willison's Weblog. Retrieved August 4, 2025, from https://simonwillison.net/2025/May/25/claude-4-system-prompt
Xiang C (2023) Man dies by suicide after talking with AI chatbot, widow says. VICE. Retrieved January 16, 2025, from https://www.vice.com/en/article/man-dies-by-suicide-after-talking-with-ai-chatbot-widow-says/
Yao D (2024) Air Canada held responsible for chatbot's hallucinations. AI Business. Retrieved January 16, 2025, from https://aibusiness.com/nlp/air-canada-held-responsible-for-chatbot-s-hallucinations
Young GB, Pigott SE (1999) Neurobiological basis of consciousness. Arch Neurol 56(2):153–157. https://doi.org/10.1001/archneur.56.2.153
Zhou L, Schellaert W, Martínez-Plumed F, Moros-Daval Y, Ferri C, Hernández-Orallo J (2024) Larger and more instructable language models become less reliable. Nature 634(8032):61–68. https://doi.org/10.1038/s41586-024-07930-y

Acknowledgements

This research was funded by the National Science Centre, Poland, and is the result of the research project no. 2022/45/N/HS5/00871. The open access of the publication has been supported by a grant from the Faculty of Law and Administration under the Strategic Programme Excellence Initiative at Jagiellonian University. The authors would like to thank Andrzej Uhl for his careful reading, help with language proofreading, and valuable comments that made this work better. Andrzej Porębski would like to thank the Kazimierz Twardowski Philosophical Society of Lviv for the opportunity to discuss a preliminary version of the arguments put forward in the study during the Round Table "Mind-Body Problem: History – the Lviv-Warsaw School – Contemporaneity". The work of Andrzej Porębski was supported by the Foundation for Polish Science (FNP).

Author contributions

The original conception and design of the paper were developed by A.P., while J.F. suggested its significant extensions. The first draft of the manuscript was written primarily by A.P., except for the "Potential limitations" section, which was written primarily by J.F. Both authors have revised and edited the manuscript. Both authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Ethical approval

The study did not involve human participants or animal subjects.

Informed consent

The study did not involve human participants.

Additional information

Correspondence and requests for materials should be addressed to Andrzej Porębski.