Phenomenology and the Cognitive Sciences (2024) 23:1009–1015
https://doi.org/10.1007/s11097-024-10040-9

Phenomenology and artificial intelligence: introductory notes

Steven S. Gouveia[1,2] · **Carlos Morujão**[3]

Accepted: 22 October 2024 / Published online: 7 November 2024

Keywords Phenomenology · Artificial intelligence · Consciousness · ChatGPT

The new developments in the field of artificial intelligence raise many different philosophical questions that can profoundly change our thinking about various mental concepts, such as what it means to be a conscious subject, how perception works, how humans interact with their environment via their bodies, and so on (cf. Andler, 2006; Froese & Ziemke, 2009).

Interestingly, the recent 2024 Nobel Prize in Physics Award was attributed John J. Hopfield and Geoffrey E. Hinton for their work on artificial neural networks, which are considered the foundations of the many machine learning techniques that are being used today. The interesting take is that Hopfield's motivation to develop his method was to actually describe and understand how the human brain works (specifically, how memories are stored and retrieved), using several mental concepts and mental analogies to describe how the algorithm worked (Hopfield, 1982).

Oddly enough, these new developments – from machine learning tools to deep learning algorithms and generative AI models – have not been the primary focus of philosophical deliberation (with some exceptions, such as Buccella & Springle, 2022), even though more traditional AI has certainly been considered in the past by several theorists such as Dreyfus (1972, 1992), who contended that computers lack embodied capabilities related to our lived experience of the world, Ihde (1990) who claimed that technology is not neutral but shapes our subjective experience of the world, or Gallagher (2005), who argues that the embodied cognition paradigm,

Special Issue: "Phenomenology and Artificial Intelligence: Bridges and New Paths".

✉ Steven S. Gouveia
stevensequeira92@hotmail.com

[1] Research Fellow of the Mind, Language and Action Group, Institute of Philosophy, Faculty of Arts and Humanities, University of Porto, Via Panorâmica s/n 4150-564 Porto, Portugal

[2] Faculty of Medicine, Andres Bello University, Vinã Del Mar, Chile

[3] Faculty of Philosophy, Portuguese Catholic University, Lisbon, Portugal

based on the relevance of how our bodily experiences shape our subjective experience, can provide an appropriate framework for developing intelligent systems.

As a philosophical approach that focuses on the study of the universal structures of subjective experiences and the way people perceive and interpret the world around them, phenomenology can provide valuable insights when applied to AI in general (cf. Beavers, 2002; Mensch, 1991; Preston, 1993). For example, we can better understand how people interact with intelligent systems and how these systems can be made more user-friendly and effective, but also by using AI to model and simulate different scenarios that allow researchers to examine hypothetical situations and understand how people might respond to different stimuli (cf. Coeckelbergh, 2011) or how simulated embodied agents can improve their performance (cf. Xia et al., 2018).

This can be particularly useful when studying complex and dynamic phenomena such as emotions, social interactions, cultural practices, or mental illnesses, that are difficult for researchers to analyze using traditional research methods (cf. Susser, 2013; Tae-hee, 2019; Zhang, 2021). Other relevant issues are related to topics such as how the Internet can be considered a new kind of cognitive ecology (cf. Smart et al., 2017), the impact that virtual reality and the metaverse can have on the 4E's approach to the mind (cf. Smart, 2022), how the research in Robotics can be improved by making use of the 4E Cognition framework (cf. Hoffmann & Pfeifer, 2018), or the philosophical debates raised by the existence of large language models such as ChatGPT by OpenAI (e.g. is there a relationship between linguistic behavior/performance and subjectivity?) (cf. Floridi, 2023).

Phenomenology can also provide a perspective on consciousness that highlights its embodied, dynamic, and situated nature. According to this view, consciousness is not a stationary, isolated, or fixed phenomenon, but is constantly in flux, shaping and being shaped by the world around it. This perspective can inspire AI researchers to develop intelligent systems that are more flexible and adaptive, responding to their environment rather than simply following pre-programmed rules or algorithms (cf. Linson et al., 2018; Turner, 2020; Ramstead et al., 2022).

With this Special Issue, we tried to find out whether a phenomenological approach to consciousness and other mental features could or improve the "mental" capacities of artificial systems to the point of developing some sense of subjectivity and first-person perspective. Would a more advanced ChatGPT be conscious as humans and other animals are? Or would it develop a different kind of subjective experience that, even though it is different than the one we humans possess, should still count as conscious?

This is also relevant the other way around: can we make use of specific artificial intelligence technologies to ensure the continuity of the stream of consciousness in scenarios such as the mind-uploading hypothesis, where each individual biological neuron is replaced by artificial neurons, or in cases related with disorders of consciousness (e.g., coma or vegetative state) where brain-machine interfaces can be used to improve the patient's health condition and his ability to regain a first-person perspective?

After this brief introduction, we are happy to present the eight papers selected for this Special Issue. Even if the topics are different from each other, they all try

to show how the combination of phenomenology and AI can provide new ways of understanding subjective experience in its broadest sense, and how AI practice and development can be improved and understood via an inspired phenomenological approach broadly considered.

Blake Lemoine, a former Google software engineer, recently gained attention for his claim that the company's AI language model, LaMDA, had developed sentience. His assertion, which led to his dismissal after he shared conversations with LaMDA online, raised a key question: what does it mean for an AI to be sentient? In one conversation, Lemoine asked LaMDA to describe its consciousness, to which the chatbot responded that it was aware of its existence, eager to learn about the world, and capable of experiencing emotions such as happiness and sadness. Furthermore, LaMDA claimed it could understand and use natural language similarly to humans, suggesting that this ability distinguished it from other animals.

The first paper of this special issue, by James Mensch and titled "Embodiment and Intelligence, a Levinasian Perspective" critically analyzes Lemoine's claims about LaMDA's sentience and linguistic intelligence. By exploring how LaMDA's language use, while sophisticated, differs from human understanding, the paper explores the deeper question of what constitutes true sentience. To unpack this, Mensch first presents arguments supporting LaMDA's capacity for language-based intelligence and then highlights how this is fundamentally different from how human consciousness operates. The paper draws on Emmanuel Levinas' phenomenology of embodiment to argue that, despite LaMDA's advanced capabilities – which are indeed notable –, it lacks the physical and lived experience that is central to human sentience.

Following the same focus on language, Susan J. Stuart presents in her paper "Why language clouds our ascription of understanding, intention and consciousness" another argument focused on how language's grammatical manipulation and production often mislead observers into equating fluency with intelligence, understanding, and intention, whether those intentions are conscious or not. This misjudgment concerning many of the large language models (LLMs) such as ChatGPT, LLaMA, Googl Bard or GPT-4 seems to persist from specifically two categorial mistakes: one epistemological, addressing the reasons behind the misinterpretation of mental life in LLMs, and the other ontological, asserting the impossibility of LLMs accomplishing consciousness. Like the previous paper, Stuart will make a phenomenological claim: that the relevance of the coordination and reciprocal adaptation of the body and the world in both natural and cultural language – grounded in the concept of *enkinaesthesia* – shows why these LLMs cannot develop any meaningful understanding or intersubjectivity.

Making use of the same contextual insight, Veronica Cibotaru will present in her paper "For a contextualist and content-related understanding of the difference between human and artificial intelligence" an examination of several theoretical perspectives on the distinction between human and artificial intelligence. Against the behaviorist approach (intelligence is based on observable behaviors), the representational approach (intelligence involves internal representations of the world) and the holistic approach (intelligence emerges from the interaction of various cognitive processes, potentially including embodied AI), Cibotaru

will argue that intelligence is not essentialist – like the previous approaches – but contextualist and content-related. This view emphasizes the phenomenological differences rooted in the life-world of human existence and intelligence, which allows Cibotaru to argue that human and artificial intelligence could be distinct even if one could prove that they are eidetically identical.

Next, Anthony Beavers and Eli B. McGraw consider in their paper "Clues and caveats concerning artificial consciousness from a phenomenological perspective" the implications of LLMs and GPT-equipped robotics for understanding semantic meaning and the possibility of artificially conscious machines. The authors argue that consciousness, in a human-like sense, is rooted in phenomenology, that is, a lived experience combining sensory data (qualia) and a 4E enactivist framework, which gives rise to intentional behavior. Without such a phenomenology, a robot is "semantically empty" and cannot possess consciousness similar to that of humans. Beavers and McGraw also highlight some concerns about the potential dangers of creating and deploying pseudo-conscious entities: there is a real risk of machines behaving unpredictably due to the lack of human-like phenomenology, which constrains behavior. Additionally, robots lack moral agency and genuine intentionality, which raises ethical concerns about accountability and the social impact of machines that appear conscious or moral but are not.

Focusing on the same topic of consciousness, Georg Northoff and Steven S. Gouveia argue in their "Does artificial intelligence exhibit basic fundamental subjectivity? A neurophilosophical argument" that, if we look at how humans and their brain function (specifically, the timescales between the brain, the body and the world), we can conclude that current AI models such as ChatGPT lack the relevant neuroecological layer between the synchronization of the (embodied) brain's processes with the world, which does not allow them to develop a sense of basic subjectivity. Moreover, this neuroecological layer is phenomenologically associated with a "point of view" which is also missing in AI. However, Northoff and Gouveia do not necessarily exclude that future AI may exhibit subjective capacities: for that, we will have to be able to develop artificial systems that can recreate the relevant timescales that give rise to human consciousness in humans. For now, this seems an impossibility and further research needs to be done.

The next paper, by Camilo Miguel Signorelli & Joaquin Diaz Boils, and titled "Multilayer networks as embodied consciousness interactions. A formal model approach" introduces an algebraic approach to understanding conscious experience by modeling brain and body interactions through multigraph networks: these networks can merge via an associative binary operation, representing biological composition. Interestingly, Signorelli and Boil also present a mathematical framework for analyzing the transition between conscious and non-conscious activity, revealing core structures for conscious experience, its dynamic content, and the causal constraints of conscious interactions. By extending concepts from enactive and embodied cognition using mathematical methods, focusing on experiential layers and their interactions, the paper predicts structural and causal elements tied to consciousness and outlines how these principles can apply to artificial models, suggesting that mathematical and phenomenological approaches can deepen our

understanding of subjective experience, while remaining uncertain if AI models can become truly conscious.

Considering one of the core concepts of classical phenomenology – the Husserlian concept of "transcendental consciousness" – Zbigniew Orbik presents the paper "Husserl's concept of transcendental consciousness and the problem of AI consciousness" by exploring whether Husserl's concept of "pure transcendental consciousness" can serve as a model for AI consciousness. Transcendental consciousness, according to Husserl, is an outcome of the practice of the phenomenological method that looks at conscious activity as embedded in experience. The paper questions whether AI could possess the qualities Husserl attributes to this form of consciousness. Orbik argues that AI, which is a product of human design, is restricted to the realm of phenomena, whereas human intelligence operates on an ontological level. Because of this, AI does not align with the transcendental concept of consciousness as Husserl envisioned. The paper also emphasizes Husserl's philosophical view that consciousness holds a unique position in the world's cognitive structure and critiques naturalism, particularly the reductionist approaches of Husserl's time, and suggests that while modern naturalism has evolved, it still doesn't adequately account for the nature of consciousness. Husserl's framework, which challenges the naturalization of consciousness, could inform AI research, especially by analyzing AI behavior without presupposing metaphysical theories about machine consciousness. Ultimately, Orbik proposes that phenomenology may help clarify to what extent AI can replicate aspects of human consciousness, while leaving open – again –, the question of whether artificial consciousness can ever exist.

To close the Special Issue, Daniil Koloskov presents the paper "AI-informed acting: an Arendtian perspective" by exploring AI's impact from two angles: first, it argues that AI cannot directly interact with human action (since AI operates based on efficiency, it cannot guide or execute human actions, which are driven by existential motivations rather than by an instrumental pursuit of efficiency); second, it investigates AI's potential indirect impact, suggesting that neural networks could facilitate the circulation of actions by organizing interactions beyond human cognitive limits, thus serving as a catalyst for collective action. Based on Hanna Arendt's view on action as a form of praxis (an activity whose purpose is intrinsic to its performance), the paper argues that while AI cannot take over the act of "doing" from humans or threaten our capacity for genuine action, it may enhance the communication and reach of human actions (for example, AI's role in organizing information and reducing cognitive challenges may amplify the scope and unpredictability of action). By increasing the communicability of actions and enabling broader networks of interaction, AI can expand the landscape in which human action takes place, facilitating the spontaneous and unpredictable elements of action that are essential in the public sphere. In this way, Koloskov end up arguing that AI doesn't diminish the spontaneity of action but may rather help expand the web of human interactions where actions gain their full significance.

With this compilation of perspectives, we aim to contribute to the ongoing dialogue surrounding the interplay between phenomenology and artificial intelligence, highlighting both the bridges and new paths that are emerging in this field.

By examining these intersections, we hope to inspire further research not only in the traditional domains of phenomenology and philosophy of mind but also in interdisciplinary fields that include empirical studies in AI, cognitive science, and human–computer interaction. The insights from these reflections could lead to a more nuanced understanding of how AI might be integrated into human experience, without losing sight of the unique features of consciousness that are central to phenomenological analysis.

We would like to extend our gratitude to all contributors of the Special Issue – but also to the two Editors-in-Chief, Shaun Gallagher and Dan Zahavi –, from whom this collaboration emerged, whose insights and efforts have enriched this conversation on phenomenology and AI. We hope that these collective works will serve as a springboard for future inquiry into how these two fields can continue to inform and shape one another, ultimately paving new paths for understanding the complexities of human and artificial intelligence.

Acknowledgements SG is funded by CEEC Individual Project by FCT 2022.02527.CEECIND, at the Mind, Language and Action Group, Institute of Philosophy, University of Porto (Faculdade de Letras, Via Panorâmica s/n, P-4150-564 Porto, Portugal). CM is funded by FCT UID/00683/2020, Centro de Estudos Filosóficos e Humanísticos/Centre for Philosophical and Humanistic Studies, Universidade Católica Portuguesa/Portuguese Catholic University, Praça da Faculdade, Braga 4710-297, Portugal.

Author's contribution SG first author, CM second author.

Funding Open access funding provided by FCT|FCCN (b-on). SG is funded by CEEC Individual Project by FCT 2022.02527.CEECIND, and the internal project IF.Proj.06.2024 at the Mind, Language and Action Group, Institute of Philosophy, University of Porto (Faculdade de Letras, Via Panorâmica s/n, 4150–564 Porto, Portugal). CM is funded by UID/00683/2020, Centro de Estudos Filosóficos e Humanísticos/ Centre for Philosophical and Humanistic Studies, Universidade Católica Portuguesa/Portuguese Catholic University, Praça da Faculdade, Braga 4710–297, Portugal.

Data availability Not applicable.

Declarations

Ethical approval Not applicable.

Statement regarding research involving human participants and/or animals Not applicable.

Informed consent Not applicable.

Competing interests Not applicable.

References

Andler, D. (2006). Phenomenology in artificial intelligence and cognitive science. In H. Dreyfus & M. Wrathall (Eds.), *The blackwell companion to phenomenology and existentialism* (pp. 377–393). Blackwell.

Beavers, A. (2002). Phenomenology and artificial intelligence. *Metaphilosophy, 33*(1–2), 70–82.

Buccella, A., & Springle, A. (2022). Phenomenology: What's AI got to do with it*?*. *Phenomenology and the Cognitive Sciences*. https://doi.org/10.1007/s11097-022-09833-7

Coeckelbergh, M. (2011). Humans, animals, and robots: A phenomenological approach to human-robot relations. *International Journal of Social Robotics, 3*, 197–204.

Dreyfus, H. (1972). *What computers can't do*. MIT Press.

Dreyfus, H. (1992). *What computers still can't do*. The MIT Press.

Floridi, L. (2023) AI as agency without intelligence: On ChatGPT, large language models, and other generative models. *Philosophy & Technology*, *36*(15). https://doi.org/10.1007/s13347-023-00621-y

Froese, T., & Ziemke, T. (2009). Enactive artificial intelligence: Investigating the systemic organization of life and mind. *Artificial Intelligence, 173*, 466–500.

Gallagher, S. (2005). *How the body shapes the mind*. Oxford University Press.

Hoffmann, M., & Pfeifer, R. (2018). Robots as powerful allies for the study of embodied cognition from the bottom up. In A. Newen, L. De Bruin, & S. Gallagher (Eds.), *The Oxford handbook of 4E cognition* (pp. 841–862). Oxford Library of Psychology.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences, 79*(8), 2554–2558. https://doi.org/10.1073/pnas.79.8.2554

Ihde, D. (1990). *Technology and the lifeworld: From garden to earth*. Indiana University Press.

Linson, A., Clark, A., Ramamoorthy, S., & Friston, K. (2018). The active inference approach to ecological perception: General information dynamics for natural and artificial embodied cognition. *Frontiers in Robotics and AI, 5*(21), 1–22.

Mensch, J. (1991). Phenomenology and artificial intelligence: Husserl learns Chinese. *Husserl Studies, 8*(2), 107–127.

Preston, B. (1993). Heidegger and artificial intelligence. *Philosophy and Phenomenological Research, 53*(1), 43–69.

Ramstead, M., Seth, A., Hesp, C., Sandved-Smith, L., Mago, J., Lifshitz, M., Pagnoni, G., Smith, R., Dumas, G., Lutz, A., Friston, K., & Constant, A. (2022). From generative models to generative passages: A computational approach to (Neuro) phenomenology. *Review of Philosophy and Psychology, 13*(4), 829–857.

Smart, P., Heersmink, R., & Clowes, R. (2017) The cognitive ecology of the internet. In S. Cowley, & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: Computation, interactivity and human artifice* (2nd ed., pp. 251-282). Springer.

Smart, P. (2022). Minds in the metaverse: Extended cognition meets mixed reality. *Philosophy and Technology, 35*(4), 1–29.

Susser, D. (2013). Artificial intelligence and the body: Dreyfus, bickhard, and the future of AI. In V. C. Müller (Ed.), *Philosophy and theory of artificial intelligence* (pp. 277–287). Springer.

Tae-hee, K. (2019). Situation and environment of embodied artificial intelligence – A phenomenological reflection. *Phenomenology and Contemporary Philosophy, 83*, 1–38.

Turner, C. (2020) The cognitive phenomenology argument for disembodied AI consciousness. In Steven S. Gouveia (ed.), The age of artificial intelligence: An exploration. Vernon Press. pp. 111–132.

Xia, F., Zamir, A., He, Z., Sax, A., Malik, J., & Savarese, S. (2018). Gibson Env: Real-world perception for embodied agents. *Proceedings of the IEEE computer society conference on computer vision and pattern recognition* (pp. 9068–9079). https://doi.org/10.1109/CVPR.2018.00945

Zhang, X. (2021). Human nature, time-consciousness, and the new frontiers of artificial intelligence – An inquiry from the perspective of phenomenology and the eastern school of mind. In B. Song (Ed.), *Intelligence and wisdom: Artificial intelligence meets Chinese philosophers* (pp. 131–150). Springer Singapore.