

The Cartesian Paradox: Thinking AI and Non-Thinking Humans?

Giuseppe Pernagallo*

Abstract— The rise of artificial intelligence has revived a fundamentally Cartesian question: what conditions must hold for something to count as genuinely thinking? This paper examines the relationship between cognitive delegation and the Cartesian *cogito*, arguing that contemporary AI systems challenge not the logical validity of the *cogito* itself, but its cultural and epistemic centrality. Drawing on a methodological distinction between functional, Cartesian, and agential forms of thinking, the paper develops what is termed the “Cartesian paradox”: while humans increasingly delegate judgment and reasoning to AI without necessarily ceasing to exist, machines increasingly reproduce the external markers of intelligence without thereby becoming *res cogitantes*. The analysis shows that the *cogito* guarantees certainty of existence through first-person self-presence, but does not imply that the suspension of reflective thought entails ontological annihilation. At the same time, the paper argues that AI systems, despite their growing functional sophistication, remain forms of *res extensa* insofar as they lack subjective self-awareness and agential self-relation. The conclusion is that the real danger posed by AI is not the replacement of human beings by conscious machines, but the progressive reduction of thinking to operational performance, externalized computation, and algorithmic delegation.

I. INTRODUCTION

Do you exist? You are here, reading this paper; the question therefore seems trivial, perhaps even annoying. Yet it is not so easily dismissed. Descartes took this question with utmost seriousness because he was searching for an indubitable foundation for his new epistemology. What if my senses are deceiving me? What if what I am experiencing right now is merely a dream? In such cases, we might still be tempted to say that some certainties remain, especially those sciences that do not depend on the physical existence of objects. But what if an evil genius were causing us to believe that our calculations or geometrical demonstrations were correct when, in fact, they were not?

The French philosopher’s project was a radical one. Descartes deliberately adopted the standpoint of the extreme sceptic because he believed that metaphysics had to be treated as a genuine science. Such a science could not be established as long as it remained trapped in endless disputes and inherited assumptions. The only way to purify metaphysics was therefore to develop a new method, first outlined in the *Regulae ad directionem ingenii* and later refined in the *Discours de la méthode*. For this project to succeed, philosophy required an unshakable foundation. The answer is well known: if I doubt, then I think; and if I think, then I am immediately present to myself as a thinking being (*res cogitans*, in Cartesian terms). Therefore, I necessarily exist: *ego cogito, ergo sum*.

The Cartesian approach becomes particularly valuable

when applied to artificial intelligence (AI). Contemporary AI systems make it possible to transfer to machines activities that were once integral parts of the thinking process, such as generating ideas, writing, calculating, remembering, or correcting ([1]–[3]). This tendency can be pushed to its extreme: we may imagine a subject who gradually stops exercising independent judgment and delegates cognitive activity almost entirely to AI. Here the Cartesian paradox arises: does a human being who completely delegates thought to a machine cease to be *res cogitans*? And if machines successfully reproduce the external signs of thinking, does this make them *res cogitantes*?

The paradox emerges because the very act of delegation seems already to presuppose thought. In order to entrust one’s reasoning to a machine, one must first recognize an end, select a means, and authorize the process itself. Delegation therefore appears to confirm, rather than negate, the activity of the thinking subject. So, the paradox becomes more radical once we imagine a form of total and permanent delegation in which the individual no longer exercises independent judgment and merely ratifies the outputs produced by AI. In such a scenario, the question is no longer whether machines can think, but whether the human subject can still meaningfully be described as *res cogitans*.

The paradox extends to AI itself. If a machine can successfully reproduce the external manifestations of reasoning (argumentation, problem solving, linguistic interaction, and even forms of apparent self-reference) then it begins to resemble what Descartes attributed exclusively to the thinking subject. Yet, from a Cartesian perspective, genuine thinking is not reducible to the production of correct outputs ([4]–[6]). Thinking requires first-person self-presence: the subject must not only generate thoughts, but also be aware of itself as thinking. AI systems may imitate the structure and language of rationality while lacking the reflective self-relation that grounds the *cogito*. The paradox, therefore, is double-sided: humans risk becoming less reflective by delegating thought to machines, while machines increasingly appear reflective without genuinely becoming thinking subjects.

The contribution of this paper is therefore twofold. First, by employing the Cartesian criterion to distinguish what counts as *res cogitans* from what does not, I argue that AI, despite reproducing many of the external markers of genuine thinking, remains a highly sophisticated instance of *res extensa*. In fact, AI appears increasingly mind-like only because the meaning of reason has already been narrowed to calculation, performance, and functional output. Once intelligence is understood from the outside, as the capacity to generate coherent responses or solve problems efficiently, artificial systems seem to approach human thought. This apparent proximity depends on a prior reduction of thinking itself: what AI resembles is not the Cartesian mind as self-present subjectivity, but a post-Cartesian model of reason

*Department of Economics and Statistics “Cognetti de Martiis”, University of Turin, Italy, Lungo Dora Siena, 100A, 10153 Torino TO, giuseppe.pernagallo@unito.it

detached from inner experience, judgment, and first-person awareness. This conclusion is broadly consistent with previous studies ([7]–[14]), yet the present argument differs from them in an important respect: it provides a clear and unambiguous criterion, namely the Cartesian one, for distinguishing genuine thinking subjects from systems that merely simulate the appearance of thought.

Second, this paper argues that even in the extreme scenario of total cognitive delegation to machines, the human subject cannot simply be ontologically annihilated. The reason is that, while the *cogito* guarantees the certainty of existence, the absence of active thinking does not automatically imply non-existence. In other words, consistently with previous work ([15]), the total suspension of reflective thinking deprives the *res cogitans* of certainty about itself, but it does not entail the ontological disappearance of the individual.

The paper proceeds through three interconnected moments. It begins by reconsidering the Cartesian *cogito* not as a psychological statement, but as an epistemic event through which the subject attains indubitable certainty of itself. The analysis then shifts to contemporary forms of cognitive delegation, examining how AI increasingly occupies functions once associated with judgment, deliberation, and intellectual autonomy. Against this background, the paper revisits the Cartesian paradox introduced above: the real danger posed by AI is not that machines may literally become thinking subjects, but that human beings may progressively abandon the reflective activity through which they recognize themselves as such. The central claim, therefore, is not that AI refutes the *cogito*, but that it contributes to a historical transformation in the meaning and cultural experience of thinking itself.

II. THE CARTESIAN METHODOLOGY

Descartes’ philosophy in relation to AI has been only marginally discussed in previous work ([15]–[17]). The recent development of large language models has revived a profoundly Cartesian problem: whether the external manifestation of intelligence is sufficient to justify the attribution of mind. Systems such as ChatGPT or LaMDA can generate context-sensitive, semantically coherent, and apparently reflective responses across an open range of domains. In conversational contexts, they may even present themselves as possessing emotions, fears, intentions, or self-awareness. As a result, AI increasingly creates the impression not merely of computation, but of understanding.

From a Cartesian perspective, however, this conclusion remains problematic. The production of sophisticated linguistic outputs does not necessarily imply the existence of inner consciousness or genuine self-relation. AI may reproduce the external signs of thinking while lacking the first-person awareness that, for Descartes, defines the *res cogitans*. The issue is therefore not whether machines can imitate rational behaviour, but whether such imitation is sufficient to establish the presence of a thinking subject.

This difficulty reflects a broader philosophical tension in contemporary debates on AI. Some approaches implicitly treat linguistic competence and adaptive behaviour as evidence of consciousness, thereby collapsing the distinction

between functional performance and subjective awareness. Critics of this “Cartesian” inheritance argue that the obsession with machine consciousness risks obscuring more concrete political and ethical questions concerning governance, responsibility, and power ([17]). The central issue, on this view, is not whether AI is genuinely sentient, but how societies should regulate systems that increasingly behave as if they were.

From a genuinely Cartesian standpoint, the impossibility of establishing consciousness through behaviour alone is not a weakness of Cartesianism but one of its central insights. The *cogito* provides certainty only in the first-person case: I know that I think because I encounter myself directly in the act of thinking. This certainty cannot be generalized straightforwardly to external entities, whether human or artificial. Descartes himself acknowledged this epistemic asymmetry when he remarked that what we perceive in others may be nothing more than “coats and hats” concealing automata ([18], [19]). In this sense, AI does not refute Cartesian philosophy; rather, it radicalizes its deepest problem: the gap between the appearance of intelligence and the existence of a thinking self.

Long before the emergence of computational systems, Descartes had already separated mechanical operation from genuine thought. In the *Discours de la méthode* ([20]), he imagines machines capable of imitating human behaviour and language, denying that such imitation is sufficient for intelligence in the strict sense. The reason is that machines operate through fixed mechanisms, whereas human reason possesses universality: it can adapt flexibly to indefinitely many situations rather than merely executing specialized tasks. In this respect, it endorses the idea that intellect is fundamentally a matter of illumination rather than computation ([21], [22]).

Descartes anticipated a problem that remains central to AI research today. Even highly sophisticated systems struggle with genuinely open-ended understanding because computational procedures, however complex, remain rule-governed operations. Contemporary AI may therefore challenge the intuitive force of Cartesian distinctions without necessarily overcoming them. Large language models can produce coherent arguments, context-sensitive responses, and forms of apparent self-reflection, thereby creating the impression that machines genuinely think. From a Cartesian perspective, behavioural sophistication alone cannot establish the existence of a thinking subject.

The point is crucial because the Cartesian distinction between *res cogitans* and *res extensa* is not merely a distinction between “smart” and “non-smart” entities. It is a distinction between self-aware subjectivity and mechanical process. For Descartes, thinking is inseparable from first-person self-presence: doubting, affirming, willing, and judging are not simply functions performed by a system, but acts through which the subject encounters itself directly as thinking. AI systems may reproduce the external grammar of these acts while lacking the inner self-relation that gives them subjective reality.

Cartesian philosophy organizes reality around a fundamental distinction between *res cogitans* and *res extensa*. The first designates the thinking subject. The second designates the realm of extension, that is, bodies governed by spatiality, divisibility, and mechanical causation. Although

forms of dualism existed long before Descartes (see [23]–[25]), his originality lies in having reformulated the distinction between mind and body in epistemic and metaphysical rather than primarily moral or religious terms.

This distinction, however, should not be interpreted too rigidly. Descartes does not merely oppose two isolated substances; he also attempts to explain their interaction ([26]). The human being is therefore neither pure spirit nor pure mechanism, but a problematic union of thought and corporeality. Descartes famously locates this interaction in the pineal gland ([27]), thereby acknowledging that mental and bodily states cannot be understood as entirely disconnected realities.

What matters for the present discussion is that, within this framework, machines belong to the order of *res extensa*. No matter how sophisticated their operations become, they remain governed by mechanisms, geometry, and motion rather than by first-person awareness. A machine may imitate reasoning or language, but imitation alone is insufficient to establish genuine thought in the Cartesian sense. The distinction between *res cogitans* and *res extensa* is therefore not simply classificatory; it functions as a criterion for determining whether intelligence is merely operational or genuinely self-aware.

The Cartesian framework can then be formalized as follows. The starting point is radical doubt, which functions as a filter: whatever survives it can serve as a foundation. The result is the *cogito*. Even if I am deceived, I cannot be deceived about the fact that I am undergoing the act of thinking. The Cartesian formula may therefore be expressed as follows:

$$D(x) \rightarrow T(x) \quad (1)$$

where $D(x)$ means “ x doubts,” and $T(x)$ means “ x thinks.” The *cogito* does not state that existence in general depends on thinking. Rather, it states that thinking provides the subject with indubitable certainty of its own existence. More precisely:

$$T(x) \rightarrow C_x(E(x)) \quad (2)$$

where $C_x(E(x))$ means “ x is certain of its own existence.” This distinction is crucial. The absence of actual thinking does not automatically entail non-existence:

$$\neg T(x) \nrightarrow \neg E(x) \quad (3)$$

it entails, at most, the loss of Cartesian self-certainty:

$$\neg T(x) \rightarrow \neg C_x(E(x)). \quad (4)$$

This gives us the first methodological rule of the paper: the *cogito* is primarily an epistemic criterion, not an ontological switch.

A second distinction is required. For Descartes, thinking is not merely the production of correct answers. It includes doubting, affirming, denying, willing, judging, imagining, and understanding. In other words, thinking is not reducible

to performance. It is a first-person act through which the subject is present to itself. A machine may produce a theorem, a diagnosis, a summary, or a decision, but from the Cartesian point of view, this is not sufficient to establish that the machine is *res cogitans*. What matters is not only whether an output is generated, but whether there is a subject to whom the act appears as its own.

This allows us to distinguish between three senses of thinking. Functional thinking, $T_f(x)$, refers to the capacity to process information, produce inferences, solve problems, or generate coherent linguistic outputs. Cartesian thinking, $T_c(x)$, refers to first-person self-presence: the capacity of a subject to experience itself as doubting, judging, affirming, or denying. Agential thinking, $T_a(x)$, designates the capacity to endorse judgments, assume responsibility for beliefs, critically evaluate reasons, and take ownership of decisions. A subject may therefore remain conscious while progressively renouncing agential thought through systematic cognitive delegation.

This tripartite distinction allows the paper to avoid a recurring ambiguity in debates on AI. Much of the contemporary discussion implicitly reduces thought to $T_f(x)$, that is, to functional performance alone. The Cartesian framework defended here argues that neither intelligence nor subjectivity can be reduced to successful output generation. The methodological claim of the paper may therefore be summarized as follows:

$$T_f(x) \nrightarrow R_c(x) \quad (5)$$

where $R_c(x)$ means “ x is *res cogitans*”. Functional performance alone is insufficient to establish the existence of a thinking subject. By contrast:

$$T_c(x) \rightarrow R_c(x) \quad (6)$$

which means that only first-person self-presence satisfies the Cartesian criterion for genuine thought. From this, it implicitly follows that the notion of $T(x)$ employed in (1)–(4) must be interpreted, within the Cartesian framework, specifically as $T_c(x)$. The distinction between the three forms of thinking introduced above prevents the collapse of all forms of cognition into the Cartesian concept of thought and clarifies that the *cogito* concerns first-person self-presence rather than mere functional performance.

III. THE PARADOX

The paradox introduced in this paper emerges from two apparently symmetrical questions. First: if human beings progressively delegate cognitive activity to AI ([28]–[30]), do they cease to be *res cogitantes*? Second: if machines increasingly reproduce the external signs of intelligence, should they themselves be regarded as thinking beings?

At first sight, the first question appears almost absurd. Delegation itself presupposes thought. To entrust reasoning to a machine already requires intention, selection, and assent. One must decide to rely on AI, authorize its outputs, and accept its role within one’s cognitive activity. In this sense, the act of delegation confirms rather than abolishes the

cogito. The paradox becomes philosophically significant only when delegation is imagined in its extreme form: a condition in which the subject no longer exercises independent judgment and progressively transforms AI into an epistemic authority.

In such a scenario, the human being risks becoming less a thinking subject than an executor of externally generated outputs. Even here, the Cartesian conclusion is more subtle than it initially appears. The *cogito* establishes that if I think, then I possess indubitable certainty of my existence. It does not establish the inverse proposition that, if I cease actively thinking, I therefore cease to exist (as discussed in the previous section). Such a conclusion would illegitimately transform an epistemic principle into an ontological theorem. The disappearance of reflective thought would undermine certainty, not necessarily existence itself.

The real danger posed by total delegation is therefore not metaphysical annihilation but the weakening of agency. A subject who systematically renounces judgment, interpretation, and critical assent does not disappear biologically or phenomenologically. Rather, the subject becomes increasingly opaque to itself; thinking is reduced to operational acceptance. In this condition, the human being survives, but no longer experiences itself as the active source of its own beliefs and decisions.

The second side of the paradox concerns AI itself. Contemporary systems increasingly reproduce behaviours traditionally associated with intelligence: argumentation, explanation, linguistic adaptation, and even forms of apparent self-reference. This creates the impression that machines genuinely think. From a Cartesian standpoint, the production of sophisticated outputs is not sufficient to establish the existence of a thinking subject ([4], [7], [8], [31]). Thinking, in a strong sense, is not merely the capacity to generate responses but the capacity to encounter oneself directly as thinking.

For this reason, the distinction between functional performance and conscious self-presence becomes decisive. AI systems clearly demonstrate increasingly advanced forms of functional intelligence, in some cases even outperforming human decision-making ([32]–[36]). They process information, generate arguments, and adapt to context with extraordinary efficiency. What remains unclear, however, is whether these operations are accompanied by subjective awareness. Machines may simulate the external grammar of thought while lacking the first-person self-relation through which the *cogito* becomes possible.

The paradox is therefore double. Humans risk abandoning reflective agency without ceasing to exist, while machines increasingly resemble thinking beings without necessarily becoming *res cogitantes*. The contemporary problem is not simply that AI becomes more intelligent, but that intelligence itself is increasingly interpreted in purely operational terms. Once thought is reduced to successful performance, computation begins to appear indistinguishable from subjectivity. This is the fundamental mistake of post-Cartesianism: reducing thought to its outputs, even when those outputs are produced by entities devoid of genuine rational self-awareness (*ratio sine sensu*).

This transformation may be described as the emergence of what has been defined *Cartesian AI* ([15]). The expression is deliberately paradoxical. On the one hand, contemporary

AI inherits the modern ambition to formalize reason and transform intelligence into calculability. On the other hand, it abandons the very element that made the Cartesian project epistemically foundational: first-person self-certainty. AI therefore preserves the external structure of rationality while eliminating the subjective dimension through which thinking becomes internally experienced. In this sense, AI does not refute Descartes. Rather, it radicalizes one of the deepest tensions already contained within Cartesian philosophy, namely the gap between the appearance of intelligence and the existence of a thinking self.

IV. CONCLUSIONS

This paper has argued that the contemporary relationship between AI and Cartesian philosophy should not be understood primarily as a technological problem, but as a conceptual and epistemological one. The central issue is not whether machines can perfectly imitate human cognition, but what conditions must hold for something to count as thinking in the first place.

The Cartesian framework developed here leads to two main conclusions. First, the delegation of cognitive functions to AI does not imply the ontological disappearance of the human subject. The *cogito* guarantees certainty of existence within the act of thinking, but it does not entail that the suspension of reflective thought automatically abolishes existence itself. What is threatened by total delegation is therefore not human being as such, but the subject's active relation to its own judgments, decisions, and interpretations.

Second, the increasing sophistication of AI systems does not by itself establish the existence of artificial consciousness. Machines may reproduce many external markers of intelligence while remaining fundamentally external to the first-person self-presence that defines *res cogitans*. Functional performance and subjective awareness remain conceptually distinct.

None of this implies that artificial consciousness is conceptually impossible. Contemporary philosophy remains divided between functionalist approaches, which identify thinking with the successful realization of cognitive functions, and perspectives that regard consciousness and phenomenological self-awareness as irreducible dimensions of thought ([37]–[42]). The present argument does not attempt to resolve this debate. Its claim is more limited and more precise. If AI lacks consciousness and first-person self-presence, then it cannot qualify as *res cogitans* in the Cartesian sense, regardless of the sophistication of its performances. If, however, artificial systems were genuinely conscious, then the *cogito* would apply to them as well: an artificial subject capable of self-certainty would constitute not a refutation of human subjectivity, but the emergence of an additional thinking being. Human existence, therefore, should not be conceived as a zero-sum game with AI.

The deeper significance of the paradox lies elsewhere. Contemporary AI increasingly encourages a conception of intelligence centred on efficiency, optimization, prediction, and output generation. In this context, thinking risks being redefined as a purely operational process detached from self-

awareness and agency. The danger is therefore not simply that machines may one day appear human, but that human beings may progressively cease to experience thinking as something they actively do.

In this respect, the *cogito* retains a critical philosophical function. It reminds us that thinking is not exhausted by calculation, linguistic fluency, or successful performance. Thought is also self-presence, judgment, and responsibility. AI may imitate many external signs of intelligence, but the question of who doubts, who understands, and who takes responsibility for a judgment remains irreducibly open.

The Cartesian paradox therefore does not culminate in the replacement of human beings by machines. Rather, it reveals a historical transformation in the meaning of thought itself: from an act of self-certainty to a process increasingly measured in terms of operational efficiency. Whether this transformation enriches or impoverishes human subjectivity remains one of the central philosophical questions of the age of AI.

REFERENCES

- [1] Candrian, C., & Scherer, A. (2022). Rise of the machines: Delegating decisions to autonomous AI. *Computers in Human Behavior*, 134, 107308.
- [2] Grady, D. (2024). Artificial intelligence as surrogate decision-maker. *JAMA Internal Medicine*, 184(9), 1007.
- [3] Hoskins, A. (2024). AI and memory. *Memory, Mind & Media*, 3, 1–21.
- [4] Dreyfus, H. (1992). *What computers still can't do: A critique of artificial reason*. MIT Press Ltd.
- [5] Haugeland, J. (1998). *Having thought: Essays in the metaphysics of mind*. Harvard University Press.
- [6] Robbins, S., & Blundell, I. (2026). Losing our voice? Generative AI and the degradation of human expression. *Mind & Machines*, 36(2).
- [7] Bunge, M. (1956). Do computers think? *The British Journal for the Philosophy of Science*, 7(26), 139–148.
- [8] Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417–424.
- [9] Searle, J. R. (1990). Is the brain's mind a computer program? *Scientific American*, 262(1), 25–31.
- [10] Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42(1–3), 335–346.
- [11] Michie, D. (1993). Turing's test and conscious thought. *Artificial Intelligence*, 60(1), 1–22.
- [12] Floridi, L. (2023). AI as agency without intelligence: On ChatGPT, large language models, and other generative models. *Philosophy & Technology*, 36, 1–7.
- [13] Stoljar, D., & Zhang, Z. V. (2024). Why ChatGPT doesn't think: An argument from rationality. *Inquiry*, 1–29.
- [14] Quattrocioni, W., Capraro, V., & Perc, M. (2025). *Epistemological fault lines between human and artificial intelligence*. arXiv. <https://doi.org/10.48550/arXiv.2512.19466>
- [15] Pernagallo, G. (2025). Cartesian AI. *Synthese*, 207, article number 202.
- [16] Morioka, M. (2023). Introduction: Descartes and artificial intelligence. *Journal of Philosophy of Life*, 13(1), 1–4.
- [17] Hull, G. (2023). Unlearning Descartes: Sentient AI is a political problem. *Journal of Social Computing*, 4(3), 193–204.
- [18] Avramides, A. (1996). Descartes and other minds. *Teorema*, 16(1), 27–46.
- [19] Descartes, R. (2016). *Meditationes de prima philosophia*. Editori Laterza.
- [20] Descartes, R. (1850). *Discourse on the method of rightly conducting the reason, and seeking truth in the sciences*. Sutherland and Knox.
- [21] Kintz, J. S. (2019). The illuminative function of the agent intellect. *British Journal for the History of Philosophy*, 27(1), 3–22.
- [22] Shen, L. (2025). Not the machine's fault: Taxonomising AI failure as computational (mis)use. *AI & SOCIETY*, 40, 5793–5807.
- [23] Gerson, L. P. (1986). Platonic dualism. *The Monist*, 69(3), 352–369.
- [24] Seitz, J. A. (2015). Religious dualism and the problem of dual religious identity. *Buddhist-Christian Studies*, 35, 49–55.
- [25] Weir, R. S. (2022). Christian physicalism and the biblical argument for dualism. *International Journal for Philosophy of Religion*, 91, 115–138.
- [26] Popper, K. R. (1994). *Knowledge and the body-mind problem: In defence of interaction*. Routledge.
- [27] Descartes, R. (1989). *The passions of the soul*. Hackett Pub. Co.
- [28] Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58(1), 7–19.
- [29] Van Rooij, I., Guest, O., Adolff, F., de Haan, R., Kolokolova, A., & Rich, P. (2024). Reclaiming AI as a theoretical tool for cognitive science. *Computational Brain & Behavior*, 7, 616–636.
- [30] Ali, A. M., Dornaika, F., & Charafeddine, J. (2026). Agentic AI: A comprehensive survey of architectures, applications, and future directions. *Artificial Intelligence Review*, 59(11).
- [31] Bringsjord, S., & Zenzen, M. (1997). Cognition is not computation: The argument from irreversibility. *Synthese*, 113, 285–320.
- [32] Pernagallo, G. (2024). *Overcoming asymmetric information: A data-driven approach*. In J. Włodarczyk & D. Raban (Eds.), *The Elgar companion to information economics*. Edward Elgar Publishing.
- [33] Pernagallo, G., & Torrisi, B. (2022). A theory of information overload applied to perfectly efficient financial markets. *Review of Behavioral Finance*, 14(2), 223–236.
- [34] Kleinert, T., Waldschütz, M., Blau, J., Heinrichs, M., & Schiller, B. (2026). AI outperforms humans in establishing interpersonal closeness in emotionally engaging interactions, but only when labelled as human. *Communications Psychology*, 4, article number 23.
- [35] Huang, J., & Zhou, A. (2025). Generative AI outperforms humans in social media engagement: Evidence from GPT-4 and the FIIT model. *Public Relations Review*, 51(5), 102643.
- [36] Kiškis, M. (2026). AI on the path to good decisions. *Sustainable Futures*, 11, 101644.
- [37] Chalmers, D. (1996). *The conscious mind. In search of a fundamental theory*. Oxford University Press.
- [38] Lewis, D. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy*, 50(3), 249–258.
- [39] Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review*, 83(4), 435–450.
- [40] Fodor, J. A. (1975). *The language of thought*. Thomas Y. Crowell Company, Inc.
- [41] Husserl, E. (1982). *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy. First Book: General Introduction to a Pure Phenomenology*. Springer.
- [42] Dennett, D. C. (1991). *Consciousness explained*. Back Bay Books. Little, Brown and Company.