Rocznik Naukowy Kujawsko-Pomorskiej Szkoły Wyższej w Bydgoszczy.
Transdyscyplinarne Studia o Kulturze (i) Edukacji, 2023, nr 18 specjalny, z. 1

ARTIFICIAL INTELLIGENCE AND UNDERSTANDING. A HERMENEUTIC PERSPECTIVE

dr Tomasz Kalaga
Akademia Kujawsko-Pomorska
e-mail: t.kalaga@akp.bydgoszcz.pl, https://orcid.org/0000-0002-8648-4004

Summary: The present article situates the question of textual interpretation in the context of the contemporary rise of artificial intelligence. Drawing on Martin Heidegger's distinction between calculative and meditative thinking and discussing the implications of the Turing test and Chinese room experiment, the author argues that the absence of intentionality in machine-generated responses necessitates a reconsideration of the seemingly outdated hermeneutic approach to textual interpretation. Recalling the traditions of Schleiermacher and Dilthey, the paper emphasizes understanding not merely as decoding meaning, but as grasping the lived experience (Erlebnis) behind a text. In a world increasingly shaped by artificial agents capable of mimicking language and thought, hermeneutics invites us to reconsider what it means to interpret, to understand, and ultimately, to be human. The article calls for a renewed focus on the human experience as the core of meaningful interpretation.

Key words: artificial intelligence, interpretation, intentionality, calculative and meditative thinking, Erlebnis, hermeneutics.

Throughout the winter and summer semesters of 1951 and 1952 at the University of Freiburg, Martin Heidegger, one of the most influential philosophers of the twentieth century, conducted a series of lectures which was later to be published as a volume entitled *What is Called Thinking?* The provocative question mark raises expectations of learning the answer to the titular question, a definition – perhaps even the definition – and if not that, then at least some kind of a framing for such a complex, composite faculty of a human being. The reader has a right to expect this answer above all from a thinker who approximately twenty years earlier had published *Being and Time*; a work that in a revolutionary manner redefines the concept of a human being through an elaborate philosophical analysis of its relations with time, world and the ultimate horizon of its existence – death. The book, considered by both his contemporaries and modern scholars to be one of the most important works of the era is, *par excellence,* a eulogy of thinking, in form and content alike – the meticulously structured intricacies of the multi-layered argument all concern, in one manner or another, Dasein as a being essentiated by what can only be termed in the common idiom a modality of thought.

And yet, nearly a quarter of a century later, Heidegger writes

> We come to know what it means to think when we ourselves try to think. If the attempt is to be successful, we must be ready to learn thinking.
>
> As soon as we allow ourselves to become involved in such learning, we have admitted that we are not yet capable of thinking. [Heidegger 1968, p. 5]

That thought-provoking statement on thought, another step taken in the late stage of the life-long journey of philosophizing stands in stark juxtaposition to a question considered eighteen years earlier by a British mathematician Alan Turing in his now canonical text "Computing Machinery and Intelligence" – "Can machines think?" [Turing 1950, p. 433]. It is certainly one of the moments where the forking out between what Heidegger names *meditative* and *calculative* thinking [Heidegger 1966] becomes starkly apparent in its radical differentiation of the essencing of a human being, a differentiation which one may already, from the perspective of the present times, begin to perceive as a hallmark of hermeneutics of the future.

What came to be known as the Turing test, originally named by its author "the imitation game," despite its various modern criticisms (for instance, [Heyes, Ford, 1995], [Vardi 2014], [McDermott 2014]) enjoyed enormous popularity as a conceptual design in the days before digital technology was developed well enough for serious practical work on artificial intelligence to be undertaken and is still perceived by the research community as an attractive benchmark for the so-called strong AI [Gonçalves 2022]. In the Turing test, the task of the machine (in all actuality – a computer program) is to deceive its interlocutor by giving them the impression that they are talking to a human being. The test is blind test, where the human interlocutor communicates with two other interlocutors without seeing them directly – some sort of a physical barrier is essential – for example, in terms of today's technology a computer or a smartphone screen. Based on the questions the interlocutor asks and the answers they receive, they are supposed to judge which of their conversationalists is human and which is not. If the machine's answers are indistinguishable from the answers a human would give, the test is passed by the artificial intelligence.

Of significance here is not so much whether the Turing test is a reliable means of testing the capabilities of AI, but the shift in perspective that marks the aforementioned moment of bifurcation between meditative and calculative modes and is exemplified by Turing's departure from the original question – "Can machines think?" – as "too meaningless to deserve discussion" [Turing 1950, p. 442] in favour of a differently phrased question – "Are there imaginable digital computers that would do well in the imitation game?" [Turing 1950, p. 442]. The shift is significant as it, in a certain sense, lowers the bar of the test: from an exercise that would indicate the presence of consciousness or self-awareness in the "thinking machine" to determining the quality of imitation of such. As Carter writes: "Keep in mind that the claim is not that passing the Turing test is sufficient for having a mind. The thought is that passing the Turing test gives us good grounds to suppose that the test subject has a mental life." [Carter 2007, p. 111, emphasis mine].

In practice, these good grounds translate into very concrete directives for artificial intelligence design. As Russel and Norvig write in Artificial Intelligence. A Modern Approach, to pass the Turing test

> the computer would need the following capabilities:
>
> - natural language processing to communicate successfully in a human language;
> - knowledge representation to store what it knows or hears;
> - automated reasoning to answer questions and to draw new conclusions;
> - machine learning to adapt to new circumstances and to detect and extrapolate patterns. [Russell, Norvig 2021, p. 32]

and to pass the total Turing test which requires physical interaction with the environment

> a robot will need
>
> - computer vision and speech recognition to perceive the world;
> - robotics to manipulate objects and move about. [Russell, Norvig 2021, p. 33]

All six of the above bulletpoints constitute today the core of artificial intelligence research, research which is, above all, directed at producing a calculative semblance of thinking, capable of winning the imitation game, yet remaining what Catherine Havasi calls the profitable “low-hanging fruit” as opposed to the “high-hanging fruit” of strong AI[1].

A critique of sorts, or rather an identification of certain weaknesses of the Turing Test was carried out by American philosopher John Searle in what is known as the Chinese room thought experiment [Searle 1980]. The experiment (a blind test again) was designed by Searle to argue against a feasible possibility of strong AI, ultimately concluding that a semblance of understanding on the part of a machine remains just that – an illusory semblance – as its emulation of understanding lacks a crucial component – that of intentionality. In essence, the computer is likened to a human who, in a series of input-output responses imitates knowledge of the Chinese language by manipulating Chinese characters according to a set of prescribed rules but without understanding their actual meaning. In this way the computer generates feedback that is no different from the information a human fluent in Chinese would give and thus may succeed in convincing its interlocutor that it is a thinking being capable of providing meaningful responses to questions given. Yet naturally, actual thinking, in this case exemplified by understanding, simply does not occur “[…] because the formal symbol manipulations by themselves don’t have any intentionality; they are quite meaningless; they aren’t even symbol manipulations, since the symbols don’t symbolize anything. In the linguistic jargon, they have only a syntax but no semantics. Such intentionality as computers appear to have is solely in the minds of those who program them and those who use them, those who send in the input and those who interpret the output” [Searle 1980, p. 422].

And yet, mankind invents what is meant to pass for “thinking machines.” Whether this is a question of supplementing human incapacity, striving for a grander scientific achievement or a simple ego-trip is in the long run immaterial as of the third decade of the twenty-first century they stand, in some form, among us.

[1] *Artificial Inteligence: An Inhuman Future?* Full Panel Discussion, Oxford Union. https://www.youtube.com/watch?v=uqeqnE7CLr8&t=236s

Take, for instance, Sophia, a moving, talking robot capable of imitating human interaction to a degree that in 2017 it was granted a citizenship of Saudi Arabia.[2] Sophia is shaped like a woman from the waist up (what it is from the waist down remains a mystery to the general public), it speaks in a female voice, its face imitates human expressions. It gives the impression of being aware of its interlocutor, answers questions intelligently, and can even tell contextual jokes. However frivolous the notion of bestowing a citizenship in Saudia Arabia may sound and is by majority interpreted as a publicity stunt indented to attract AI research to boost the country's status and increase economic benefits, granting this privilege to a robot certainly has a symbolic dimension – for the first time in the history of mankind, a technological product of human civilization, a representative of "artificial intelligence" becomes a citizen of an independent state.

Symbolic gestures aside, the proliferation AI in contemporary life, the ever-increasing presence of bots or various mutations of ChatGPT software, which is now capable of producing texts coherent and informative enough to imitate a human author, effectuates a hermeneutically paradoxical situation: *an imitation of thinking becomes a subject of intentionally directed human understanding*.

Hermeneutics is otherwise known as the art of interpretation, a foundation for understanding in contexts of appearance of expressions of thought, experience, emotional states, fantasies, all that in the nineteenth century fell under the broad umbrella term *Geisteswissenshaften*, the sciences of the spirit, or as we would put it today, the humanities. The word "hermeneutics" derives from ancient Greek, it is thus rooted in Hellenic culture, commonly called the cradle of Western civilization. *Hermēneia* is a noun which means interpretation, and the name *hermeios* referred to a priest of the Delphic oracle [Palmer 1969, p. 13]. The etymological source here clearly points to Hermes, who, as we remember, in Greek mythology served as a messenger of the gods and was "associated with the function of transmuting what is beyond human understanding into a form that human intelligence can grasp" [Palmer 1969, p. 13]. Hermes was the first archetypal interpreter – translating the language of the gods into human language, he simultaneously rendered the content accessible to a mortal mind. Not without significance is the fact that Hermes was credited with inventing writing and passing it onto mankind as a gift for representing and, above all, recording meaning, one of many reasons for his identification with an even more ancient deity of Egypt – Thoth, who "had several faces, belonged to several eras, lived in several homes," caught in "the discordant tangle of mythological accounts [which] should not be neglected" [Derrida 1981, p. 86].

For many centuries, hermeneutics remained primarily the art of interpretation – that is, of translation. Its task was *to make the hermetic hermeneutic*, that is, to turn the hidden or obscure into the transparent and explained (hermetic is yet another word which owes its origin to Hermes, referring to knowledge restricted and accessible only to select few). Hermeneutics should be regarded here in its multiplicity, as there was no single coherent field of this art, but each sphere of human activity requiring interpretation had its own methodology specific to its tasks and goals – thus there were legal, literary, historical, and biblical hermeneutics. The common denominator

[2] https://en.wikipedia.org/wiki/Sophia_(robot)

was the presence of a text and the overarching goal of its explanation. Hermeneutics sought to make the text understandable – accessible to reason.

In the nineteenth century, a radical turn occurs in hermeneutics. The Enlightenment's celebration of reason, while sustaining progress in the sciences, casts its own shadow, provoking a reaction that pupates into one of the most powerful and fertile humanistic movements in the contemporary history of Western thought – Romanticism, which – above and beyond the principles of reason and objective, intellectual cognition – affirms the human being in its spiritual dimension. It is on this ground that the metamorphosis of hermeneutics takes place as Friedrich Schleiermacher, a German philosopher and theologian, undertakes the task of unifying the scattered hermeneutic methodologies under the aegis of a new goal. In keeping with the spirit of the age, Schleiermacher postulates that hermeneutics should first and foremost become the art of *understanding* [Palmer 1969, pp. 84–97].

It is by no means understanding defined as seeking a logical solution to a problem or a scrupulous analysis of the structures of the object of research. Such methodological approaches belong to the domain of the sciences. At stake here are not natural phenomena, but texts written by the human hand – products of the human spirit, human consciousness, human emotions, sensations, thoughts or imagination. In Schleiermacher's view, the primary goal of hermeneutics becomes to understand the other; more specifically, to understand the message of the other that reaches us through a written text. Such a statement may initially seem trivial; after all, we assume a priori, hardly giving it a second thought, that, as it were, on the other side of the text stands its author, a thinking, feeling human being. The triviality of this statement (somewhat encapsulated by the far too often asked sacramental question "What did the author have in mind?" negotiated by literary theory in various ways over the past century) ceases to be so obvious, however, in light of the brief discussion on artificial intelligence that opened this text.

With the current rate of development of technology, the level of intensity (and funding!) of AI research and the advent of quantum computers, it seems feasible that in the not-too-distant future algorithms will be created which can satisfactorily pass the Turing test and demonstrate such a sophisticated ability to manipulate human language systems and have a large enough data set at their disposal that the Chinese Room effect will fade into the background as more of a philosophical curiosity than any actual practical hindrance. We will perhaps find ourselves in a situation where it will be impossible to distinguish between a human interlocutor and highly sophisticated software – furthermore, the latter may prove to be a more engaging, witty, and erudite interlocutor.

It is slowly becoming clear that perhaps the main problem birthed by the advancement of artificial intelligence is not the fear of its rebellion, taken to its extreme in apocalyptic visions of enslavement or destruction of the human kind. Rather, it is the question of humanity itself, and, more specifically, the question of *the distinguishing difference*. What distinguishes and, at the same time, defines a human being against an intelligent machine, algorithm or program? And here the question of hermeneutic understanding acquires a new depth. Continually at stake is the matter of interpretation – which is always deeply embedded in mediation. Communication, after all, is never direct in the sense of being semiotically unmediated – not even with another human being; the primary means of communication is language in its more or less complex forms. From such a perspective, there is no significant difference between interpreting a story written by a human and one generated by software.

The difference only occurs when we consider again the purpose which defined and directed Schleiermacher's investigations – it is not the meaning of the text itself that is the object of understanding, but the thought which resides behind it – to understand a person through a text. That text will always be subject to the moulds and grafts of language itself, which will shape it in ways that are often unpredictable and unintended by the author. Hence the ultimate goal of Romantic hermeneutics, aptly encapsulated in the phrase: "to understand the author better than he himself."[3] It would be difficult today, for a number of reasons, to try to resurrect this goal in the form given by Schleiermacher. Poststructuralist theories have provided too many pertinent arguments for this to be possible. Nevertheless, in light of current technological changes, perhaps once again there arises a need to direct hermeneutic thought toward the self behind the text – be it human or artificial.

For this purpose, we may find some signposts in the writings of Wilhelm Dilthey, in a many ways a spiritual disciple of Schleiermacher. Most contemporary theories of text derived, in one manner or another, from the common stem of (post)structuralism perceive language as an innately semiotic phenomenon and a text, woven from the linguistic fabric, is itself an innately ambiguous artifact, open to multiplicity of readings via the (perhaps infinite) "play of language." Yet Dilthey, remote from structuralist roots, emphasizes a different semiotic aspect – a text is, according to the philosopher, primarily an external manifestation of mental activity grounded in and originating from Erlebnis – lived experience [Palmer, 1969, pp. 100–114]. Thus, before the question of interpreting the sign itself arises, what is important is its source, the internal, mental activity of the sign sender, or, to simplify – behind the message is someone's thought. Dilthey locates that thought in its broader context, foregrounding its – painfully human – source: the experience of life lived.

> The ultimate root of worldview is life [...] I also experience a certain inner state of tranquility [...] in it I internalize other people and things not only as realities that remain in causal relationships with me and each other: life's connections lead away from me in all directions, I relate to people and things, assume a stance towards them, meet their demands and expect something from them. Some bring me happiness, expand my existence and heighten my powers, while others pressure and limit me. […] A friend is a force that intensifies a person's own existence, every family member has a specific place in his life [...] A bench by the entrance, a shady tree, the house and the garden all have their own character and meaning. In this way, the life of each individual produces a world proper to itself. […] From reflection on life, life experience is created. [Dilthey1987, p.121, trans. mine]

This is the source of sign transmission according to Dilthey – life understood as an active, inner experiencing of relations with the world external to a human being. This is how colloquial utterance, everyday communication is created, but this is also, or perhaps, above all, the origin of literature, music, art. Thus we raise the bar high – if such a test were to be devised, the Dilthey test would be constructed according to significantly different criteria than the Turing test.

[3] For the genealogy of the phrase, often attributed either to Schleiermacher himself or Wilhelm Dilthey, see the short but illuminating study by O. F. Bollnow, 1979.

However, what remains all the more puzzling in the light above is Heidegger's statement which opened the present remarks: "We are not yet capable of thinking." What, then, is this thinking proper, or in other words, when already and how will we be able to say of ourselves that we are already thinking? The answer which is not really an answer, Heidegger slowly unfolds over a thirty-year span of work via hundreds of pages of writing. Yet what he offers is but a hint, a signpost the search for the essence of thinking. In conclusion, I will evoke two moments of thought that might provide a little insight into the direction pointed out by the thinker.

The first is a compacted recollection from Division I of *Being and Time*, V.31. *Being there as understanding* [Heidegger, 1996, pp. 182–188], a fragment that predates the statement on thinking by over a decade and concerns the notion of understanding. There, Heidegger recognizes that understanding is not really an activity, as activity is a volitional event. We can perform a certain action or not: I can sit down but I do not have to. I can eat a sandwich, I can look to the left, I can whistle, sing. Understanding does not belong to this sphere – it is an existential, an integral part of what we define by the term "man." A human being exists by understanding, that is, by interpreting. One cannot, just like that, stop understanding. This is not about understanding correctly or incorrectly – it is about understanding as establishing oneself in relation to the world. Man is always an entity in the world. Through understanding, we establish in one gesture both ourselves and the world in which we live.

The second moment occurs in a 1955 lecture "Memorial Address," frequently referred to as simply Gelassenheit. A curious parallel arises – that text was delivered in the face of technological changes that may be perceived just as radical and momentous as the contemporary advent of AI. A few years earlier, the world had learned about the destructive power of the atomic bomb, capable of wreaking destruction on a scale never imagined before. At that time, the first nuclear power plants were also being erected in an effort to tame forms of energy that contain the potential to annihilate the humankind and its world. There, Heidegger also refers to the notion of thinking and the thinking proper, which we have not yet achieved, he contrasts with calculative thinking – the inevitable and most potent threat of the era. Calculating thinking is perfect, rigorous, logical, capable of creating extremely complex intellectual structures and apparatuses whose purpose is to comprehend and manipulate the reality around us. We reap tangible benefits from this thinking in the form of comforts brought about by the progress of technology. However, its essence and, at the same time, its greatest threat, is the treatment of all that which we as humans come into contact with, including ourselves, as resource. Heidegger writes: "The world now appears as an object open to the attacks of calculative thought, attacks that nothing is believed able any longer to resist. Nature becomes a gigantic gasoline station, an energy source for modern technology and industry." [Heidegger 1966, p. 50]. In this perspective, a river becomes a water transport route, an ancient oak tree in a forest becomes raw material, a homestead remains a building, a human being is labor force, a consumer, a collection of data fished out of the global network by ceaselessly scouring algorithms.

The thinking towards which Heidegger directs us is, first and foremost, concern. Concern for being, and therefore, simultaneous concern for both mankind and its world. Man is the shepherd of being, Heidegger writes. It is obvious that without the world there would be no humankind. Less obvious is that without humankind, there would be no world.

Bibliography

- Carter M. (2007), Minds and Computers. An Introduction to the Philosophy of Artificial Intelligence. Edinburgh: Edinburgh University Press.
- Derrida J. (1981) Dissemination. Translated, with Introduction and Additional Notes, by Barbara Johnson. Chicago: The University of Chicago Press.
- Dilthey W. (1987) O istocie filozofii. (1987), Warszawa: PWN.
- Gonçalves B. (2022), "The Turing Test is a Thought Experiment." Minds and Machines, vol. 33, pp. 1–31. https://doi.org/10.1007/s11023-022-09616-8
- Hayes P. & Ford K. (1995). "Turing Test Considered Harmful." In Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, pp. 972–977. https://www.ijcai.org/Proceedings/95-1/Papers/125.pdf
- Heidegger M. (1966), Discourse on Thinking. Translated by John M. Anderson & E. Hans Freund. New York: Harper & Row, Publishers.
- Heidegger M. (1968), What is Called Thinking. Translated by J. Glenn Gray. New York, Evanston, and London: Harper & Row, Publishers.
- Heidegger M. (1996), Being and Time. Translated by John Macquarrie and Edward Robinson. Oxford: Blackwell Publishing.
- McDermott D. (2014), "What Was Alan Turing's Imitation Game?" The Critique. http://www.thecritique.com/articles/what-was-alan-turings-imitation-game/
- Palmer R. (1969), Hermeneutics. Evanston: Northwestern University Press.
- Russell S., Norvig P. (2021), Artificial Intelligence. A Modern Approach. Hoboken: Pearson.
- Searle J. (1980), "Minds, Brains, and Programs." The Behavioral and Brain Sciences, vol. 3, pp. 417–457.
- Turing A. M. (1950), "Computing Machinery and Intelligence." Mind. A Quarterly Review of Psychology and Philosophy, vol. LIX, no. 236, pp. 433–460.
- Vardi M. Y. (2014), "Would Turing have passed the Turing Test?" Communications of the ACM, vol. 57, no. 9. https://doi.org/10.1145/2643596

SZTUCZNA INTELIGENCJA I ROZUMIENIE. PERSPEKTYWA HERMENEUTYCZNA

Streszczenie: Niniejszy artykuł lokuje problematykę interpretacji tekstu w kontekście współczesnego rozwoju sztucznej inteligencji. Opierając się na rozróżnieniu Martina Heideggera między myśleniem rachującym i kontemplacyjnym oraz omawiając implikacje testu Turinga i eksperymentu chińskiego pokoju, autor sugeruje, że nieobecność intencjonalności w tekstach generowanych przez AI wymaga ponownego zastanowienia się nad pozornie przestarzałym hermeneutycznym podejściem do interpretacji. Nawiązując do tradycji Schleiermachera i Diltheya, artykuł podkreśla, że rozumienie nie polega jedynie na dekodowaniu znaczenia, ale na uchwyceniu doświadczenia życiowego (Erlebnis) stojącego za tekstem. W świecie coraz bardziej kształtowanym przez algorytmy zdolne do naśladowania języka i myśli, hermeneutyka zachęca nas do ponownego przemyślenia tego, co oznacza interpretować, rozumieć i ostatecznie być człowiekiem oraz do traktowania przeżycia i doświadczenia ludzkiego jako kluczowych elementów w procesie interpretacji.

Słowa kluczowe: sztuczna inteligencja, interpretacja, intencjonalność, myślenie rachujące i myślenie kontemplacyjne, Erlebnis, hermeneutyka.