FILOSOFIJA. SOCIOLOGIJA
2025. T. 36. Nr. 1, p. 14–22

ISSN 0235-7186 eISSN 2424-4546
DOI: https://doi.org/10.6001/fil-soc.2025.36.1.2

Large Language Models and the Enhancement of Human Cognition: Some Theoretical Insights

AISTĖ DIRŽYTĖ
Department of Philosophy and Cultural Studies, Faculty of Creative Industries, Vilnius Gediminas Technical University, 1 Trakų Street, 00132 Vilnius, Lithuania
Email: aiste.dirzyte@vilniustech.lt

This essay explores the possible contribution of Large Language Models (LLMs) to human cognition. It investigates whether human cognition can be enhanced by advanced AI systems such as LLMs. Can LLMs make people as learners smarter, or, on the contrary, make them reason/think less? The author discusses the concepts of human and artificial intelligence and examines LLMs as advanced AI systems, which use deep learning techniques and can be considered as excelling in neural network architectures, data volume, generalisation and scalability. The author suggests that while LLMs can assist in facilitating numerous cognitive tasks, more research and philosophical inquiry is needed to understand whether such kind of AI assistance would make people cultivate human intelligence more, and not less. Presumably, Large Language Models (LLMs) can contribute to human intelligence and cognition just under strict (addressed existing limitations, questioning prompting, time-sensitivity, etc.) conditions. However, it is important that these theoretical considerations could be verified by experimental research.

Keywords: Large Language Models (LLMs), artificial intelligence, human intelligence, cognition, 'extended mind' theory

INTRODUCTION

Philosophers have for ages analysed the nature of consciousness, language, reasoning, knowledge and learning (Jori 2022; Annas 2023). In the last decades, this analysis was enriched with questions on the function and meaning of technologies (Grant 2023). The emergence of advanced artificial intelligence systems like Large Language Models (LLMs) adds new aspects to these debates, to consider whether these models can extend human intelligence, whether they truly 'understand' language, and what the ethical concerns are, related to their deployment. From a philosophical viewpoint, LLMs represent a new scope for awareness whether technologies can enhance human cognition.

In examining AI and LLMs through philosophical lenses, ranging from epistemology to the philosophy of mind, and from ethics to hermeneutics, both cognitive enhancement

potential and also substantial limitations rooted in the nature of artificial intelligence can be disclosed (Candiotto 2023; Ongaro et al. 2024). While LLMs can serve as powerful tools that can contribute to cognition through nearly unlimited processing capacities, it essentially operates in a realm of syntactic analysis, lacking moral agency and the reflective qualities of consciousness. The question onward is how to employ LLMs in ways that enrich human cognition, while remaining vigilant to the threats of bias, overreliance, and possible erosion of critical thinking. As LLMs develop, ongoing philosophical discourse will be necessary for guiding responsible progress and ensuring that the transformative potential of LLMs is neither underestimated nor overestimated.

One of the most promising applications of LLMs, related to the questions on cognitive enhancement, lies in education. Vygotsky's (1980) concept of the 'zone of proximal development' draws attention to the significance of scaffolding in learning (Vygotsky 1980), and based on this, LLMs presumably can act as scaffolding tools that help learners manage complex problems. Yet, it is not clear if LLMs can serve as tutors or resources that guide learners toward higher levels of knowledge. While LLMs can generate outputs based on billions of datasets, produce comprehension exercises or provide immediate feedback, these models offer on-demand resources, and there is the risk that learners may over rely on automated solutions, outsource too much cognitive labour, and, without engaging critically, this can obstruct genuine cognitive growth.

This article supports the balanced view of transhumanism (Grant 2023) to the agency of AI and explores whether human cognition can be enhanced by technologies, such as LLMs. Can LLMs make people smarter, or, on the contrary, make them reason less? Philosophical discourse on LLMs can reveal that LLMs have potential to dramatically extend human cognitive capabilities, aligning in part with the extended mind thesis.

PHILOSOPHICAL INSIGHTS ON COGNITION AND THE ROLE OF LLMS

The recent philosophical discourse on LLMs is primarily centred around the concepts of 'true' intelligence, the frontiers of practical LLMs applications, and the issues related to ethics or biases (Trepczyński 2023). LLMs are based on the statistical learning of the patterns of language, and some scholars doubt whether a 'mechanistic cognition' can be described as genuine intelligence or comprehension. However, LLMs might model decision-making and other cognitive processes, challenging traditional views of AI. Moreover, LLMs rely on statistical associations, which indicates that knowledge can emerge from data-driven synthesis rather than logical deduction, prompting philosophical reflection on the nature of reasoning and the limits of its practical applications.

The first philosophical consideration concerning LLMs is epistemological, related to the nature, origin, and limits of knowledge, the assumptions about how knowledge is acquired and understood. Since Plato's *Theaetetus* (Plato 1997), in Western epistemology, knowledge was described as 'justified true belief', and this was the central conception adopted by numerous theorists (Hintikka 2007). Following this, can information, generated by LLMs, be described as knowledge? LLMs can generate outputs that appear to be both correct and relevant, but do these outputs constitute 'justified true belief'? Gettier (1963) established that knowledge requires more than correct statements – it demands robust justification, and correctness alone does not guarantee justification (Gettier 1963). Furthermore, the answers generated by LLMs appear to be correct, but it does not mean that they necessarily are. LLMs predict words based on statistical patterns in training data, produce probabilistic predictions of the next word,

phrase, or sentence based on patterns in vast training corpora without the reflective, justificatory grounding that characterises traditional epistemic agents. So, despite LLMs can provide answers, they lack the reflective mechanism that grounds justification in an epistemic agent's belief structure and, besides, can provide fictitious answers. To the extent that LLMs generate 'knowledge' and may provide valuable 'information', it is arguably knowledge-by-proxy for the human user who interprets and evaluates the output, and whether LLMs can contribute to genuine 'knowledge' depends on the validation and critical interpretation by human users.

The subsequent consideration lies in hermeneutics, which emphasises the 'fusion of horizons' between the interpreter and the text. Hans-Georg Gadamer's (1975) hermeneutical philosophy, outlined in *Truth and Method*, suggests that comprehension emerges from the 'fusion of horizons' between the interpreter and the text during an interpretive process that is historically and existentially situated (Gadamer 1975). LLMs can automate certain interpretive acts but do they share in the hermeneutical circle of question and answer? Although LLMs can mimic certain interpretive processes while answering queries or summarising texts, their 'understanding' is purely statistical, based on merely pre-trained data, lacking the historical or existential dimension central to hermeneutics. Despite the user's query shapes LLMs response, the model's 'horizon' is purely statistical, missing the existential and historical contexts that Gadamer sees as constitutive of interpretation. As a result, LLMs role in interpretation can be significant yet incomplete without human contextualisation, which highlights both the strengths and the limitations of LLMs as a possible tool for the enhancement of human cognition.

Furthermore, certain considerations concern phenomenology of mind, namely, the concepts of intelligence, consciousness, and the extended mind. John Searle's (1984) famous 'Chinese Room' argument in *Minds, Brains and Science* (Searle 1984) suggests that the syntactic manipulation of symbols could not equate to true semantic understanding. LLMs resemble the Chinese Room operator: they appear to navigate language with an impressive fluency, they manipulate language inputs according to learned rules, yet they operate in a manner parallel to Searle's hypothetical operator, manipulating symbols without grasping their meaning, which refers to the concern that the syntactic manipulation of symbols does not entail semantic understanding. This philosophical lens points out that LLMs lack intentionality, the subjective, first-person perspective that is integral to conscious understanding, therefore, the subjective aspect of cognition is missing.

On the other hand, Andy Clark and David Chalmers' 'extended mind' theory (1998) posits that cognitive processes can extend beyond the biological brain, incorporating external artifacts and tools (Clark, Chalmers 1998). From this perspective, LLMs can be seen as an extension of one's cognition, which can function as a supplementary cognitive resource, allowing humans to offload certain tasks (e.g. summarisation, quick retrieval of information). Such a perspective locates cognition in the interaction between an individual and the external tool (Candiotto 2023; Ongaro et al. 2024). This 'extended mind' framework suggests that if used thoughtfully, LLMs can augment human cognition by expanding the range and speed of human intellectual operations. By distributing selected tasks, humans may augment their capacity to think more broadly and creatively; yet the question remains whether this enhancement is purely instrumental or whether it alters the very structure of human cognition.

Besides, there are several other significant philosophical considerations. Jürgen Habermas (1984), in his *Theory of Communicative Action*, emphasised the significance of communicative rationality within the public sphere (Habermas 1984). Along with this model, if LLMs

become widespread knowledge facilitators, do they bolster or undermine critical discourse? Do they amplify or diminish human agency in dialogue? Theoretically, LLMs might ease access to information, yet there is also a risk of creating a dependence on AI systems, developed by tech groups. This consequently raises concerns about intellectual autonomy and the possible erosion of critical thinking, especially if users treat LLMs' outputs as authoritative without a careful examination. Thus, there is a risk of dependence if users solely accept LLM outputs uncritically, eroding the spirit of reasoned debate and intellectual autonomy. However, LLMs can also empower users by immediately providing insights and references on the questions they ask. Following this, it appears that LLMs can lead to both human cognition empowerment and dependence.

Further, some considerations are related to algorithmic biases and the concepts of justice and fairness. Philosophers of technology, following in the footsteps of Langdon Winner's (1980) observation that artifacts can embody politics, argue that technologies inherently embody social and political values (Winner 1980). Subsequently, LLMs might inherit political and social biases from their training datasets, which existed in their training data, reflecting systemic inequities or prejudices (Liu et al. 2022; Stańczak et al. 2023). Addressing these biases requires an ethical framework that recognises how apparently 'neutral' algorithms can enable injustices. If these biases function unchecked, they can perpetuate systemic injustices. Researchers and practitioners must therefore undertake deliberate measures, ranging from data curation to transparent model design, to ensure that these tools adopt more equitable outcomes. The ethically responsible development of LLMs demands transparent data curation, bias mitigation, and clear accountability frameworks to ensure that their outputs align with values of fairness, justice and equity.

In addition, some philosophical considerations concern practical applications of LLMs (Trepczyński 2023). Specifically, 'philosophical skills' and creativity, as the function of cognition, are the key topics to address. The philosophical questions here are as follows: Can LLMs be described as intellectual partners, which help writers and thinkers brainstorm ideas, synthesise sources, or experiment with novel forms of expression? If so, and if these actions arise from a non-conscious, pattern-based system, does this creative process lose its authenticity? To answer this question, interdisciplinary endeavours are needed. LLMs definitely provide rapid brainstorming or generate new ideas, holding promise as creativity boosters. However, philosophically, it still needs to be answered whether creativity mediated by non-conscious tools diminishes the authenticity of creative acts.

COGNITION AS THE PROCESS OF CHANGE

Learning, as the process of change and acquisition of knowledge by reasoning (Schneider 2024), could be considered one of the main human activities from the infancy nearly to the moment of death. Despite numerous studies, which claim that learning makes learners smarter, the links between learning and intelligence are still ambiguous. Some scholars argue that contribution of learning to intelligence is just a myth (Haier 2014), while some studies provide evidence on the opposite: e.g. a recent study concluded that prolonged intensive training in creative problem-solving can lead to substantial and positive effects on human intelligence (Stankov, Lee 2020).

For centuries, non-religious individuals have been perceiving human intelligence as the highest intelligence in the Universe. Religious people have also believed in human supremacy in comparison to other biological diversity. Historically, this belief can be dated back

to Aristotle's *Historia animalium* (Aristotle 2024). Since 1758, it was fuelled by Carolus Linnaeus taxonomy (Harvey 2007), which distinguished *homo sapiens* from other *homo* species or apes by numerous, and most importantly, cognitive features, as denoted in the Latin word *sapiens* ('discerning, wise, judicious, discrete').

Recently, evolutionary biology proved the conservative biases of scientists (Katz 2019) and no meaning in taking one animal higher than another (Darwin 2008). However, since the rapid progression of Artificial Intelligence (AI) systems, and, namely, Large Language Models (LLMs), a new scholar discussion has fired: which intelligence is higher – human or artificial? (Korteling et al. 2021).

Presuming that human intelligence is higher, artificial intelligence can be conceptualised as a 'cognitive helper', which just directly and rapidly responds to instructions and provides the information that the individual was searching for. As such, it cannot contribute to human intelligence, as it serves just as cognitive processing extension to the extent the user defines. Moreover, this 'cognitive helper' can distort knowledge and hallucinate, providing inaccurate or not valid information.

Presuming that AI is higher, as it has billions of units reaching processing capacities, human intelligence can be conceptualised as a 'biologically limited agent', which is determined by biological variables and has fundamentally different cognitive abilities than AI systems. In this case, AI immensely prevails over human intelligence in numerous cognitive tasks, largely even more than *homo sapiens* prevails over apes.

However, the presumptions on 'higher' or 'lower' intelligence might resemble the concepts of 'higher' and 'lower' in evolutionary biology, which proved that this kind labelling is unmeaningful. Human and artificial intelligence evolved differently; there are several fundamental distinctions between biological and artificial intelligence. Thus, their comparison is meaningful just to identify complementarities and capitalise the strengths of cognition of both.

HUMAN VERSUS ARTIFICIAL INTELLIGENCE

Intelligence, derived from the Latin words *intelligentia* ('insight, ability to know') and *intellegere* ('to understand'), is one aspect of the human mind, described as resourcefulness, cleverness, the ability to solve problems, to notice connections, express thoughts, memorise, create, understand causes, anticipate consequences and learn from experience (Tirri, Nokelainen 2011). People vary in their intelligence, cognitive functioning, cognitive abilities, and capabilities to find solutions to problems. However, human intelligence is multidimensional rather than a general capacity for conceptualisation and problem-solving (Cavas, Cavas 2020). The theory of intellectual diversity created by H. Gardner in the 1990s, proposed analysing human intelligence as a multifaceted factor, a set of linguistic, logical-mathematical, musical, spatial, bodily-kinesthetic, interpersonal and intrapersonal abilities (Gardner 2024). Among these abilities, cognitive capacities are evolutionary most recent, 'embryonal' on the time scale of evolution, reaching possibly not more than 100 thousand years to compare with 'ancient' neural-biological 'intelligence' for survival functions (Korteling et al. 2021), and therefore limited: capacity of human working memory is approximately just 10–50 bits per second, and humans cannot simultaneously execute cognitive tasks that require deliberation and attention (Beltramini 2019). In this regard, AI, which refers to the simulation of human-like intelligence by machines or computer systems, outperforms humans incomparably: it processes large volumes of data at speeds far beyond human capabilities, never gets tired or emotional

to make wrong decisions, it can 'multitask' using algorithms and mathematical models, can run millions of times more complex calculations than humans can (Siemens et al. 2022; Soria Zurita et al. 2022; da Costa 2024). AI learns through training on billions of datasets, generates outputs based on existing patterns or data, and relies on data-driven logic. However, although designed for specific tasks, it interacts based on pre-programmed responses and is susceptible to errors and biases in its training data or algorithms. Human intelligence, despite also being prone to biases (Dror 2020), is flexible in responses and involves both logical reasoning and emotional, intuitive and subjective decision-making processes. Human intelligence adapts to new information not only through 'abstract thinking', but also through experience-based learning and is capable of social understanding, ethical/moral consideration, self-consciousness, generalisation of knowledge in novel emotional contexts, creativity and originality by synthesising disparate ideas or inventing entirely new concepts. Thus, AI mimics many cognitive functions of human intelligence but operates fundamentally differently from it. Some scholars argue that AI lacks mechanisms underpinning true cognition and can just mirror the depth of human thought (Fokas 2023), despite outperforming in computational tasks (Asch et al. 2022; Gyory et al. 2022). Human intelligence is flexible, largely modelled by evolution, biology and social interactions, while AI is static unless explicitly updated. The complementarity of both can work for the benefit of humanity, as evidenced by recent studies in the field of medicine (Asch et al. 2022; Barron et al. 2024; Habicht et al. 2024).

The analysis of the nature of LLMs also raises questions about the nature of 'understanding', management of bias, and the maintenance of critical autonomy. While LLMs excel at information processing and can assist humans in data screening, attempts of data interpretation and some creative tasks, from the philosophical point of view, they fall short of genuine consciousness or self-reflective understanding. Employing LLMs strengths while controlling risks will require continued interdisciplinary dialogue.

LIMITATIONS OF LLMS AND FUTURE DIRECTIONS

Despite strengths, LLMs have significant limitations, which, if not properly addressed, may have a distorting effect on human cognition. Firstly, LLMs can be prone to errors or 'hallucinations', factually incorrect or logically flawed outputs, as its architecture is based just on statistical patterns. Next, LLMs can adopt inconsistent tutoring strategies, like 'leaking' answers instead of guiding learners to solutions. In addition, LLMs struggle with identifying specific errors in learner responses and providing effective feedback (Daheim et al. 2024). Moreover, LLMs still can fail in adapting to the learner's unique needs or tracking long-term progress (Neshaei et al. 2024) and in indicating the confidence level of answers, making it challenging for learners to verify the information provided. Besides, updating LLMs to incorporate new knowledge is limited (Wu et al. 2024). Therefore, while LLMs presumably can serve as scalable, personalised learning tutors, their limitations highlight the need for critical evaluation and continuous improvement by human intelligence.

In the future, learners probably would benefit most from time-sensitive LLMs, designed to provide outputs that are relevant and accurate in the context of time-sensitive tasks. Time-sensitive LLMs enable differentiation between past, present and future, and have mechanisms to integrate new information over time. The development of time-sensitive LLMs could be based on fine-tuning with recent data, real-time data integration, memory-augmented architectures, and temporal embeddings, but the challenges like data freshness, possibility of 'catastrophic forgetting', context ambiguity, or accuracy concerns should be addressed.

Generally, in the future, the role of LLMs as a contributor to the enhancement of human cognition depends on the success of addressing existing challenges and limitations. At present, LLMs still might offer incorrect or oversimplified explanations of complex phenomena, it still can reproduce or amplify biases, posing risks of stereotyping, and, most importantly, LLMs-generated answers still seem very confident, and users can disproportionally trust them. These limitations can contribute to reduced critical thinking and overall poorer self-regulated learning. Consequently, there would be no positive effect on human intelligence.

CONCLUSIONS

The appearance of LLMs raises philosophical questions regarding the nature of knowledge, the assumptions of 'understanding', and the ethical concerns, and therefore demands sustained philosophical inquiry to preserve the spirit of critical reflection and ethical discourse. LLMs present opportunities for complementing human cognition, aligning with the 'extended mind' thesis in their capacity to distribute cognitive labour.

Cognitive processes are central to humans throughout their life, but the links between learning and human intelligence are still vague. Human intelligence can be conceptualised as a 'biologically limited agent', a multidimensional capacity for problem-solving. Artificial intelligence can be conceptualised as a 'cognitive helper', serving as cognitive processing extension, prevailing over human intelligence. Human intelligence is flexible, largely modelled by evolution, biology and social interactions, while AI is static unless explicitly updated, and the complementarity of both can work for the benefit of humanity.

LLMs, as advanced AI systems, which use deep learning techniques, can be considered as excelling in neural network architectures, data volume, generalisation and scalability. LLMs can assist in facilitating numerous cognitive tasks, but more research and philosophical inquiry is needed to understand whether such kind of AI assistance would make people cultivate human intelligence more, and not less. Presumably, Large Language Models (LLMs) can contribute to human intelligence and cognition just under strict (addressed existing limitations, questioning prompting, time-sensitivity, etc.) conditions. However, it is important that these theoretical considerations could be verified by experimental research.

Received 29 November 2024
Accepted 16 January 2025

References

1. Annas, J. 2023. *Hellenistic Philosophy of Mind*. DOI: 10.5840/ancientphil199414162
2. Aristotle. 2024. 'Historia animalium, or The History of Animals', in *The Marvels of the World*. DOI: 10.9783/9780812297812-038
3. Asch, F. M.; Descamps, T.; Sarwar, R.; Karagodin, I.; Singulane, C. C.; Xie, M.; Tucay, E. S.; Tude Rodrigues, A. C.; Vasquez-Ortiz, Z. Y.; Monaghan, M. J.; Ordonez Salazar, B. A.; Soulat-Dufour, L.; Alizadehasl, A.; Mostafavi, A.; Moreo, A.; Citro, R.; Narang, A.; Wu, C.; Addetia, K.; Upton, R.; Woodward, G. M.; Lang, R. M.; WASE-COVID Investigators. 2022. 'Human Versus Artificial Intelligence–based Echocardiographic Analysis as a Predictor of Outcomes: An Analysis from the World Alliance Societies of Echocardiography COVID Study', *Journal of the American Society of Echocardiography* 35(12). DOI: 10.1016/j.echo.2022.07.004
4. Barron, K.; Blaivas, M.; Blaivas, L.; Sadler, J.; Deal, I. 2024. 'Bedside Ultrasound to Identify and Predict Severity of Dysphagia Following Ischemic Stroke: Human Versus Artificial Intelligence', *Ultrasound in Medicine and Biology* 50(1). DOI: 10.1016/j.ultrasmedbio.2023.09.008
5. Beltramini, E. 2019. 'Life 3.0. Being Human in the Age of Artificial Intelligence, by Max Tegmark', *Religion & Theology* 26(1–2). DOI: 10.1163/15743012-02601006

6. Candiotto, L. 2023. 'What I Cannot do Without You. Towards a Truly Embedded and Embodied Account of the Socially Extended Mind', *Phenomenology and the Cognitive Sciences* 22(4). DOI: 10.1007/s11097-022-09862-2
7. Cavas, B.; Cavas, P. 2020. *Multiple Intelligences Theory–Howard Gardner*. DOI: 10.1007/978-3-030-43620-9_27
8. Clark, A.; Chalmers, D. 1998. 'The Extended Mind', *Analysis* 58(1): 7–19.
9. da Costa, L. da S. G. M. 2024. 'Artificial Intelligence and the Transdisciplinary Human Mediation of HPTD-M', *Transdisciplinary Journal of Engineering and Science* 15. DOI: 10.22545/2024/00243
10. Daheim, N.; Macina, J.; Kapur, M.; Gurevych, I.; Sachan, M. 2024. *Stepwise Verification and Remediation of Student Reasoning Errors with Large Language Model Tutors*. Available at: https://arxiv.org/abs/2407.09136
11. Darwin, C. 2008. *The Descent of Man, and Selection in Relation to Sex*. DOI: 10.1038/011305a0
12. Dror, I. E. 2020. 'Cognitive and Human Factors in Expert Decision Making: Six Fallacies and the Eight Sources of Bias', *Analytical Chemistry* 92(12). DOI: 10.1021/acs.analchem.0c00704
13. Fokas, A. S. 2023. 'Can Artificial Intelligence Reach Human Thought?', *PNAS Nexus* 2(12). DOI: 10.1093/pnasnexus/pgad409
14. Gadamer, H. G. 1975. 'Truth and Method', *Journal of Aesthetics and Art Criticism* 36(4): 487–490.
15. Gardner, H. 2024. *The Essential Howard Gardner on Education*. New York, NY: Teachers College Press.
16. Gettier, E. L. 1963. 'Is Justified True Belief Knowledge?', *Analysis* 23(6): 121–123.
17. Grant, A. S. 2023. 'Will Human Potential Carry us Beyond Human? A Humanistic Inquiry into Transhumanism', *Journal of Humanistic Psychology* 63(1). DOI: 10.1177/0022167819832385
18. Gyory, J. T.; Soria Zurita, N. F.; Martin, J.; Balon, C.; McComb, C.; Kotovsky, K.; Cagan, J. 2022. 'Human Versus Artificial Intelligence: A Data-driven Approach to Real-time Process Management During Complex Engineering Design', *Journal of Mechanical Design* 144(2). DOI: 10.1115/1.4052488
19. Habermas, J. 1984. *The Theory of Communicative Action*. Polity.
20. Habicht, J.; Viswanathan, S.; Carrington, B.; Hauser, T. U.; Harper, R.; Rollwage, M. 2024. 'Closing the Accessibility Gap to Mental Health Treatment with a Personalized Self-referral Chatbot', *Nature Medicine* 30(2). DOI: 10.1038/s41591-023-02766-x
21. Haier, R. J. 2014. 'Increased Intelligence is a Myth (so far)', *Frontiers in Systems Neuroscience* 8(MAR). DOI: 10.3389/fnsys.2014.00034
22. Harvey, M. S. 2007. 'The Smaller Arachnid Orders: Diversity, Descriptions and Distributions from Linnaeus to the Present (1758 to 2007)', *Zootaxa* (1668). DOI: 10.11646/zootaxa.1668.1.19
23. Hintikka, J. 2007. *Socratic Epistemology: Explorations of Knowledge-seeking by Questioning*. DOI: 10.1017/CBO9780511619298
24. Jori, A. 2022. 'Aristotle's Philosophy of Mind', *Axiomathes* 32(6). DOI: 10.1007/s10516-019-09451-0
25. Katz, P. S. 2019. 'The Conservative Bias of Life Scientists', *Current Biology* 29(14). DOI: 10.1016/j.cub.2019.05.066
26. Korteling, J. E. (Hans); van de Boer-Visschedijk, G. C.; Blankendaal, R. A. M.; Boonekamp, R. C.; Eikelboom, A. R. 2021. 'Human-versus Artificial Intelligence', *Frontiers in Artificial Intelligence* 4. DOI: 10.3389/frai.2021.622364
27. Liu, R.; Jia, C.; Wei, J.; Xu, G.; Vosoughi, S. 2022. 'Quantifying and Alleviating Political Bias in Language Models', *Artificial Intelligence* 304. DOI: 10.1016/j.artint.2021.103654
28. Neshaei, S. P.; Davis, R. L.; Hazimeh, A.; Lazarevski, B.; Dillenbourg, P.; Käser, T. 2024. *Towards Modeling Learner Performance with Large Language Models*. Available at: http://arxiv.org/abs/2403.14661
29. Ongaro, G.; Hardman, D.; Deschenaux, I. 2024. 'Why the Extended Mind is Nothing Special but is Central', *Phenomenology and the Cognitive Sciences* 23(4). DOI: 10.1007/s11097-022-09827-5
30. Plato. 1997. *Plato: Complete Works*. Hackett Publishing Co.
31. Schneider, K. 2024. 'What Is Learning?', *Psychology* 15(05): 779–799. DOI: 10.4236/psych.2024.155047
32. Searle, J. 1984. *Minds, Brains and Science*. Harvard University Press.
33. Siemens, G.; Marmolejo-Ramos, F.; Gabriel, F.; Medeiros, K.; Marrone, R.; Joksimovic, S.; de Laat, M. 2022. 'Human and Artificial Cognition', *Computers and Education: Artificial Intelligence* 3. DOI: 10.1016/j.caeai.2022.100107
34. Soria Zurita, N. F.; Gyory, J. T.; Balon, C.; Martin, J.; Kotovsky, K.; Cagan, J.; McComb, C. 2022. 'Data on the Human Versus Artificial Intelligence Process Management Experiment', *Data in Brief* 41. DOI: 10.1016/j.dib.2022.107917

35. Stańczak, K.; Choudhury, S. R.; Pimentel, T.; Cotterell, R.; Augenstein, I. 2023. 'Quantifying Gender Bias Towards Politicians in Cross-lingual Language Models', *PLoS ONE* 18 (11 November). DOI: 10.1371/journal.pone.0277640
36. Stankov, L.; Lee, J. 2020. 'We Can Boost IQ: Revisiting Kvashchev's Experiment', *Journal of Intelligence* 8(4). DOI: 10.3390/jintelligence8040041
37. Tirri, K.; Nokelainen, P. 2011. 'Multiple Intelligences Profiling Questionnaire', in *Measuring Multiple Intelligences and Moral Sensitivities in Education*, 1–13. DOI: 10.1007/978-94-6091-758-5_1
38. Trepczyński, M. 2023. 'Religion, Theology, and Philosophical Skills of Llm–powered Chatbots', *Disputatio Philosophica* 25(1). DOI: 10.32701/dp.25.1.2
39. Vygotsky, L. 1980. 'Mastery of Memory and Thinking', in *Mind in Society*. Harvard University Press, 38–51. DOI: 10.2307/j.ctvjf9vz4.8
40. Winner, L. 1980. 'Do Artefacts Have Politics?', *Daedalus* 109(1): 121–136.
41. Wu, T.; Luo, L.; Li, Y.-F.; Pan, S.; Vu, T.-T.; Haffari, G. 2024. *Continual Learning for Large Language Models: A Survey*. Available at: https://arxiv.org/abs/2402.01364

AISTĖ DIRŽYTĖ

Didieji kalbos modeliai ir žmogaus pažinimo tobulinimas: kelios teorinės įžvalgos

Santrauka

Didžiųjų kalbos modelių (DKM) atsiradimas kelia filosofinių klausimų apie pažinimo ir žinojimo prigimtį bei prielaidas, todėl reikalauja nuoseklių tarpdisciplininių, taip pat ir filosofinių, tyrimų. DKM suteikia galimybių praplėsti pažinimo procesus, atliepdami „išplėstojo proto" tezės idėją – pagalbą paskirstant užduotis tarp žmogaus ir įrankio.

Šiame straipsnyje nagrinėjamas didžiųjų kalbos modelių galimas indėlis į žmogaus pažinimą. Nagrinėjamas klausimas, ar pažangios dirbtinio intelekto (DI) sistemos gali sustiprinti žmogaus mokymąsi ir mąstymą, o gal priešingai – sumažina mąstymo ir argumentavimo pastangas.

Ieškant atsakymo į šiuos klausimus, aptariamos žmogaus ir dirbtinio intelekto sąvokos, analizuojamos DKM galimybės ir ribos.

Straipsnyje teigiama, kad žmogaus intelektą galima apibrėžti kaip biologiškai determinuotą daugialypį gebėjimą spręsti problemas, o dirbtinį intelektą galima traktuoti kaip „kognityvinį pagalbininką", kuris veikia kaip pažintinių procesų plėtinys, galintis apdoroti milžiniškus informacijos kiekius ir tam tikrose srityse turintis potencialą pranokti žmogaus intelektą. Žmogaus intelektas yra lankstus, daugiausia nulemtas evoliucijos, biologijos ir socialinių sąveikų, o DI yra statiškas, jei nėra specialiai atnaujinamas, tačiau šių abiejų intelektų sąveika gali būti naudinga žmonijai.

DKM, kaip pažangios DI sistemos, kurios veikia giliųjų neuroninių tinklų principais, išsiskiria tinklo architektūrų sudėtingumu, dideliais duomenų kiekiais, plačiu apibendrinimo lygiu. Jie gali padėti atlikti daugybę kognityvinių užduočių, tačiau reikia daugiau tyrimų ir gilesnės filosofinės analizės norint suprasti, ar tokia DI pagalba skatina žmones labiau, o ne mažiau, lavinti mąstymą. Apibendrinant, didieji kalbos modeliai turi vertę žmogaus pažinimo tobulinimui tik esant griežtoms sąlygoms. Svarbu, kad šias teorines įžvalgas patvirtintų eksperimentiniai tyrimai.

Raktažodžiai: didieji kalbos modeliai (DKM), dirbtinis intelektas, žmogaus intelektas, kognicijos, „išplėstojo proto" teorija