UDC 165:004.8
DOI: https://doi.org/10.21564/2663-5704.68.349837

Lykhatskyi Andrii, PhD student at the Cultural Studies and Philosophy of Culture Department, Odesa Polytechnic National University, Ukraine
e-mail: bofh@bofh.od.ua
ORCID iD: 0009-0002-3735-2862

Chorna Lidiya, Doctor of Philosophical Sciences, Associate Professor, Professor at the Information Activities and Media Communications Department, Polytechnic National University, Odesa, Ukraine
e-mail: chorlidiya@gmail.com
ORCID iD: 0000-0001-8527-2416

AI AS COGNITIVE EXTENSION: RESHAPING HUMAN KNOWING

This study explores AI as a cognitive extension that integrates into human thinking, forming hybrid architectures with transformative potential for knowledge production. It identifies three key epistemic virtues for effective collaboration: critical prompting, algorithmic literacy, and epistemic discernment. Responsible use is essential to preserve human agency, avoid illusions of understanding, and prevent scientific monocultures. The work offers philosophical foundations for AI integration in education, science, and governance.

Keywords: *artificial intelligence, cognitive extension, extended mind, epistemology, epistemic virtues, philosophy of technology, human-AI collaboration, scientific discovery.*

Problem Statement. A radiologist reviews chest X-rays with an AI system highlighting potential anomalies. Neither the physician's medical training nor the algorithm's pattern recognition alone produces the diagnosis; instead, knowledge emerges from their collaboration, a phenomenon that challenges traditional assumptions about where thinking begins and ends.

This pattern repeats across professions. Climate researchers use machine learning to detect atmospheric patterns spanning decades, while software developers write code alongside AI assistants that complete their thoughts. These interactions represent more than mere assistance, signaling fundamental changes in how human cognition operates.

What happens when the tools we use to think become part of thinking itself? This article examines AI as **cognitive extension**, where artificial systems integrate into human cognitive processes rather than simply supporting them. Drawing on the extended mind hypothesis and building upon previous analysis of hybrid epistemology, we argue that AI systems become functional components of human cognitive architectures [1; 2].

This integration creates hybrid forms of knowing with significant augmentative potential, but only if used responsibly. Recent evidence reveals both opportunities and challenges: while AI extends analytical capabilities beyond biological limits, heavy reliance correlates with reduced critical thinking and risks creating illusions of understanding [3; 4]. To grasp these dynamics, we must move beyond simple human-versus-machine narratives and instead examine specific mechanisms of cognitive coupling while developing intellectual skills suited to hybrid cognition. We employ an interdisciplinary approach, combining philosophical analysis of cognitive extension theory with epistemological critique of AI-mediated knowledge production and examination of transformations in scientific practice. The analysis draws on contemporary philosophy of technology, philosophy of science and virtue epistemology to investigate both the ontological status of AI-enhanced cognition and its normative implications for intellectual practice.

Current Research Landscape. Otto keeps a notebook. He has Alzheimer's, so he writes down important information: addresses, phone numbers, appointment times. When he needs the Museum of Modern Art's location, he flips through pages until he finds it, while Inga, who doesn't have memory problems, simply recalls the address from her biological memory.

Clark and Chalmers posed a provocative question: What's the functional difference between Otto's external notebook and Inga's internal memory [1]? Both store information, both provide access when needed and both guide decision-making and action. If we accept that Inga's biological memory constitutes part of her cognitive process, why not Otto's notebook?

Their answer helped establish the **extended mind hypothesis**: cognitive processes can extend beyond skull boundaries to include external tools, but this requires meeting specific criteria. The external resource must be constantly available, generally reliable, easily accessible and previously endorsed by the user. Otto's notebook qualifies because he always carries it, trusts its contents, can quickly find information and consciously chose to record that information.

Modern AI systems satisfy these criteria with unprecedented sophistication. Consider how a pathologist now analyzes tissue samples: the AI doesn't just provide isolated suggestions but integrates into diagnostic reasoning, available through hospital interfaces around the clock, reliable across thousands of validated cases, offering immediate access to complex pattern recognition and endorsed through medical approval processes.

Moreover, AI extends beyond simple information storage. Large language models process queries, generate novel text and engage in multi-turn conversations; they don't just retrieve data but manipulate, combine and transform information in real-time. This points toward what cybernetics researchers call an **exocortex**: an artificial extension to biological cognition that provides additional thinking capabilities [5].

The term gained popularity through science fiction, particularly Charles Stross's novel *Accelerando* [6], but serious academic attention followed. Current implementations leverage familiar interfaces rather than futuristic brain implants, yet they demonstrate the concept's practical potential. Clark describes humans as «natural-born cyborgs,» constantly merging with technology to enhance cognitive capacity, and today's AI systems represent early exocortex implementations pointing toward futures where boundaries between internal and external thought may blur entirely [7].

Consider this progression: we began with simple tools where hammers extend physical reach and written language extends memory. Now AI systems augment our analytical capabilities, pattern recognition and creative generation, making the boundary between self and tool increasingly less obvious with each step.

Article Objectives. The main objective of this article is to extend the extended mind hypothesis to contemporary AI systems and, on this basis, to formulate a set of epistemic virtues necessary for responsible human-AI cognitive integration. Specifically, the study aims to: (1) provide philosophical foundations for understanding AI as a functional component of hybrid cognitive architectures; (2) identify and justify three key epistemic virtues (critical prompting, algorithmic literacy, epistemic discernment) that extend traditional intellectual virtues to address AI-specific challenges; (3) analyse both the augmentative potential and the risks of AI-mediated cognition (offloading, opacity, illusions of understanding, scientific monocultures); (4) offer normative implications for the integration of AI in education, scientific practice, and governance.

Presentation of the main material. The integration mentioned earlier enables **cognitive offloading**, the delegation of mental tasks to external systems that frees human resources for other functions. Programmers using AI assistants focus on architecture while algorithms handle implementation details, researchers employ AI to process literature in ways that enable synthesis and students use language models to clarify concepts in ways that accelerate learning.

Yet concerns about capacity erosion need examination. Consistent delegation might weaken core intellectual abilities through disuse while creating dependency that leaves us vulnerable when external resources fail. Historical anxieties about writing weakening memory or calculators eroding mathematical thinking proved largely unfounded, as these tools freed cognitive resources for higher-order tasks; however, AI represents qualitatively different offloading.

Previous tools delegated narrow functions, whereas current AI systems handle broad reasoning processes. Recent evidence reveals significant negative correlations between frequent AI usage and critical thinking scores (n=666 participants, 50 interviews), with cognitive offloading as the mediating mechanism [3]. Higher education buffers some effects, but younger users and heavy dependents show the greatest vulnerability.

While these findings remain primarily correlational rather than definitively causal, they underscore the need for proactive cultivation of intellectual virtues. Evidence increasingly suggests AI-mediated offloading may erode intellectual capacities differently than previous technologies, making compensatory skills not optional enhancement but necessary safeguard. The epistemic virtues we propose address this challenge directly.

Equally significant is AI opacity, where systems operate through mechanisms we cannot fully inspect [8]. Deep learning discovers statistical patterns across massive datasets rather than following human-traceable logic, creating **epistemic vulnerability**: the adoption of beliefs based on outputs we cannot thoroughly evaluate. Moreover, human and AI biases interact dynamically, with biased AI potentially amplifying human cognitive biases in feedback loops that traditional safeguards fail to address [9].

When medical AI recommends treatments, legal algorithms assess defendants or financial systems approve loans, opacity becomes ethically problematic. Distributed cognition in AI-supported environments introduces cognitive overload, loss of situational awareness and impaired coordination, all challenges requiring new intellectual capabilities beyond traditional critical thinking [10].

Developing Intellectual Skills for Hybrid Thinking. When AI becomes part of thinking rather than separate from it, traditional intellectual skills prove insufficient. Open-mindedness and critical thinking remain essential, yet hybrid cognition demands additional capabilities. Three skills become particularly important: critical prompting, algorithmic literacy and epistemic discernment [11, 12]. These extend traditional intellectual virtues (careful inquiry, intellectual humility, responsible belief formation) to address AI-specific challenges like algorithmic opacity, cognitive offloading and distributed thinking across human-machine systems.

The Art of Critical Prompting. «Tell me about environmental law.» This prompt generates generic overviews useful for basic orientation but insufficient for serious analysis. Compare it with: «Analyze three landmark cases from the International Court of Justice where environmental protection conflicted with economic development, focusing on how judges balanced competing claims and the precedents they established.»

The difference illustrates **critical prompting**: designing inquiries that guide AI systems toward reliable, relevant outputs. This skill goes beyond technical knowledge to include strategic thinking about information needs and communication clarity.

Effective prompting requires precision through unambiguous language that clearly defines scope and intent (vague requests produce vague responses). Contextual framing provides background information that helps AI understand complex requests appropriately, while constraint specification shapes outputs through explicit format requirements or analytical frameworks.

Perhaps most importantly, skilled prompting treats AI interaction as dialogue rather than simple query-response. Initial outputs inform refined questions, and unexpected responses suggest new inquiry directions. This iterative refinement resembles Socratic questioning, a method of drawing out knowledge through successive refinements that helps AI systems generate their most valuable contributions.

A legal researcher investigating climate litigation might begin broadly: «What are the main types of climate change lawsuits?» The AI's response reveals several categories: rights-based claims, procedural challenges and corporate liability suits. Each category suggests focused follow-up questions: «Explain how courts have interpreted the right to a healthy environment in climate cases, with specific examples from different jurisdictions.»

This approach prevents two common problems. First, it avoids information overload from receiving too much unfocused material to process effectively. Second, it reduces hallucination risk, as AI systems generate more accurate responses when questions include specific constraints and contextual guidance.

Understanding Algorithmic Logic. **Algorithmic literacy** doesn't require programming expertise, but it demands functional understanding of AI capabilities and limitations. When IBM Watson recommended cancer treatments that contradicted medical consensus, the problem wasn't technical malfunction but stemmed from misaligned training data and optimization criteria. Oncologists who understood these limitations could properly evaluate Watson's suggestions rather than blindly following them.

Large language models operate probabilistically, generating statistically likely word sequences rather than retrieving factual information. Such models excel at pattern recognition and text generation, but struggle with precise calculations or real-time information. Understanding this distinction helps users leverage AI strengths while compensating for weaknesses.

These systems inherit biases from training data, potentially amplifying societal prejudices in hiring algorithms, criminal justice assessments and medical diagnoses [13]. Recognition of bias sources enables critical evaluation rather than naive acceptance.

Different models excel in different domains. Creative writing models may lack factual accuracy, analytical models might struggle with nuanced communication and specialized models outperform general-purpose ones in specific tasks. Algorithmic literacy includes knowing which tools work best for particular purposes.

Consider a journalist researching climate impacts. A general language model might provide compelling statistics about rising sea levels complete with authoritative-sounding explanations. Algorithmic literacy suggests verification strategies: cross-referencing claims with authoritative sources (like IPCC reports), checking for internal consistency in the AI's reasoning, identifying potential biases in how climate data gets presented.

This knowledge enables calibrating trust appropriately: neither blind faith nor cynical rejection, but informed confidence based on understanding AI's actual capabilities.

Epistemic discernment integrates critical prompting and algorithmic literacy into practical evaluation skills. When AI generates information, how do we assess its reliability and value?

Source verification provides the foundation. AI-generated claims require systematic checking against established authorities – not to dismiss AI outputs, but to treat them as hypotheses requiring confirmation rather than established facts.

Bias detection involves recognizing how training data prejudices might shape outputs in subtle ways. A model trained primarily on Western academic literature might have cultural blind spots when discussing global issues, while historical biases in medical research might influence AI recommendations about treatment effectiveness across demographic groups.

Coherence checking evaluates whether AI responses maintain logical consistency throughout complex arguments. Does the reasoning flow logically from premises to conclusions? Do different parts of a response contradict each other? Are claims supported by appropriate evidence?

Perhaps most importantly, epistemic discernment maintains intellectual humility, a healthy skepticism about both AI capabilities and one's own judgment. Neither humans nor machines possess perfect knowledge, and effective collaboration requires recognizing the limitations of both biological and artificial cognition.

These three virtues work synergistically. Skilled prompting elicits higher-quality AI outputs, algorithmic literacy calibrates appropriate trust levels and epistemic discernment ensures responsible evaluation of results. Together, they can enable cognitive extension that may genuinely enhance human thinking rather than replacing it with algorithmic dependency. Table 1 summarizes these skills with practical examples.

Table 1

Essential intellectual skills for AI-enhanced cognition*

Skill	Definition	Application Example	Problem Addressed
Critical Prompting	Designing precise, contextualized queries that guide AI toward useful outputs	Refining «tell me about AI» to «explain transformer architecture in LLMs for non-technical audiences, focusing on attention mechanisms»	Vague, irrelevant or shallow responses
Algorithmic Literacy	Understanding AI capabilities, limitations, biases without technical expertise	Recognizing that an LLM might generate plausible but fake citations and verifying them independently	Black box opacity; misplaced trust
Epistemic Discernment	Critically evaluating AI-generated content for reliability and value	Cross-referencing AI climate statistics with IPCC reports while checking for presentation bias	Misinformation; uncritical acceptance

** Source: created by author*

AI's Transformation of Scientific Practice. Individual cognitive extension represents just one dimension of change; AI also reshapes collective scientific practice, accelerating discovery while raising fundamental questions about explanation and understanding.

From Theory to Pattern. Traditional science follows familiar rhythms: researchers observe phenomena, develop hypotheses, design experiments, collect data and analyze results. Human insight drives each stage. Intuition suggests which questions matter, creativity designs revealing experiments and interpretation makes sense of findings.

AI introduces a different approach entirely. Instead of starting with human theories, machine learning systems analyze massive datasets to identify patterns that would escape human notice, and these patterns then suggest hypotheses for experimental testing.

AlphaFold exemplifies this methodological shift [14]. Rather than beginning with biological theories about protein folding, the system analyzed structural relationships across thousands of known proteins and discovered patterns between amino acid sequences and three-dimensional shapes that solved a fifty-year-old problem in computational biology. The breakthrough emerged from computational pattern recognition rather than theoretical insight.

Similar transformations appear across scientific domains. Materials scientists use AI to predict novel compound properties before synthesis, screening millions of possibilities rather than following chemical intuition. Climate researchers employ machine learning to identify atmospheric patterns in decades of observational data, discovering relationships too complex for traditional analysis. Drug discovery screens molecular interactions at unprecedented scales, identifying therapeutic candidates through statistical correlation rather than mechanistic understanding.

Each case inverts traditional methodology: data patterns generate hypotheses rather than hypotheses being tested against data. This doesn't eliminate human involvement but changes its character, shifting from initial theorizing to subsequent interpretation and validation. But what does this mean for scientific knowledge? Is pattern recognition sufficient for understanding? Can we trust predictions without explanatory mechanisms? How do we validate findings when the discovery process exceeds human comprehension?

The Challenge of Explanation. AI's predictive power often comes at a cost: comprehensibility. When deep learning models identify promising drug candidates, they typically cannot explain why these molecules might work. The system recognizes statistical patterns across massive chemical databases without providing causal explanations that humans can evaluate.

This creates what researchers term an **explanatory gap** [15]. We may know *that* a particular treatment shows promise because the AI predicts success, but we may not understand *why* in terms of underlying biological mechanisms. The gap challenges traditional assumptions about scientific knowledge, which emphasized not just prediction but causal understanding.

Consider cancer research. AI systems can analyze thousands of tumor samples to identify patients likely to respond to specific treatments. These predictions often prove accurate, enabling personalized therapy that saves lives. Yet the systems rarely explain their reasoning in terms of molecular pathways or cellular mechanisms that oncologists can verify independently.

This introduces what some philosophers call «performative science,» knowledge validated primarily through successful outcomes rather than theoretical understanding. The approach works pragmatically but challenges long-held ideals about scientific explanation.

Recent analysis reveals deeper risks beyond simple gaps in causal knowledge. Research demonstrates that AI tools risk creating **illusions of understanding** in scientific research, where predictive success leads scientists to overestimate their explanatory knowledge [4]. When AI systems identify promising patterns or generate plausible hypotheses, researchers may experience subjective feelings of comprehension without genuine mechanistic insight. These illusions can obscure the formation of **scientific monocultures** – environments where certain methods, questions and viewpoints dominate, making science less innovative and more vulnerable to systematic errors.

Furthermore, AI-driven pattern recognition may induce what some term «illusions of pursuitworthiness,» directing research toward phenomena that appear promising computationally but lack true explanatory depth. In complex domains like molecular biology, the explanatory gap may prove permanent rather than bridgeable, requiring fundamental reconceptualization of what constitutes scientific explanation in an age of machine learning.

Different research communities respond to these challenges differently. Some embrace predictive power regardless of explanation, arguing that successful treatments matter more than theoretical understanding. Others insist that genuine scientific knowledge requires comprehensible causal mechanisms. Many seek middle ground, using AI predictions to guide research while working to develop explanatory frameworks after the fact, yet this middle path becomes precarious when illusions of understanding prevent recognition of explanatory deficits.

Bridging explanatory gaps represents a new frontier for scientific inquiry, though one filled with risk. It requires researchers to develop novel theoretical frameworks that make sense of AI-driven discoveries while remaining vigilant against false confidence. Sometimes the AI's «opacity» reflects reality's genuine complexity, patterns too intricate for existing theories to accommodate. The challenge becomes expanding human understanding to match machine recognition capabilities while avoiding monocultures that favor computational tractability over genuine explanation.

This dynamic creates intriguing feedback loops: AI identifies patterns that suggest new theoretical directions, human researchers develop explanatory frameworks to account for these patterns and improved theories guide better AI training and interpretation. The process potentially accelerates both discovery and understanding, though the pace of pattern recognition often outstrips explanatory development – risking the illusions documented by recent empirical research.

Conclusion. AI transforms both scientific and everyday thinking, demanding new intellectual virtues. This article's central contribution extends virtue epistemology to hybrid cognition, identifying capabilities essential for AI-augmented knowledge production.

Critical prompting, algorithmic literacy and epistemic discernment extend traditional intellectual virtues (curiosity, open-mindedness, intellectual humility) to meet the specific challenges of human-AI collaboration. Classical virtue epistemology focuses on individual thinking; these hybrid virtues address thinking distributed across biological minds and artificial systems. In such extended systems, virtue includes not just internal dispositions but also practices for engaging external computational resources responsibly.

Recent empirical evidence validates this virtue-based approach. As AI systems grow more sophisticated and persuasive, technical safeguards alone prove insufficient [3, 4]. Individuals must develop intellectual character traits enabling critical engagement with machine outputs while resisting both uncritical acceptance and reflexive rejection. These virtues represent necessary conditions for genuine knowledge in the age of cognitive hybridization, not optional enhancements.

Toward a Hermeneutics of Opacity. Building on the recognition of essential epistemic opacity, where the path from input to output in deep learning systems remains fundamentally inaccessible even in principle, we confront a deeper interpretive challenge [8, 16]. Rather than pursuing ever-greater transparency through post-hoc explainability techniques that often trade off accuracy, an alternative stance becomes necessary: one that treats opaque AI outputs as phenomena amenable to **hermeneutic engagement** without reduction to underlying mechanisms.

We propose the concept of a **hermeneutics of opacity** – a distinctive interpretive practice that learns to understand model conclusions relationally. This happens through iterative attention to surface patterns, distributed representations, emergent behaviors and statistical textures, even as the generative logic remains irreducible to humanly traceable steps. This approach draws inspiration from hermeneutic traditions that grapple with texts whose full meaning exceeds authorial intention or explicit reconstruction, but adapts them to the specific epistemic texture of large-scale neural architectures [17].

In practice, such a hermeneutics might involve tracking how outputs shift across prompt variations, mapping conceptual clusters in latent space via probing techniques, attending to analogical resonances between model behavior and domain phenomena and cultivating sensitivity to the model's «style» of reasoning as an emergent signature rather than a defect. Far from passive acceptance of black-box verdicts, this represents an active and critical mode of sense-making that preserves human agency while acknowledging the structural limits of mechanistic explanation.

This proposal is preliminary and will be developed more fully in forthcoming work. For the present analysis, it underscores that epistemic discernment in hybrid cognition must extend beyond verification and bias detection to include hermeneutic competence: the cultivated ability to read opaque systems productively without demanding their full demystification.

Implications and Future Directions. Understanding AI as cognitive extension provides a framework for responsible navigation of this transformation. Rather than viewing AI as external threat or mere productivity tool, this perspective recognizes it as a new medium for thought itself, one that reshapes thinking processes rather than simply assisting existing ones.

Human-AI integration creates hybrid cognitive architectures with significant augmentative potential. Pattern recognition operates at scales exceeding biological cognition, information processing accelerates beyond unaided limits and creative generation combines human insight with algorithmic exploration. These capabilities represent genuine expansion of intellectual capacity – contingent on cultivating appropriate epistemic virtues.

Successfully navigating hybrid cognition requires the intellectual capabilities we propose. Critical prompting enables effective AI collaboration, algorithmic literacy provides functional understanding of system capabilities and limitations and epistemic discernment integrates these into evaluation frameworks that maintain human agency while leveraging machine capabilities. Together, these skills harness pattern-recognition power while preserving critical oversight, helping ensure human judgment remains the ultimate arbiter of knowledge.

Scientific practice illustrates both promise and challenges. AI identifies patterns at unprecedented scales and generates hypotheses across domains, accelerating discovery. Yet explanatory gaps persist: we often know *what* works without understanding *why*. Bridging these gaps requires developing theoretical frameworks that make sense of AI-driven discoveries while cultivating hermeneutic competence for productive engagement with opacity.

Several questions demand continued attention. How can educational systems cultivate AI-specific intellectual skills across populations? What governance frameworks ensure responsible cognitive extension while preserving innovation? How might hybrid intelligence serve human flourishing? These questions lack simple answers but demand serious engagement from researchers, educators, policymakers and citizens alike.

The co-evolution of human intelligence with artificial offers unprecedented opportunities for expanding our collective capacity to understand the world and improve it. Medical diagnoses become more accurate. Scientific discovery accelerates. Creative expression explores new territories, and educational access expands. Success depends on thoughtful integration that maintains what makes human thinking valuable while leveraging machine capabilities.

The framework of cognitive extension provides philosophical foundations for navigating this transformation – benefiting educational institutions, technology designers and policymakers alike. The choices we make about cognitive extension today will shape human thinking for generations. Neither uncritical embrace nor reflexive rejection serves us well; instead, we need approaches that recognize potential and acknowledge challenges, guided by epistemic virtues adequate to the hybrid age.

REFERENCES

1. Clark, A., Chalmers, D. (1998). The extended mind. *Analysis, 58 (1), 7–19*. doi: https://doi.org/10.1093/analys/58.1.7
2. Lykhatskyi, A. (2025). Hybrid epistemology: Emergent knowledge forms in the age of human-AI cognitive integration. *The Bulletin of Yaroslav Mudryi National Law University. Series: Philosophy, philosophies of law, political science, sociology, 3 (66)*, *148–158.* doi: 10.21564/2663–5704.66.337968
3. Gerlich, M. (2025). AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. *Societies, 15(1), 6*. doi: https://doi.org/10.3390/soc15010006
4. Messeri, L., Crockett, M. J. (2024). Artificial intelligence and illusions of understanding in scientific research. *Nature, 627, 49–58*. doi: https://doi.org/10.1038/s41586-024-07146-0
5. Bonaci, T., Herron, J., Matlack, C., Chizeck, H. J. (2014). Securing the exocortex: A twenty-first century cybernetics challenge. *Conference on Norbert Wiener in the 21st Century (21CW)*. IEEE, *1–8.* doi: https://doi.org/10.1109/NORBERT.2014.6893912
6. Stross, C. (2005). Accelerando. London: Ace Books.
7. Clark, A. (2003). Natural-born cyborgs: Minds, technologies, and the future of human intelligence. Oxford: Oxford University Press.
8. Burrell, J. (2016). How the machine ’thinks’: Understanding opacity in machine learning algorithms. *Big Data & Society, 3(1), 1–12*. doi: https://doi.org/10.1177/2053951715622512
9. Felten von, N. (2025). Beyond isolation: Towards an interactionist perspective on human cognitive bias and AI bias. arXiv. doi: https://doi.org/0.48550/arXiv.2504.18759
10. Jacobsen, R. M., Wester, J., Djernæs, H. B., Berkel van, N. (2025). Distributed cognition for AI-supported remote operations: Challenges and research directions. arXiv. doi: https://doi.org/10.48550/arXiv.2504.14996
11. Baehr, J. (2011). The inquiring mind: On intellectual virtues and virtue epistemology. Oxford: Oxford University Press.
12. Lim, C. (2025). DeBiasMe: De-biasing human-AI interactions with metacognitive AIED interventions. arXiv: doi: https://doi.org/10.48550/arXiv.2504.16770

13. Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. New York: NYU Press.
14. Jumpe, J., Evans, R., Pritzel, A., Green, T. et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature, 596, 7873, 583–589.* doi: https://doi.org/10.1038/s41586-021-03819-2
15. Sullivan, E. (2022). Understanding from machine learning models. *The British Journal for the Philosophy of Science, 73(1), 109–133*. doi: https://doi.org/10.1093/bjps/axz035
16. Humphreys, P. (2009). The philosophical novelty of computer simulation methods. *Synthese*. *169(3)*. *Springer, 615–626*. doi: https://doi.org/10.1007/s11229-008-9435-2
17. Gadamer, H.-G. (2004). Truth and method. London; New-York: Continuum.

Лихацький Андрій Васильович, аспірант кафедри культурології та філософії культури, Національний університет «Одеська політехніка», Україна

Чорна Лідія Валеріївна, доктор філософських наук, доцент, професор кафедри інформаційної діяльності та медіакомунікацій, Національний університет «Одеська політехніка», Україна

ШТУЧНИЙ ІНТЕЛЕКТ ЯК КОГНІТИВНЕ РОЗШИРЕННЯ: ПЕРЕОСМИСЛЕННЯ ЛЮДСЬКОГО ЗНАННЯ

Ця стаття розглядає штучний інтелект як когнітивне розширення, що інтегрується в людське мислення, створюючи гібридні архітектури з великим потенціалом для виробництва знань. Визначено три ключові інтелектуальні чесноти для ефективної співпраці: критичне промотування, алгоритмічна грамотність та епістемічна розбірливість. Відповідальне використання необхідне, щоб зберегти людську агентність, уникнути ілюзії розуміння та наукових монокультур. Робота пропонує філософські засади відповідальної інтеграції ШІ в освіту, науку та управління технологіями.

Ключові слова: *штучний інтелект, когнітивне розширення, розширений розум, епістемологія, епістемічні чесноти, філософія техніки, співпраця людини та ШІ.*