UDC: 165:004.8:316.4
DOI: https://doi.org/https://doi.org/10.21564/2663-5704.66.337968

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

HYBRID EPISTEMOLOGY: EMERGENT KNOWLEDGE FORMS IN THE AGE OF HUMAN-AI COGNITIVE INTEGRATION

This article introduces hybrid epistemology as a framework for understanding how knowledge emerges when humans and AI systems function as cognitive partners. Traditional philosophy treated knowledge as something individual minds acquire through experience or reasoning. That view no longer captures contemporary reality. The investigation examines three critical challenges. How do we evaluate knowledge claims when the reasoning process involves opaque algorithms? What happens when trust networks replace traditional methods of justification? How does cognitive responsibility get distributed across human-machine partnerships? These questions matter because hybrid knowledge systems now influence medical diagnoses, legal decisions, and policy choices that directly affect people's lives.

Keywords: *artificial intelligence, epistemology, distributed cognition, extended mind, epistemic opacity, cognitive agency, human-machine interaction*

Problem Statement. Imagine a normal morning in a contemporary medical facility. Alongside an AI diagnosis system that has examined thousands of identical photos, the doctor examines X-rays. A small pattern that he might have overlooked is flagged by the system. His background in clinical practice aids in placing the discovery in the perspective of the patient's larger medical background. The knowledge comes from their cooperation; neither the doctor nor the AI system could have arrived at the same diagnostic conclusion on their own.

This situation exemplifies a major change in the process of knowledge creation. Conventional philosophy made the assumption that people's minds functioned independently to comprehend the universe. This idea – that knowledge belonged to individual thinking beings who could clearly distinguish their own thoughts from external reality – was embodied in Descartes' well-known «I think, therefore I am» statement. The knowledge practices of today paint a different picture.

The problem goes beyond simple tool use. When the doctor uses a stethoscope, he remains the primary knower – the instrument merely amplifies what he can perceive. AI systems don't just amplify human perception. They actively contribute to the development of novel understandings. The diagnostic algorithm does more than just process the doctor's data. Rather, it finds patterns in medical images that are beyond human perception, producing discoveries that change the way doctors think about illness.

Why is this important? Because AI systems now have a say in important societal decisions. University admissions, criminal penalties, and loan approvals are all influenced by algorithms. Public health interventions, urban planning, and investment initiatives are all guided by predictive models. Machine learning is being used more and more in scientific research to find patterns in complicated datasets, such as those used in climate modeling and genomics. As knowledge arises from human-AI collaborations, issues of justice, accountability, and dependability cease to be theoretical conundrums and instead become pressing practical issues.

Current Research Landscape. Although researchers have taken a variety of approaches to the relationship between AI and knowledge, the body of extant research shows notable gaps. The foundation for extended mind theory was established by Clark and Chalmers. According to their hypothesis, cognitive processes frequently utilize tools and structures from the environment in addition to those found inside the brain [1]. Consider how we use GPS to drive or rely on smartphones to recall phone numbers – these devices become a part of our cognitive machinery. However, Clark and Chalmers concentrated on passive instruments like calculators and notebooks. AI systems pose a distinct difficulty. Instead of merely supporting human thought, machine learning algorithms actively provide insights that transform entire domains of knowledge, such as when they identify previously undiscovered astrophysical events or new medicine molecules.

Distributed cognition research looked at how teams address complicated problems collaboratively. Hutchins demonstrated that navigation aboard navy ships requires complicated synchronization of persons, instruments, and procedures [2]. No single person has all of the knowledge required for successful navigation. However, studies on distributed cognition have not effectively addressed AI systems, which operate as qualitatively distinct cognitive partners similar to human teammates.

The focus of science and technology studies has been on how technologies actively influence the creation of knowledge. According to Latour's actor-network theory, non-human actors – from written texts to lab equipment – contribute to the creation of scientific knowledge [3]. To handle AI systems that can surpass human cognitive abilities in particular domains, these frameworks must be updated. Deeper epistemological issues were overlooked in recent work on XAI, which addressed algorithmic opacity's technical issues. What does it signify for our understanding of knowledge itself when an AI system generates accurate predictions using methods that are difficult for humans to understand? Technical solutions alone are insufficient to address these philosophical issues.

Fairness and accountability have dominated discussions on AI ethics, but no formal frameworks for comprehending AI-mediated knowledge have been developed. Comparably, virtue epistemology has started investigating the application of conventional intellectual qualities to technological settings; however, this research needs to be expanded to include hybrid human-AI systems, where cognitive work is divided among various agent kinds.

Article Objectives. As a paradigm for comprehending knowledge that arises from human-AI cognitive partnerships, this approach develops hybrid epistemology. The investigation is guided by three goals. The study first explains how modern AI systems operate as cognitive agents as opposed to passive instruments. This discovery calls into question long-held philosophical beliefs regarding knowledge as a personal accomplishment. Second, the approach looks at issues with epistemic opacity, or circumstances in which precise knowledge is produced by computing processes that are difficult for humans to understand. Third, the study investigates the distribution of cognitive responsibility in human-machine assemblages.

The method is theoretical-conceptual and incorporates concepts from cognitive science, science and technology studies, and philosophy of knowledge. This work builds conceptual underpinnings for comprehending epistemic shifts in technology-mediated contexts rather than providing empirical facts.

Presentation of the main material. Clark and Chalmers changed philosophical thinking about cognition with their «Extended Mind» thesis. They argued that cognitive processes routinely extend beyond brain boundaries to incorporate environmental structures and tools. Your smartphone storing contacts? That's part of your memory system. GPS guidance? A part of your spatial awareness. These tools become intrinsic elements of cognitive processes rather than only aiding in thought. However, AI systems are more advanced than the passive calculators and notebooks used in Clark and Chalmers' initial examples.

Today's AI agents act as the cognitively active partners, so they are able to adapt, generate hypotheses, and learn on their own. Instead of using tools in a static manner, this active quality fosters dynamic relationships. Think about how radiologists currently use diagnostic algorithms: the AI actively finds patterns that alter clinical understanding rather than merely amplifying human awareness. Consider DeepMind's AlphaFold technology as a concrete example. Instead of just processing data provided by human scientists, AlphaFold learns to anticipate protein structures by studying evolutionary linkages and physical restrictions in ways that are beyond human comprehension [4]. The generated predictions required validation using experimental approaches. What developed was a knowledge-production cycle that combined machine learning, human expertise, and laboratory verification to produce an integrated epistemic system.

This shift calls into question long-held epistemological beliefs about where knowledge resides. In hybrid contexts, knowledge is not contained in isolated human minds, but in interactive systems that include humans, AI, data, interfaces, and interpretative frameworks. The locus of knowing shifts from the person to the integrated system, and knowledge becomes an emergent feature that transcends its basic aspects.

The transition from classical to hybrid epistemology involves fundamental shifts across multiple dimensions (table 1). This framework demonstrates that hybrid epistemology does more than just tweak traditional methodologies, it adds fundamentally distinct organizing ideas. Perhaps most importantly, the fundamental unit of epistemic analysis shifts from the individual to the system.

Table 1

Key Differences Between Classical and Hybrid Epistemology*

Dimension	Classical Epistemology	Hybrid Epistemology
Knowing Subject	Individual, autonomous human agent	Distributed human-AI cognitive system
Knowledge Location	Internal mental states/beliefs	Network of interactions and processes
Justification	Transparent reasoning chains	Institutional trust and validation networks
Truth Conditions	Correspondence to objective reality	Emergent properties of system dynamics
Epistemic Authority	Individual expertise and credentials	Collective human-AI performance
Temporal Structure	Discrete moments of knowing	Continuous adaptive learning
Opacity Tolerance	Transparency required for justification	Functional opacity accepted with validation

** Source: created by author.*

Here's a major issue: many AI systems are «black boxes» whose underlying functions are difficult to explain [5; 6]. A deep learning program may correctly identify skin cancer from images, but explaining how it reached that result is practically hard. The system processes thousands of features and interactions that humans cannot understand. This raises an epistemic challenge. Traditional concepts of justified knowledge frequently assume that knowers can articulate the reasons for their beliefs. But what happens when real knowledge comes from computational processes that are beyond human comprehension?

The problem has inspired the creation of alternate validation strategies. Rather than requiring explicit reasoning, «epistemic trust» and «institutional justification» authenticate information through belief in sources, institutions, or procedures [7, 8]. Justification increasingly depends on complex networks of trust relationships rather than individual comprehension of evidence. Consider high-stakes domains where this matters most. When AI systems influence medical diagnoses, legal judgments, or defense decisions, questions of responsibility become simultaneously ethical and political [9; 10]. Who bears responsibility for knowledge created by hybrid systems? The human expert who depends on algorithmic insights? Who designed the system? The institutions that use it?

The opacity problem goes beyond technological considerations. As proprietary AI systems increasingly mediate knowledge production, epistemic fairness problems arise, such as who has access to these systems and whose interests they are serving [11; 12]. Efforts to reduce opacity through explainable AI (XAI) have produced significant technological advancements, but they frequently overlook the underlying philosophical point [13]. Although XAI techniques can offer post-hoc explanations for individual actions, they rarely make the underlying computational processes apparent in ways that meet traditional epistemic justification standards. The basic tension remains: some types of correct information may need accepting functional opacity.

The temporal organization of knowledge generation in hybrid systems deviates significantly from standard epistemological theories. Traditional understandings of knowledge frequently assumed distinct moments of insight or discovery, followed by long periods of settled conviction. A scientist would make a discovery, publish results, and the knowledge would remain relatively stable until new evidence emerged. This model reflected the human timescale of cognition – deliberate, sequential, bounded by biological constraints.

Knowledge in hybrid systems demonstrates temporal stratification, which refers to layered temporal processes in which human deliberation, computational processing, and system adaptability all occur on separate timelines. Consider how this emerges in practice: an AI system can process millions of similar patterns in minutes, whereas a human expert may spend weeks analysing a complex dataset. However, years of professional experience help humans improve their contextual understanding and interpretive framework. These various temporal rhythms provide knowledge that exists concurrently across several time horizons.

The stratification is most visible in fields such as financial trading, where algorithmic computers make thousands of decisions per second based on pattern recognition, while human oversight functions on minutes-to-hours timelines and regulatory frameworks change over months or years. Knowledge claims in such systems cannot be evaluated using conventional epistemological frameworks that presuppose synchronous reasoning processes.

AI systems create temporal complications by allowing for continuous learning and adaptability. While human experts can update their expertise through periodic contact with new literature or experience, AI systems can assimilate new data on a continual basis, resulting in incremental modifications in decision patterns that may not be immediately evident to human partners [14]. This raises questions about epistemic responsibility: when knowledge arises from systems that change over time, establishing which version of the system produced specific insights becomes critical for validation and accountability. Furthermore, hybrid knowledge systems demonstrate retroactive epistemic adjustment – the ability to change previous findings in response to new patterns discovered through extended data analysis.

Traditional virtue epistemology emphasized individual qualities such as intellectual humility, curiosity, and critical thinking [15]. Hybrid epistemology requires perceiving these virtues as distributed properties of human-AI systems rather than individual characteristics.

Calibrated trust emerges as a fundamental virtue – the capacity to discern when to rely on AI systems and when to prioritize human judgment [16]. This differs markedly from blind faith or wholesale skepticism toward automated systems. Calibrated trust requires developing sophisticated meta-cognitive awareness of the boundaries and capabilities of both human and artificial cognition. A radiologist exercising calibrated trust might recognize that AI systems excel at detecting subtle pattern variations in medical images that human eyes might miss, while simultaneously understanding that contextual factors – patient history, unusual presentation, rare conditions – require human interpretive expertise. This entails becoming aware of the limitations of both human and artificial cognition, as well as comprehending how their combination might compensate for individual inadequacies. The development of calibrated trust involves practical experience with system failures and successes. Professionals must learn to recognize when algorithmic confidence scores correlate with accuracy, when edge cases might confound automated analysis, and how to integrate computational insights with domain expertise. This learning process cannot be reduced to simple rules but emerges through repeated interaction with hybrid systems across varied contexts.

Interpretive flexibility is another essential virtue – the capacity to switch between multiple modes of explanation and reasoning as required by different contexts and audiences. Practitioners working with AI systems must be able to switch between computational and human-centered modes of reasoning without losing critical insights from either area.

Beyond modifying methods of information acquisition, AI systems significantly change the configuration of subjectivity. Human subjects who use AI for cognitive support are no longer independent sources of understanding, but rather functions within collective cognition – nodes in an extended network that combines human experience, technology computing, and social circumstances to produce an integrated cognitive system.

This transformation is consistent with Marx's idea of «species-being» (Gattungswesen) and Heidegger's concept of «being-in-the-world» (In-der-Welt-sein), which both emphasize that individual consciousness is always enmeshed in networks of meaning and interpretation [17; 18]. Don Ihde's postphenomenology takes this theory to technological mediation, illustrating how human-technology relationships fundamentally shape perceptual and cognitive experience rather than just supplementing it [19; 20].

The hybrid subject is formed as a result of cognitive delegation and integration. As people increasingly delegate cognitive processes to AI systems, ranging from memory and arithmetic to complex pattern recognition and decision guiding, the functional and phenomenological distinctions between human and machine cognition become less clear [21]. This process occurs gradually and often imperceptibly. Consider how GPS navigation has affected spatial cognition. Drivers are increasingly relying on algorithmic route optimisation instead of creating mental spatial maps of their surroundings. The cognitive function of navigation is split between human decision-making and automated pathfinding. Over time, this delegation alters not only how people travel, but also how they perceive and comprehend spatial relationships. The distinction between human spatial awareness and algorithmic help becomes increasingly blurred. Similar similarities appear across the professional fields.. Financial analysts integrate algorithmic trading insights into their market understanding. Scientists use machine learning pattern spotting in their experimental reasoning. Doctors use diagnostic AI recommendations alongside clinical judgement. In each scenario, delegation entails more than just job automation; it signifies a rearrangement of cognitive architecture in which human expertise and computing capabilities are interconnected components of knowledge generation.

This interdependence raises fundamental questions about cognitive autonomy. When professional expertise becomes inseparable from algorithmic mediation, what constitutes independent human knowledge? The hybrid subject does not simply use AI tools but exists in constant cognitive collaboration with artificial systems, creating new forms of agency that transcend traditional human-machine distinctions.

The emergence of the hybrid subject transforms traditional notions of epistemic autonomy and agency. When knowledge emerges from human-AI partnerships, questions of intellectual ownership, originality, and authority become complex [22, 23]. Who «knows» when a radiologist and an AI system collaboratively diagnose a medical condition? These questions cannot be resolved through simple attribution to either human or machine but require reconceptualizing agency as distributed across the hybrid system.

Furthermore, the formation of hybrid subjects raises important questions about human cognitive development and education in AI-saturated environments. As young people increasingly develop their cognitive capacities in constant interaction with AI systems, their intellectual formation occurs through fundamentally different processes than those of previous generations [24; 25].

Conclusions. This investigation developed hybrid epistemology as a framework for understanding knowledge production when humans and AI systems work as cognitive partners. Three key findings were revealed by the analysis.

Contemporary AI systems act as cognitive actors rather than passive instruments. This conclusion calls into question classical philosophical assumptions about knowledge as an individual human endeavor. When AI systems actively contribute to the generation of insights, the old concept of the autonomous knower becomes ineffective. Philosophical frameworks must now support distributed cognition between human and machine systems. The fundamental unit of epistemic analysis moves from the individual to the integrated system.

Second, epistemic opacity introduces additional problems to knowledge validation. Many AI systems employ computational techniques that are beyond human comprehension. Traditional approaches to epistemic justification assumed that knowers could express the rationales for their views. When this becomes impossible, other validation methods develop. Institutional trust networks and validation protocols enable new forms of epistemic justification. This transition is particularly important in sectors where knowledge has a direct impact on human welfare, raising new challenges to accountability and governance.

Third, hybrid knowledge systems are temporally divided. Human thought and computer processing take various time frames. AI systems can constantly assimilate new data, whereas human specialists must update their understanding through repeated exposure to new material. This raises epistemic responsibility concerns: when knowledge is drawn from systems that change over time, knowing which version of the system produced specific insights is critical for validity and accountability.

These findings indicate particular research directions. Empirical research might examine how hybrid epistemology evolves in practice by undertaking ethnographic investigations into human-AI interactions in certain sectors. Philosophical study might lead to the creation of new epistemic justification theories that support both transparency and trust-based validation.

Rather than treating AI as either enhancement or threat, hybrid epistemology suggests a more nuanced approach. Collaboration between humans and machines leads to the emergence of new types of understanding. Successfully managing this transition involves both technological competence and philosophical analysis of how knowledge is evolving. The paradigm provided here establishes conceptual foundations for comprehending these shifts while also leading to future theoretical and empirical research.

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. Hutchins, E. (1995). Cognition in the wild. MIT Press.
3. Latour, B. (1987). Science in action: How to follow scientists and engineers through society. Harvard University Press.
4. Jumper, J., Evans, R., Pritzel, A., 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
5. 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
6. Humphreys, P. (2009). The philosophical novelty of computer simulation methods. *Synthese*, 169(3), 615–626. URL: https://doi.org/10.1007/s11229-008-9435-2
7. Goldman, A. I. (2001). Experts: Which ones should you trust? *Philosophy and Phenomenological Research*, *63(1), 85–110.* URL: https://doi.org/10.1111/j.1933–1592.2001.tb00093.x
8. Hardwig, J. (1991). The role of trust in knowledge. *The Journal of Philosophy*, *88(12), 693–708.* URL: https://doi.org/10.2307/2027007
9. Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*, *6(3), 175–183.* doi: https://doi.org/10.1007/s10676-004-3422-1
10. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society, 3(2), 1–21.* doi: https://doi.org/10.1177/2053951716679679
11. Fricker, M. (2007). Epistemic injustice: Power and the ethics of knowing. Oxford University Press. doi: https://doi.org/10.1093/acprof:oso/9780198237907.001.0001
12. O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Crown.

13. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1(5), 206–215.* doi: https://doi.org/10.1038/s42256-019-0048-x
14. Russell, S. (2019). Human compatible: Artificial intelligence and the problem of control. Viking.
15. Zagzebski, L. T. (1996). Virtues of the mind: An inquiry into the nature of virtue and the ethical foundations of knowledge. Cambridge University Press.
16. Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, *57(3), 407–434.* doi: https://doi.org/10.1177/0018720814547570
17. Marx, K. (1988). Economic and philosophic manuscripts of 1844 (M. Milligan, Trans.). Prometheus Books.
18. Heidegger, M. (1962). Being and time (J. Macquarrie & E. Robinson, Trans.). Blackwell.
19. Ihde, D. (1990). Technology and the lifeworld: From garden to earth. Indiana University Press.
20. Ihde, D. (2009). Postphenomenology and technoscience: The Peking University lectures. SUNY Press.
21. Hayles, N. K. (2017). Unthought: The power of the cognitive nonconscious. University of Chicago Press. doi: https://doi.org/10.7208/chicago/9780226447919.001.0001
22. Boden, M. A. (2004). The creative mind: Myths and mechanisms (2nd ed.). Routledge.
23. Floridi, L., & Sanders, J. W. (2004). On the morality of artificial agents. *Minds and Machines*, 14(3), 349–379. doi: https://doi.org/10.1023/B:MIND.0000035461.63578.9d
24. Livingstone, S., & Helsper, E. J. (2007). Gradations in digital inclusion: Children, young people and the digital divide. *New Media & Society*, *9(4), 671–696.* doi: https://doi.org/10.1177/1461444807080335
25. Selwyn, N. (2019). What's the problem with learning analytics? *Journal of Learning Analytics*, *6(3), 11–19.* doi: https://doi.org/10.18608/jla.2019.63.3

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

ГІБРИДНА ЕПІСТЕМОЛОГІЯ: ЕМЕРДЖЕНТНІ ФОРМИ ЗНАННЯ В ЕПОХУ КОГНІТИВНОЇ ІНТЕГРАЦІЇ ЛЮДИНИ ТА ШТУЧНОГО ІНТЕЛЕКТУ

У статті представлено гібридну епістемологію як підхід до розуміння того, як формується знання у процесі когнітивної взаємодії людини та штучного інтелекту. Традиційна філософія розглядала знання як те, що окрема особа здобуває через досвід або раціональне мислення. Однак цей погляд уже не відображає сучас-

ної реальності. У дослідженні розглядаються три ключові виклики: як оцінювати пізнавальні твердження, якщо процес мислення містить непрозорі алгоритми? Що відбувається, коли мережі довіри замінюють традиційні методи обґрунтування? Як розподіляється когнітивна відповідальність у партнерстві між людиною та машиною? Ці питання мають значення, оскільки гібридні системи знання вже впливають на медичні діагнози, юридичні рішення та політичні вибори, що безпосередньо стосуються життя людей.

Ключові слова: *штучний інтелект, епістемологія, розподілене пізнання, розширений розум, епістемічна непрозорість, когнітивна агенція, взаємодія людини та машини.*