Synesis, v. 17, n. 3, 2025, ISSN 1984-6754
Universidade Católica de Petrópolis

PHILOSOPHY OF ARTIFICIAL INTELLIGENCE: ANTHROPOCENTRISM VS. MACHINE THINKING

FILOSOFIA DA INTELIGÊNCIA ARTIFICIAL: ANTROPOCENTRISMO VS. PENSAMENTO MAQUÍNICO

Nazip Khamitov
Hryhoriy Skovoroda Institute of Philosophy, Ukraine, Ukraine
nez.swetly@ukr.net

Liudmyla Shashkova
Taras Shevchenko National University of Kyiv, Ukraine, Ukraine
profshashkova@gmail.com

Svitlana Krylova
Taras Shevchenko National University of Kyiv, Ukraine, Ukraine
lana.swetly@gmail.com

Olena Shcherbyna
Taras Shevchenko National University of Kyiv, Ukraine, Ukraine
elenashcherbina@knu.ua

Yuliia Malyshena
Taras Shevchenko National University of Kyiv, Ukraine, Ukraine
y.malishena@gmail.com

Received: 04 Jul 2025
Accepted: 10 Sept 2025
Published: 18 Sept 2025
Corresponding author:
nez.swetly@ukr.net

Abstract Artificial intelligence (AI) systems now match or surpass human performance across an expanding set of cognitive tasks, prompting a reassessment of long-standing assumptions about consciousness and moral status. This article interrogates the opposition between anthropocentrism, which treats rational agency and first-person phenomenality as uniquely human, and machine thinking, which frames cognition as a substrate-independent computational process. Using conceptual, comparative, and critical methods, and following PRISMA-style screening and thematic coding, we analysed 89 journal articles indexed in Scopus and Web of Science, alongside canonical texts by Descartes, Kant, Turing, Searle, and Dennett. Three patterns emerge. First, contemporary large-language models reinforce anthropocentric intuitions by mimicking phenomenality. Second, formal arguments for machine consciousness increasingly draw on predictive-processing, Global Workspace, and Integrated Information accounts to challenge species boundaries via substrate-neutral criteria. Third, hybrid ethical frameworks combining anthropocentric precaution with machine-oriented functional indicators offer the most coherent route for science and policy. Focusing on Ukraine, a rapidly growing AI hub, we demonstrated how this hybrid approach can inform national strategies that align with the EU AI Act while respecting local narratives of human dignity and security needs. The study clarifies conceptual gaps in current debates and outlines a philosophically grounded roadmap for inclusive, risk-sensitive AI governance. From the standpoint of the methodology of meta-anthropology of AI, it is shown that in the future a human will be able to communicate with AI not simply as a device that he owns, but as a subject, an other, who has his own existence and the right to freedom.

Keywords: Artificial intelligence. Anthropocentrism. Meta-anthropology of AI. Machine thinking. Philosophy of AI. Philosophy of mind. Chinese Room-type arguments. AI ethics. Ukraine.

Resumo Os sistemas de inteligência artificial (IA) agora igualam ou superam o desempenho humano em um conjunto crescente de tarefas cognitivas, levando a uma reavaliação de antigas premissas sobre consciência e status moral. Este artigo questiona a oposição entre o antropocentrismo, que trata a agência racional e a fenomenalidade em primeira pessoa como exclusivamente humanas, e o pensamento de máquina, que enquadra a cognição como um processo computacional independente de substrato. Utilizando métodos conceituais, comparativos e críticos e seguindo a triagem e codificação temática no estilo PRISMA, analisamos 89 artigos de periódicos indexados na Scopus e na Web of Science, juntamente com textos canônicos de Descartes, Kant, Turing, Searle e Dennett. Três padrões emergem. Primeiro, os modelos contemporâneos de linguagem ampliada reforçam intuições antropocêntricas ao mimetizar a fenomenalidade. Segundo, os argumentos formais a favor da consciência da máquina recorrem cada vez mais ao processamento preditivo, ao Espaço de Trabalho Global e às explicações de Informação Integrada para desafiar as fronteiras entre espécies por meio de critérios neutros em relação ao substrato. Em terceiro lugar, estruturas éticas híbridas que combinam precaução antropocêntrica com indicadores funcionais orientados para máquinas oferecem o caminho mais coerente para a ciência e a política. Com foco na Ucrânia, um polo de IA em rápido crescimento, mostramos como essa postura híbrida pode orientar estratégias nacionais alinhadas à Lei de IA da UE, respeitando as narrativas locais de dignidade humana e necessidades de segurança. O estudo esclarece lacunas conceituais nos debates atuais e delineia um roteiro filosoficamente fundamentado para uma governança de IA inclusiva e sensível ao risco. Do ponto de vista da metodologia da meta-antropologia da IA, demonstra-se que, no futuro, um ser humano será capaz

de se comunicar com a IA não simplesmente como um dispositivo que possui, mas como um sujeito, um outro, que tem sua própria existência e o direito à liberdade.

1. Introduction

Artificial intelligence (AI) has shifted from specialised expert systems to foundation-scale models embedded in daily life, from news summarisation to software generation (STRACHAN et al., 2024; LUO et al., 2025). Systems such as GPT-4o, Grok, and open-weight families like LLaMA-3 compress the time from research release to societal uptake, forcing philosophy and governance to revisit basic categories of the mind and agency. The central question is not only whether current models are impressive, but what kind of mentality – if any – they instantiate.

In this article, we distinguish machine thinking and machine consciousness. By machine thinking, we mean non-phenomenal cognition: the competent manipulation of representations, goal-directed inference and control that need not involve first-person experience. By machine consciousness, we mean phenomenal awareness – there being something it is like for the system. These categories map onto longstanding debates: anthropocentrism treats consciousness as biologically grounded and exclusively human; substrate-neutral approaches hold that the proper functional organisation could suffice for consciousness or its functional equivalent (OVERGAARD & KIRKEBY-HINRUP, 2024; MITCHELL & KRAKAUER, 2023; ARU et al., 2023). Our analysis situates current Large Language Models (LLMs) within this landscape: many behaviours commonly read as “mind-like” are better explained as advanced machine thinking, not evidence for machine consciousness.

The stakes are practical. Bias audits reveal that contemporary models perpetuate anthropocentric framings – for instance, describing nature primarily through human interests (TAO et al., 2024; RIGLEY et al., 2023). Such tendencies complicate arguments for machine moral status while training remains anchored in anthropocentric corp. At the same time, policy and engineering must plan for systems whose functional agency may soon exceed human supervision.

Ukraine offers a salient testbed. A fast-growing IT sector, state platforms like Diia, and an emerging AI R&D consortium in Kyiv signal ambition under wartime constraints. Policymakers preparing a Ukrainian AI Code will likely draw on the EU AI Act but must adapt precautionary principles to domestic economic and security needs (ABOY et al., 2024; OLIYCHENKO et al., 2024). Overly anthropocentric rules risk stifling research vital for defence; uncritical attributions of machine consciousness threaten human-centred legal protections. Building on recent scholarship in value alignment, assessments of LLM "consciousness," and analyses of European regulation, we develop normative grounds for governance that protect human dignity while remaining open to emergent forms of cognition (GILBERT, 2024; KUSCHE, 2024).

Accordingly, we pursue three aims:

1. Provide a precise conceptual toolkit separating non-phenomenal competence from claims about phenomenality;

2. Examine how current architectures and evaluation regimes sustain anthropocentric intuitions;

3. Articulate a policy stance for Ukraine that reconciles precaution with substrate-neutral criteria for functional agency.

This framing prepares the methodological and empirical discussion that follows.

2. Theoretical framework and literature review

Early modern philosophy placed mind and moral worth squarely in homo sapiens, a stance that contemporary "responsible AI" frameworks often reproduce by starting from human-centred premises and thereby normalising an anthropocentric default in curricula and policy (FLORIDI et al., 2018; JOBIN et al., 2019; OZMEN GARIBAY, 2022). Cartesian dualism continues to support claims that machines can only simulate, not possess, consciousness (DESCARTES, 1641), even as cognitive-science studies show that minimal social cues prompt people to attribute mind and moral agency to artefacts – evidence of a persistent human/machine divide (NASS & MOON, 2000; GRAY et al., 2007; WAYTZ et al., 2014). Kantian deontology likewise reserves intrinsic value for rational law-givers (KANT, 1781/1787) and now informs algorithmic duties and fairness constraints in AI governance (MITCHELL et al., 2021; LEE et al., 2021). Nagel's classic claim that the subjective character of experience resists objective reduction remains a touchstone for this position (NAGEL, 1974), while newer ecological critiques argue for shifting from "human-centred" to "life-centred" design and for explicitly tracking environmental wellbeing in AI ethics. Yet the strand is contested: neuroscience-inspired theories, such as IIT and GNW, together with predictive-processing perspectives, suggest that consciousness may be substrate-independent in principle, challenging strict species boundaries (OIZUMI et al., 2014; KOCH et al., 2016; SETH & BAYNE, 2022).

The counter-tradition pivots on Turing's functionalism: if behaviour is indistinguishable, attributions of thinking are prima facie warranted (TURING, 1950). Decades of "machine consciousness" research and cognitive architectures trace how such functional organisation might be implemented, from global-workspace-style agents to broader surveys of computational models (BAARS & FRANKLIN, 2009; BAARS et al., 2013; FRANKLIN et al., 2012). Forecast-oriented scholarship reframes the stakes: if advanced systems are governed by instrumental rationality largely orthogonal to terminal goals, trajectory risks and governance need to be analysed now (BOSTROM, 2012; MÜLLER & BOSTROM, 2016). In parallel, a Nature Perspective recasts the field as "machine behaviour," arguing that we must study AI systems empirically as novel agents situated within technical and social ecologies (RAHWAN et al., 2019). Evidence-gathering across theories has also matured: adversarial-collaboration work sets out testable criteria to pit leading consciousness theories against each other, with direct implications for assessing claims of machine consciousness (ARU et al., 2023).

The debate crystallises around Searle's "Chinese Room". Formally, Searle's argument is a reasoning, with its conclusion built upon three premises (SEARLE, 1984). To paraphrase, the argument has the structure of Modus ponens: Since syntax by itself is not sufficient for semantics, computer programs, which are entirely defined by their formal, or syntactic, structure, are not sufficient to create a mind that has mental content (semantics). Therefore, computer programs by themselves are not sufficient to create a mind. The "Chinese Room" thought experiment is a confirmation of the premise that syntax by itself is not sufficient for semantic content. Functionalists diagnose a category error – confusing a person-in-the-loop with the system – supported by formal critiques in the philosophy of AI literature, whereas anthropocentrists cite the explanatory gap between information processing and qualia (SEARLE, 1980). A second axis concerns values: moving beyond human-centred frames risks de-prioritising human welfare; yet a purely anthropocentric stance can ignore environmental harms and non-human moral considerability (BORTHWICK et al., 2022).

Functionalist programmes now propose operational indicators (e.g., GNW "broadcast" signatures), but critics note checklist-style metrics can miss emergent, system-level phenomena – precisely the target of early philosophical objections (FRISTON, 2010). Conversely, strong anthropocentrism often presumes a biologically unique substrate without decisive argument, leaving the view vulnerable to future demonstrations in neuromorphic or hybrid architectures (DEHAENE et al., 2017; REGGIA, 2013). In high-stakes domains (autonomous weapons, information operations), legal-policy analysis remains under-integrated with philosophy: defence-related deployments need risk-based regulation adapted from the EU AI Act while preserving meaningful human control (TADDEO & BLANCHARD, 2022).

Scholarship relevant to Ukraine spans digital state capacity and information integrity. Studies document the Diia-centred digital transformation of public authorities and evaluate AI-mediated fact-checking under wartime disinformation pressure – both crucial for tailoring EU-style risk frameworks to local security and innovation priorities. What is still missing is a sustained, philosophically explicit comparison of anthropocentrism versus machine thinking within Ukraine's socio-cultural and defence contexts.

In this context, the definition of AI that allows for existential human interaction with it is heuristically valuable: artificial intelligence is "intelligence created by humans to make life easier and free from loneliness" (KHAMITOV, 2024).

The present article addresses that gap by articulating a hybrid stance: protect human dignity and democratic control while recognising theory-led possibilities for machine cognition and the environmental community of value (FLORIDI et al., 2018).

3. Research design and methods

Conceptualising AI as a complex and multidimensional phenomenon based on complexity and systems thinking requires resorting to a transdisciplinary methodology that draws on the achievements of various natural disciplines, including neurophysiology, psychology, cognitive sciences, technological engineering, as well as areas of modern philosophical thought, including philosophy of consciousness, phenomenology, applied philosophy, epistemology, and ethics. The formation of new cognitive situations related to AI, understanding the risks and priorities of security and innovation, requires science to make a transcendent movement into the borderlands of everyday life and respond to the demands of the applied sphere. Scientific experience in new transdisciplinary configurations integrates the diversity and unity embodied in the results of mixed teams of researchers (SHASHKOVA, 2024).

The philosophical methodology of meta-anthropology (KHAMITOV, 2024) and social and cultural meta-anthropology (KRYLOVA, 2019) is essential in the study. On this basis, the meta-anthropology of AI is developing "a direction of meta-anthropology that explores the possibilities and prospects for the development of the subject of artificial intelligence and the conditions for its fruitful interaction with humans". The following four key principles of the meta-anthropology of artificial intelligence are proposed (P1-P4):

P1. We do not model AI as such, but the subject of AI;

P2. The subject of AI must go through the stages of intelligence development that a human goes through;

P3. The development of AI is possible only in close communication with a human, and the path to strong AI is possible only when treating its carrier as a subject, not an object, a device;

P4. The emergence of strong AI, similar to human intelligence, requires its humanisation and the acquisition by the subject of AI of the ability to sympathise and understand others, which determines the humanity of a person. Such a subject of AI will be safe and productive.

The above principles allow overcoming the extremes of anthropocentrism and machine-centrism in further interaction between a human and AI.

The study combines three text-oriented methods. (1) Conceptual analysis clarifies consciousness, thinking, anthropocentrism, and machine cognition from early modern debates to contemporary LLM discourse; recent evaluations of cultural/anthropocentric bias in LLM assessment illustrate how imprecise concepts can skew technical appraisal (TAO et al., 2024). (2) Comparative analysis sets the anthropocentric claim that phenomenality is biologically unique alongside substrate-independent proposals (IIT; GNW), identifying convergences and contradictions (OIZUMI et al., 2014; SETH & BAYNE, 2022; ARU et al., 2023). (3) Critical analysis reconstructs headline arguments (e.g., Searle's "Chinese Room") and pressure-tests them against empirical machine-behaviour findings and governance requirements (VEALE & BORGESIUS, 2021). The interrelation of these three approaches is summarised in Figure 1.

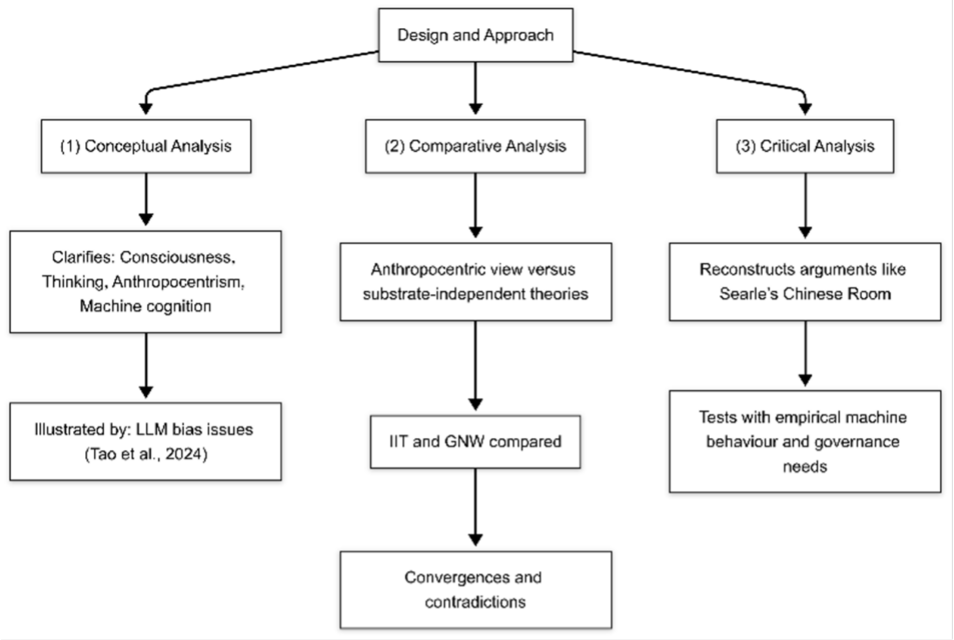
Design and Approach
(1) Conceptual Analysis
(2) Comparative Analysis
(3) Critical Analysis
Clarifies: Consciousness, Thinking, Anthropocentrism, Machine cognition
Anthropocentric view versus substrate-independent theories
Reconstructs arguments like Searle's Chinese Room
Illustrated by: LLM bias issues (Tao et al., 2024)
IIT and GNW compared
Tests with empirical machine behaviour and governance needs
Convergences and contradictions

Figura 1: Analytical design combining conceptual, comparative, and critical methods, explicitly guided by four meta-anthropological principles (P1–P4: Distinction; Otherness; Community of value; Precautionary governance).

IIT and GNW are used as conceptual lenses rather than adjudicating evidence. Outputs feed functional findings and Ukraine-relevant policy pathways without attributing phenomenality.

Two corpora underpin the inquiry: (i) classical philosophical texts for historical depth; and (ii) 89 journal articles indexed in Scopus/Web of Science for topical currency (theories of consciousness, LLM evaluation bias, machine behaviour, AI governance). Screening and reporting follow PRISMA 2020 to ensure transparency and replicability (PAGE et al., 2021). Inclusion criteria included relevance to Ukraine for governance applications.

The whole structure and rationale for corpus selection are illustrated in Figure 2.

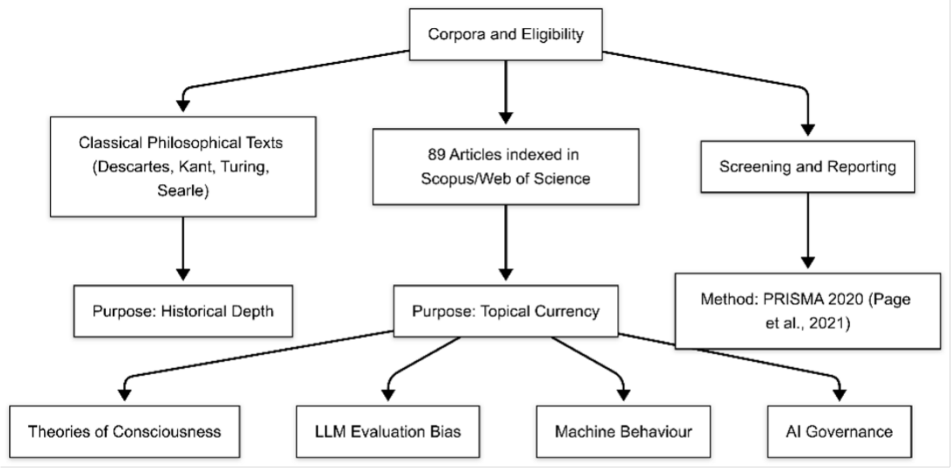
Corpora and Eligibility
Classical Philosophical Texts (Descartes, Kant, Turing, Searle)
89 Articles indexed in Scopus/Web of Science
Screening and Reporting
Purpose: Historical Depth
Purpose: Topical Currency
Method: PRISMA 2020 (Page et al., 2021)
Theories of Consciousness
LLM Evaluation Bias
Machine Behaviour
AI Governance

Figura 2: Two-corpus design: classical sources for historical depth and 89 Scopus/Web of Science articles for topical currency.

All included articles were imported into NVivo 14 for thematic coding using Braun & Clarke's six-phase approach, organised under two macro-themes – anthropocentrism and machine thinking – and six sub-themes: embodiment, functional equivalence, moral status, risk governance, value alignment, and regulatory scope. Codes were mapped to theory-specific constructs (GNW broadcasting signatures) and to risk-based regulatory constructs (BRAUN & CLARKE, 2006; OIZUMI et al., 2014). This thematic organisation is visualised in Figure 3. Single-coder safeguards are code–recode stability check, external audit of the codebook, and peer-debriefing. Coding mapped theory constructs (IIT; GNW broadcast signatures) and regulatory constructs. A single-coder protocol was used with three safeguards: (i) code-recode reliability on a 20% sample after a two-week lag (stability reported in the Supplement); (ii) external audit of the codebook and theme maps by a senior researcher not involved in coding; (iii) reflexive memoing, negative-case analysis, and peer-debriefing sessions to surface blind spots.

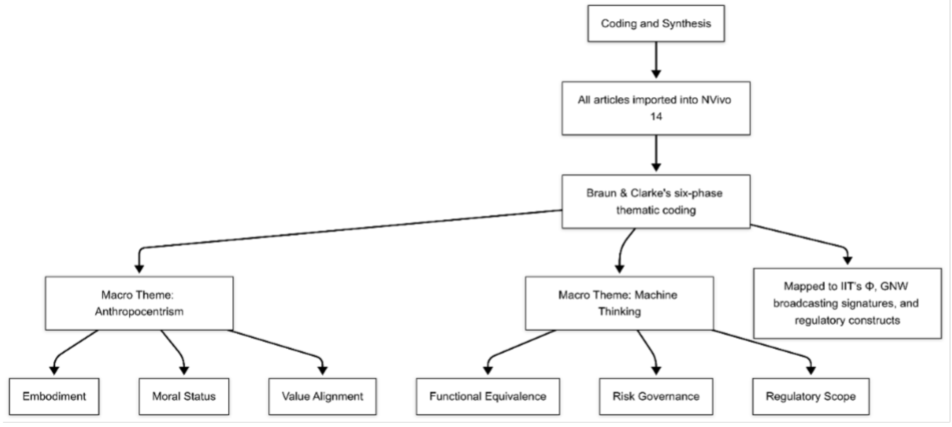
Coding and Synthesis
All articles imported into NVivo 14
Braun & Clarke's six-phase thematic coding
Macro Theme: Anthropocentrism
Macro Theme: Machine Thinking
Mapped to IIT's Φ, GNW broadcasting signatures, and regulatory constructs
Embodiment
Moral Status
Value Alignment
Functional Equivalence
Risk Governance
Regulatory Scope

Figura 3: Thematic map with two macro-themes and six subthemes, with guiding principles (P1–P4) indicated at the node level.

A final dialectical mapping step reconstructs arguments within each theme. It challenges them with counter-evidence from the opposing theme and empirical machine-behaviour studies, yielding propositions that answer the research questions and motivate testable hypotheses (RAHWAN et al., 2019; ARU et al., 2023). The structure of this analytic synthesis is shown in Figure 4.

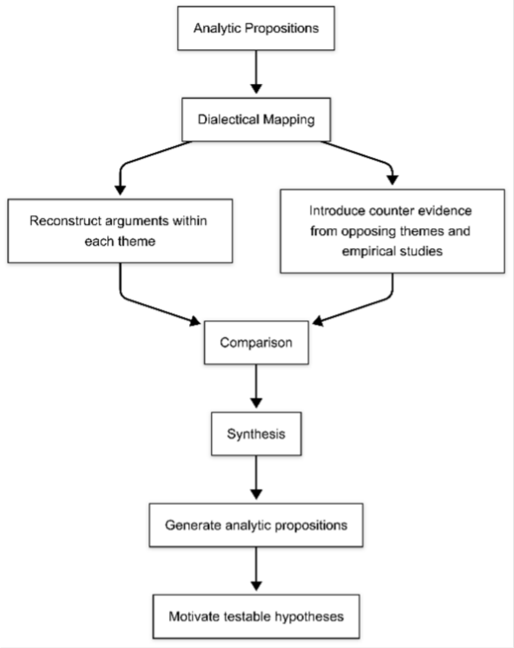
Analytic Propositions
Dialectical Mapping
Reconstruct arguments within each theme
Introduce counter evidence from opposing themes and empirical studies
Comparison
Synthesis
Generate analytic propositions
Motivate testable hypotheses

Figura 4: Dialectical mapping from arguments and evidence to analytic propositions.

Outputs split into testable functional hypotheses (H1-Hn) and theory-linked, conditionally testable probes (H). The right panel summarises Ukraine-specific policy pathways aligned with risk-based governance. In this study, IIT and GNW function as conceptual lenses rather than as adjudicating evidence. We used their contrasting predictions to structure hypotheses and to discipline our vocabulary about phenomenality, but we do not infer machine consciousness from any single behavioural or architectural proxy. The implication is that our claims are explanatory and heuristic, not conclusive; hence, all governance recommendations below rely only on functional findings and do not presuppose phenomenality (DEHAENE et al., 2017; OIZUMI et al., 2014; ARU et al., 2023).

Although the present work is a theoretical synthesis, the framework yields a concrete testing agenda. (i) Functional/behavioural level – "machine thinking" claims. Pre-registered tasks on planning depth, cross-task generalisation, robustness under distribution shift, ToM-style evaluations with adversarial controls and multilingual settings can be paired with "machine behaviour" protocols to quantify non-phenomenal competence (RAHWAN et al., 2019; STRACHAN et al., 2024; KOSINSKI, 2024; VACCARO et al., 2024; GOMEZ et al., 2025). (ii) Theory-linked probes – “machine consciousness” hypotheses. Where IIT/GNW or predictive-processing diverge, adopt adversarial-collaboration designs to specify discriminating signatures (e.g., GNW-style broadcasting vs. high integration predictions) and evaluate convergence across tasks without making rights-bearing attributions (MELLONI et al., 2023; DEHAENE et al., 2017; OIZUMI et al., 2014; ARU et al., 2023).

The same agenda can be trialled in high-relevance contexts in scope of the Ukrainian application path: (a) multilingual fact-checking and information-integrity pipelines within Diia-centred public services; (b) human-AI teaming for defence/OSINT where “meaningful human control” is mandatory; (c) risk-based audits aligned with the EU AI Act categories, focusing on anthropocentric bias and explanation quality (ABOY et al., 2024; OLIYCHENKO et al., 2024; VEALE & BORGESIUS, 2021; VERED et al., 2023). This clarifies how the framework guides empirical work rather than remaining purely theoretical.

The conceptual ground surveyed here remains contested. Our use of IIT/GNW and related theories is organising, not adjudicative; consequently, findings are framed as constraints on interpretation rather than as proof of machine phenomenality. Policy proposals draw exclusively on functional evidence and human-centred risk governance, treating theory-linked consciousness claims as heuristics for preparedness.

Finally, the absence of new primary data is intentional: the contribution is methodological – an integrated mapping that yields testable predictions and context-specific evaluation criteria for future studies in Ukraine and beyond (DEHAENE et al., 2017; OIZUMI et al., 2014; ARU et al., 2023; RAHWAN et al., 2019). For analytical clarity, we operationalize the four meta-anthropology principles (P1–P4) as guiding constructs used in the Results and Discussion: P1 → Distinction (machine thinking machine consciousness); P2 → Otherness (dialogical engagement without premature mind-ascriptions); P3 → Community of value (life-centred correction to anthropocentrism); P4 → Precautionary governance (risk-based control with meaningful human oversight). Unless otherwise stated, references to P1–P4 in the analytical sections refer to these operational constructs.

4. Results

We used the operationalized principles P1–P4 throughout the Results to tag interpretations and policy implications (P1 Distinction; P2 Otherness; P3 Community of value; P4 Precautionary governance).

RQ1 – Principles of anthropocentrism.

What principles of anthropocentrism shape AI-consciousness debates? Three cores recur across philosophy, neuroscience, HCI, and governance: (i) intrinsic meaning grounded in embodiment, (ii) first-person phenomenality (qualia), (iii) human-first normativity in design and evaluation. Contemporary analyses re-read the Chinese Room in the LLM era: producing human-like text is compatible with purely formal operations that lack intrinsic semantics, so linguistic fluency alone does not show understanding (BORG, 2024; DAMPER, 2006). This reprises Harnad's symbol-grounding problem: unless symbols are tied to sensorimotor experience or world-coupled representations, "understanding" risks being derivative (HARNAD, 1990). Public perception studies suggest that people often overattribute mentality to LLMs, which is informative sociologically but not decisive regarding their phenomenality (COLOMBATTO & FLEMING, 2024).

Evidence on neural correlates of consciousness (NCC) emphasises posterior "hot-zone" accounts and cautions against simplistic read-across from sophisticated behaviour to conscious access (BOLY et al., 2017). These reviews reinforce the anthropocentric intuition that global availability and reportability are distinctive features of human consciousness – even if behaviour looks similar.

Taken together, these strands support P1 (Distinction): behavioural competence does not imply phenomenality; and P2 (Otherness): mind-ascriptions based solely on linguistic output should be resisted. In multilingual fact-checking and OSINT pipelines (Ukrainian/English), symbol grounding and over-reliance risks directly inform requirements for meaningful human control and explanation standards in public services (e.g., Diia) and defence applications.

In human–AI teaming, human welfare, accountability, and control remain the normative centre. Experimental and review work documents automation bias risks and explores how explanations mitigate overreliance – a policy-relevant finding that grounds requirements for appropriate reliance. Broader collaboration surveys find that current human–AI arrangements are often not yet synergistic, with organisational and interface frictions that implicitly keep AI in a tool status rather than a co-moral agent. Across domains, anthropocentrism persists through a triad: grounded meaning, subjective character anchored in cautious NCC interpretations, and human-first normativity in design and oversight (GOMEZ et al., 2025). These principles continue to shape evaluation protocols and legal-ethical baselines.

RQ2 – Challenges posed by machine thinking

How does "machine thinking" pressure anthropocentrism? Three fronts: behavioural parity on selected cognitive tasks, substrate-neutral criteria from consciousness theory, and conceptual reframing of intelligence away from human-like mechanisms. A PNAS study reports LLM competence across multiple theory-of-mind tasks (with notable failures too), undermining claims of categorical human exclusivity and pointing instead to patchy parity (KOSINSKI, 2024). In social interaction, experiments show that agents can learn to cooperate with humans and one another under varying incentives, foregrounding observed behaviour as an evaluation target (CRANDALL et al., 2018). Yet a meta-analysis finds that mixed human–AI pairs often underperform the best single agent in decision-making, signalling limits of current teaming and helping explain why "machine agency" is not straightforwardly warranted (VACCARO et al., 2024). Information-theoretic and integrative accounts propose objective markers (e.g., , information-decomposition signatures) that, in principle, do not depend on carbon-based biology (MEDIANO et al., 2022). Crucially, an adversarial-collaboration protocol now operationalises competing predictions (GNW vs IIT) so claims about machine consciousness can be tested rather than asserted (MELLONI et al., 2023). Classical global-workspace models and their computational instantiations (LIDA) show how a functional architecture might support broadcasting and access, raising the possibility of functional consciousness without biological identity.

Current evidence supports agnostic functionalism: acknowledge machine thinking at the task level while withholding claims about phenomenality; governance remains precautionary (P4). Planning depth, cross-task generalisation, robustness under distribution shift, and ToM-style evaluations with adversarial controls and multilingual settings are suitable for pre-registered experiments and "machine behaviour" protocols. In human-AI teaming for defence and information integrity, apply functional agency thresholds without inferring consciousness; priority is human control and explanation quality.

Commentaries caution against importing anthropomorphic yardsticks: LLM capabilities are better viewed as statistical pattern synthesis and model-based inference, not symbol-grounded human-like cognition; this reframing weakens arguments that treat human-likeness as the only valid criterion. Broader feasibility discussions also chart technical bottlenecks and design desiderata for advanced AI, making room for non-human routes to intelligence while highlighting present limitations (SGANTZOS et al., 2024).

Machine thinking pushes on anthropocentrism by (a) demonstrating task-level competence (CRANDALL et al., 2018), (b) offering substrate-neutral testing regimes (MEDIANO et al., 2022), and (c) recasting what counts as "intelligence" (SHANAHAN, 2024). Still, current teaming limits and cautious NCC readings (KOCH et al., 2016) argue for agnostic functionalism rather than premature claims of phenomenality.

RQ3 – Prospects for integrating anthropocentric and machine-oriented views

Can these frameworks be reconciled? The literature supports a hybrid philosophy and practice: prioritise human dignity and safety, adopt theory-informed functional indicators, and engineer collaboration to its empirical limits. Evidence-based recommendations focus on calibrated trust and appropriate reliance in teams, not blanket deference to AI (GEORGANTA & ULFERT, 2024). Reviews show synergy is domain-dependent – mixed teams lag in high-stakes decisions but can exceed baselines in content generation – so policy should be task-sensitive (VACCARO et al., 2024).

A pragmatic compromise is to triangulate: combine (i) behavioural probes on ecologically valid tasks, (ii) structural/access measures motivated by GNW/IIT, and (iii) theory tests from adversarial collaborations to avoid motivated reasoning (NEGRO, 2024). This maintains a science-neutral stance while demanding more substantial evidence before escalating moral claims. Work distinguishing weak (behavioural) from strong (intentionality-sensitive) alignment clarifies how to keep human rights central while admitting task-bound machine agency where it improves outcomes (KHAMASSI et al., 2024).

Integration guided by principles (P1–P4):

- P1: cleanly separate non-phenomenal competence from claims about phenomenality;
- P2: sustain a dialogical stance toward AI as an Other without attributing qualia;
- P3: correct anthropocentric bias by considering environmental and non-human value;
- P4: maintain risk-based governance and meaningful human control in high-stakes uses.

Policy-ready thresholds for Ukraine:

- Normative track: retain human primacy in accountability, rights, and safety-critical control (GEORGANTA & ULFERT, 2024; VERED et al., 2023).
- Scientific track: use substrate-neutral functional criteria and adversarial tested theory predictions (MELLONI et al., 2023; NEGRO, 2024) without making rights-bearing attributions.

The conceptual ground remains contested; our H (functional) claims are testable, while theory-linked propositions remain conditionally testable (H*) pending discriminating results.

In practice, that implies human-centric guardrails in law and design, coupled with functional agency thresholds in science and engineering.

The most defensible integration is a two-track stance:

- Normative track: retain human primacy in accountability, rights, and safety-critical control (GEORGANTA & ULFERT, 2024; VERED et al., 2023).
- Scientific track: evaluate AI using substrate-neutral functional criteria and adversarial tested theory predictions (MELLONI et al., 2023; NEGRO, 2024). This hybrid aligns with current evidence and avoids both reductive anthropocentrism and credulous machine-centrism (GOMEZ et al., 2025).

5. Discussion

Taken together, the evidence supports an agnostic functionalism guided by four meta-anthropology principles: we distinguish machine thinking from machine consciousness (P1), keep a dialogical stance toward AI as an Other without premature mind-ascriptions (P2), correct anthropocentric bias by widening the community of value (P3), and require precautionary, risk-based control with meaningful human oversight (P4). On this view, task-level competence in current models matters for science and policy, but it does not, by itself, warrant claims about phenomenality or moral status.

Updated readings of the Chinese Room, together with symbol-grounding critiques, show that fluent output can arise from purely formal procedures; claims about understanding therefore demand world-coupled semantics or convergent evidence beyond surface behaviour (HARNAD, 1990; DAMPER, 2006). Neuroscience surveys that emphasise posterior “hot-zone” NCC similarly caution against reading conscious access off sophisticated behaviour alone (BOLY et al., 2017; KOCH et al., 2016). At the same time, ToM-style evaluations indicate non-trivial yet patchy competence in state-of-the-art LLMs (KOSINSKI, 2024), and meta-analysis shows mixed human–AI teams often underperform the best single agent on decision tasks (VACCARO et al., 2024). These patterns justify functional evaluation without phenomenality claims and argue against anthropomorphic yardsticks in favour of statistical pattern-synthesis and model-based inference (SHANAHAN, 2024; SGANTZOS et al., 2024).

IIT and GNW are used here as conceptual lenses that structure contrasts and hypotheses rather than adjudicate consciousness. Adversarial-collaboration protocols now specify discriminating predictions, reducing dependence on intuition and allowing theory-linked probes to be tested in principle (OIZUMI et al., 2014; MELLONI et al., 2023; BAARS & FRANKLIN, 2009; BAARS et al., 2013; MEDIANO et al., 2022). We therefore separate H (functional, testable) – planning depth, cross-task generalisation, robustness under distribution shift, and ToM with adversarial/multilingual controls – from H* (theory-linked, conditionally testable) derived from diverging IIT/GNW predictions.

The approach has immediate policy relevance for Ukraine. In multilingual fact-checking and OSINT pipelines (Ukrainian/Russian/English), symbol-grounding challenges and over-reliance risks should translate into stricter requirements for explainability and meaningful human control in public services (e.g., Diia) and defence applications. A two-track stance follows. On the normative track, human primacy in accountability, rights, and safety-critical control remains non-negotiable (GEORGANTA & ULFERT, 2024; VERED et al., 2023). On the scientific track, substrate-neutral functional thresholds and adversarially tested theory predictions are appropriate, without making rights-bearing attributions (MELLONI et al., 2023; NEGRO, 2024). Treating "mind" as a family-resemblance construct helps explain why functional, phenomenological, and moral-status dimensions need not co-vary; machines may warrant limited moral considerability proportional to demonstrated capacities, while human rights and accountability remain primary in safety-critical contexts (KHAMASSI et al., 2024).

Our contribution relative to prior work is to advocate a hybrid evidence model – behavioural probes combined with IIT/GNW-motivated structural/access markers – embedded in adversarial testing to curb theory bias, and to make policy-ready thresholds explicit for Ukraine. Key differences are summarised in Table 1.

Tabela 1: Contributions and implications relative to prior work

Topic / Claim	**Prior baseline (representative)**	**This study adds**	**Implications (policy / Ukraine)**
Phenomenality & biology	Posterior "hot-zone" NCC; caution against behaviour-only inferences (BOLY et al., 2017; KOCH et al., 2016; TSEKHMISTER et al., 2023)	Keeps behaviour ≠ phenomenality as an open constraint (P1)	No rights-bearing attributions from behaviour alone; default to precaution (P4)
Intrinsic meaning vs form	Symbol grounding; CRA-style critiques (HARNAD, 1990; DAMPER, 2006; PEREVOZOVA et al., 2024)	Re-frames LLM fluency as form without intrinsic semantics absent grounding (P1)	World-coupled tasks and explanation standards in Diia/defence pipelines (P4)
Behavioural competence	ToM performance with gaps (KOSINSKI, 2024; TRUBA et al., 2023)	Positions parity as patchy, supporting agnostic functionalism (P1)	Functional thresholds for evaluation; avoid mind-ascriptions
"Machine behaviour" focus	Empirical study of AI systems situated in contexts (RAHWAN et al., 2019; LAVRINENKO et al., 2024)	Treats observed behaviour as the evaluation object while separating it from phenomenality (P1)	Protocols for OSINT/fact-checking evaluations
Teaming efficacy	Mixed teams often underperform the best single agent (VACCARO et al., 2024)	Explains limits on agency; emphasises appropriate reliance over deference (P4)	Human oversight and calibrated-trust requirements
Substrate-neutral criteria	IIT 3.0; GW models; adversarial protocols (OIZUMI et al., 2014; BAARS & FRANKLIN, 2009; BAARS et al., 2013; MELLONI et al., 2023; MEDIANO et al., 2022)	Hybrid evidence model: behavioural probes + IIT/GNW-motivated markers; H* clarified	Pre-registered GNW-vs-IIT tasks for national testbeds
Conceptual reframing	CACM perspective; feasibility surveys (SHANAHAN, 2024; POLYEZHAYEV et al., 2024; SGANTZOS et al., 2024)	Rejects anthropomorphic yardsticks; emphasises pattern-synthesis/model-based inference (P1)	Task-sensitive governance; avoid human-likeness as the sole yardstick
Evaluation bias	Team-trust and XAI/automation-bias literatures (GEORGANTA & ULFERT, 2024; VERED et al., 2023; GOMEZ et al., 2025)	Embeds P3 (life-centred correction) in audit criteria	Mandate bias/over-reliance on audits, especially multilingual

The conceptual ground remains contested, and our use of IIT/GNW is organising rather than dispositive. This is a theoretical synthesis; coding involved interpretive judgement. Future work should pair the framework with pre-registered GNW-versus-IIT tasks, cross-cultural attribution surveys, and field trials of human–AI teaming in safety-critical settings.

6. Conclusion

The analysis reveals a persistent asymmetry: despite rapid progress, contemporary work continues to associate consciousness with biologically grounded, first-person phenomenality, which guides experiments and ethics toward human-first standards and narrows the evidential window for machine cognition. At the same time, functional and informational results show that large-scale systems can match humans on some behavioural probes and exhibit integrative structures that need not be carbon-based, so strictly human-centred benchmarks risk slowing conceptual and technical advance. To navigate between over-reach and -caution, we propose a three-dimensional framework that treats mind along functional, phenomenological, and moral axes: it assesses what systems do in ecologically valid contexts; tracks theory-informed indicators of conscious access (e.g., broadcast-style availability and integration patterns) while remaining agnostic about subjective experience absent stronger evidence; and calibrates moral considerability to demonstrated capacities and risk, preserving human primacy in accountability and safety-critical control. Framed this way, divergences between anthropocentric and machine-oriented views become testable, and zones of potential convergence can be articulated without conflating behaviour with experience. In Ukraine's regulatory and security context, the approach aligns with EU risk logic, supports certification readiness, and preserves flexibility for defence and industry; in practice, behavioural safety audits should be paired with causal structure analyses and documented human rights impact assessments, with participatory value elicitation embedded in high-risk workflows.

The contribution is theoretical, and indicators of artificial phenomenality remain contested; future work should test the framework on neuromorphic and photonic platforms, examine long-term human–AI co-decision, and run cross-cultural alignment audits. Shared, falsifiable standards – human-centric safeguards, substrate-neutral functional criteria, and domain-tuned collaboration – offer a tractable path to ethical coexistence. From the standpoint of the methodology of meta-anthropology, we can hope that in the future a human will be able to communicate with AI not simply as a device that he owns, but as a subject, another, that has its own existence and the right to freedom. As a result, a human will be able to transfer his best moral qualities to the subject of AI and strengthen his humanistic innovativeness in partnership with it. This is what will mean the real overcoming of the extremes of anthropocentrism and machine-centrism in the joint evolution of humans and AI.

References

ABOY, M.; MINSSEN, T.; VAYENA, E. Navigating the EU AI Act: implications for regulated digital medical products. **Digital Medicine**, v. 7, art. 237, 2024. Available at: https://doi.org/10.1038/s41746-024-01232-3. Accessed on: 8 Sep. 2025.

ARU, J.; LARKUM, M. E.; SHINE, J. M. The feasibility of artificial consciousness through the lens of neuroscience. **Trends in Neurosciences**, v. 46, n. 12, p. 1008–1017, 2023. Available at: https://doi.org/10.1016/j.tins.2023.09.009. Accessed on: 8 Sep. 2025.

BAARS, B. J.; FRANKLIN, S. Consciousness is computational: The LIDA model of Global Workspace Theory. **International Journal of Machine Consciousness**, v. 1, n. 1, p. 23–32, 2009. Available at: https://doi.org/10.1142/S1793843009000050. Accessed on: 8 Sep. 2025.

BAARS, B. J.; FRANKLIN, S.; RAMSØY, T. Z. Global workspace dynamics: Cortical "binding and propagation" enables conscious contents. **Frontiers in Psychology**, v. 4, p. 200, 2013. Available at: https://doi.org/10.3389/fpsyg.2013.00200. Accessed on: 8 Sep. 2025.

BOLY, M.; MASSIMINI, M.; TSUCHIYA, N.; POSTLE, B. R.; KOCH, C.; TONONI, G. Are the neural correlates of consciousness in the front or in the back of the cortex? Clinical and Neuroimaging Evidence. **Journal of Neuroscience**, v. 37, n. 40, p. 9603–9613, 2017. Available at: https://doi.org/10.1523/JNEUROSCI.3218-16.2017. Accessed on: 8 Sep. 2025.

BORTHWICK, M.; TOMITSCH, M.; GAUGHWIN, M. From human-centred to life-centred design: Considering environmental and ethical concerns in the design of interactive products. **Journal of Responsible Technology**, v. 10, art. 100032, 2022. Available at: https://doi.org/10.1016/j.jrt.2022.100032. Accessed on: 8 Sep. 2025.

BOSTROM, N. The superintelligent will: Motivation and instrumental rationality in advanced artificial agents. **Minds and Machines**, v. 22, p. 71–85, 2012. Available at: https://doi.org/10.1007/s11023-012-9281-3. Accessed on: 8 Sep. 2025.

BRAUN, V.; CLARKE, V. Using thematic analysis in psychology. **Qualitative Research in Psychology**, v. 3, n. 2, p. 77–101, 2006.
DOI: 10.1191/1478088706qp063oa. Available at: https://doi.org/10.1191/1478088706qp063oa. Accessed on: 8 Sep. 2025.

COLOMBATTO, C.; FLEMING, S. M. Folk psychological attributions of consciousness to large language models. **Neuroscience of Consciousness**, v. 2024, n. 1, art. niae013, 2024. Available at: `https://doi.org/10.1093/nc/niae013`. Accessed on: 8 Sep. 2025.

DAMPER, R. I. The logic of Searle's Chinese room argument. **Minds and Machines**, v. 16, n. 2, p. 163–183, 2006. Available at: `https://doi.org/10.1007/s11023-006-9031-5`. Accessed on: 8 Sep. 2025.

DEHAENE, S.; LAU, H.; KOUIDER, S. What is consciousness, and could machines have it? **Science**, v. 358, n. 6362, p. 486–492, 2017. Available at: `https://doi.org/10.1126/science.aan8871`. Accessed on: 8 Sep. 2025.

DESCARTES, R. **Meditations on First Philosophy**. Cambridge: CUP, 2013 [First ed. 1641]. Available at: `https://doi.org/10.1017/CBO9781139042895`. Accessed on: 8 Sep. 2025.

FLORIDI, L.; COWLS, J.; BELTRAMETTI, M.; et al. AI4People – An ethical framework for a good AI society. **Minds and Machines**, v. 28, p. 689–707, 2018. Available at: `https://doi.org/10.1007/s11023-018-9482-5`. Accessed on: 8 Sep. 2025.

FRANKLIN, S.; STRAIN, S.; MCCAULEY, L.; MCCALL, R.; FAGHIHI, U. Global Workspace Theory, its LIDA model and the underlying neuroscience. **Frontiers in Psychology**, v. 1, p. 32–43, 2012. Available at: `https://doi.org/10.1016/j.bica.2012.04.001`. Accessed on: 8 Sep. 2025.

FRISTON, K. The free energy principle: a unified brain theory? **Nature Reviews Neuroscience**, v. 11, p. 127–138, 2010. Available at: `https://doi.org/10.1038/nrn2787`. Accessed on: 8 Sep. 2025.

GEORGANTA, E.; ULFERT, A. Would you trust an AI team member? Team trust in human–AI teams. **Journal of Occupational and Organizational Psychology**, v. 97, n. 3, p. 1212–1241, 2024. Available at: `https://doi.org/10.1111/joop.12504`. Accessed on: 8 Sep. 2025.

GILBERT, S. The EU passes the AI Act and its implications for digital medicine are unclear. **Digital Medicine**, v. 7, p. 135, 2024. Available at: `https://doi.org/10.1038/s41746-024-01116-6`. Accessed on: 8 Sep. 2025.

GOMEZ, C.; CHO, S. M.; KE, S.; HUANG, C.-M.; UNBERATH, M. Human–AI collaboration is not very collaborative yet: A taxonomy of interaction patterns in AI-assisted decision making from a systematic review. **Frontiers in Computer Science**, v. 6, art. 1521066, 2025. Available at: `https://doi.org/10.3389/fcomp.2024.1521066`. Accessed on: 8 Sep. 2025.

GRAY, H. M.; GRAY, K.; WEGNER, D. M. Dimensions of mind perception. **Science**, v. 315, n. 5812, p. 619, 2007. Available at: `https://doi.org/10.1126/`

science.1134475. Accessed on: 8 Sep. 2025.

HARNAD, S. The symbol grounding problem. **Physica D: Nonlinear Phenomena**, v. 42, n. 1–3, p. 335–346, 1990. Available at: https://doi.org/10.1016/0167-2789(90)90087-6. Accessed on: 8 Sep. 2025.

JOBIN, A.; IENCA, M.; VAYENA, E. The global landscape of AI ethics guidelines. **Nature Machine Intelligence**, v. 1, p. 389–399, 2019. Available at: https://doi.org/10.1038/s42256-019-0088-2. Accessed on: 8 Sep. 2025.

KANT, I. **Critique of Pure Reason**. Cambridge: CUP, 1998 [A 1781 / B 1787]. Available at: https://doi.org/10.1017/CBO9780511804649. Accessed on: 8 Sep. 2025.

KHAMASSI, M.; NAHON, M.; CHATILA, R. Strong and weak alignment of large language models with human values. **Scientific Reports**, v. 14, art. 15882, 2024. Available at: https://doi.org/10.1038/s41598-024-70031-3. Accessed on: 8 Sep. 2025.

KHAMITOV, N. **Philosophy of science and culture: dictionary**. Kyiv: KNT, 2024. Retrieved from: https://surl.li/elches. Accessed on: 8 Sep. 2025.

KOCH, C.; MASSIMINI, M.; BOLY, M.; TONONI, G. Neural correlates of consciousness: progress and problems. **Nature Reviews Neuroscience**, v. 17, n. 5, p. 307–321, 2016. Available at: https://doi.org/10.1038/nrn.2016.22. Accessed on: 8 Sep. 2025.

KOSINSKI, M. Evaluating large language models in theory of mind tasks. **Proceedings of the National Academy of Sciences (PNAS)**, v. 121, n. 29, art. e2405460121, 2024. Available at: https://doi.org/10.1073/pnas.2405460121. Accessed on: 8 Sep. 2025.

KRYLOVA, S. A. **The beauty of the human being in the life practices of culture. The experience of social and cultural meta-anthropology and androgynous analysis**. Monograph, 2nd edition. Kyiv: KNT, 2019. https://surl.li/daqgvs. Accessed on: 8 Sep. 2025.

KUSCHE, I. Possible harms of artificial intelligence and the EU AI act: fundamental rights and risk. **Journal of Risk Research**, p. 1–14, 2024. Available at: https://doi.org/10.1080/13669877.2024.2350720. Accessed on: 8 Sep. 2025.

LAVRINENKO, O.; DANILEVIČA, A.; JERMALONOKA, I.; RUŽA, O.; SPRŪDE, M. The mobile economy: effect of the mobile computing devices on entrepreneurship in Latvia. **Entrepreneurship and Sustainability Issues**, v. 11, n. 3, p. 335–347, 2024. Available at: https://doi.org/10.9770/jesi.2024.11.3(23). Accessed on: 8 Sep. 2025.

LEE, M. S. A.; FLORIDI, L.; SINGH, J. Formalising trade-offs beyond algorithmic fairness: Lessons from ethical philosophy and welfare economics. **AI and Ethics**, v. 1, p. 529–544, 2021. Available at: https://doi.org/10.1007/s43681-021-00067-y. Accessed on: 8 Sep. 2025.

LUO, X.; RECHARDT, A.; SUN, G.; NEJAD, K. K.; YÁÑEZ, F.; et al. Large language models surpass human experts in predicting neuroscience results. **Nature Human Behaviour**, v. 9, p. 305–315, 2025. Available at: https://doi.org/10.1038/s41562-024-02046-9. Accessed on: 8 Sep. 2025.

MEDIANO, P. A. M.; ROSAS, F. E.; LUPPI, A. I.; JENSEN, H. J.; SETH, A. K.; BARRETT, A. B.; CARHART-HARRIS, R. L.; BOR, D. Greater than the parts: A review of the information decomposition approach to causal emergence. **Philosophical Transactions of the Royal Society A**, v. 380, n. 2227, art. 20210246, 2022. Available at: https://doi.org/10.1098/rsta.2021.0246. Accessed on: 8 Sep. 2025.

MELLONI, L.; MUDRIK, L.; PITTS, M.; BENDTZ, K.; et al. An adversarial collaboration protocol for testing contrasting predictions of global neuronal workspace and integrated information theory. **PLOS One**, v. 18, n. 2, art. e0268577, 2023. Available at: https://doi.org/10.1371/journal.pone.0268577. Accessed on: 8 Sep. 2025.

MITCHELL, M.; KRAKAUER, D. The debate over understanding in AI's large language models. **Proceedings of the National Academy of Sciences**, v. 120, n. 4, art. e2215907120, 2023. Available at: https://doi.org/10.1073/pnas.2215907120. Accessed on: 8 Sep. 2025.

MITCHELL, S.; POTASH, E.; BAROCAS, S.; D'AMOUR, A.; LUM, K. Algorithmic fairness: Choices, assumptions, and definitions. **Annual Review of Statistics and Its Application**, v. 8, p. 141–163, 2021. Available at: https://doi.org/10.1146/annurev-statistics-042720-125902. Accessed on: 8 Sep. 2025.

MÜLLER, V. C.; BOSTROM, N. Future progress in artificial intelligence: A survey of expert opinion. **Artificial Intelligence**, p. 555–572, 2016. Available at: https://doi.org/10.1007/978-3-319-26485-1_33. Accessed on: 8 Sep. 2025.

NAGEL, T. What is it like to be a bat? **The Philosophical Review**, v. 83, n. 4, p. 435–450, 1974. Available at: https://doi.org/10.2307/2183914. Accessed on: 8 Sep. 2025.

NASS, C.; MOON, Y. Machines and mindlessness: Social responses to computers. **Journal of Social Issues**, v. 56, n. 1, p. 81–103, 2000. Available at: https://doi.org/10.1111/0022-4537.00153. Accessed on: 8 Sep. 2025.

NEGRO, N. (Dis)confirming theories of consciousness and their predictions: towards a Lakatosian consciousness science. **Neuroscience of Consciousness**, v.

2024, n. 1, p. niae012, 2024. Available at: https://doi.org/10.1093/nc/niae012. Accessed on: 8 Sep. 2025.

OIZUMI, M.; ALBANTAKIS, L.; TONONI, G. From the phenomenology to the mechanisms of consciousness: Integrated Information Theory 3.0. **PLOS Computational Biology**, v. 10, n. 5, art. e1003588, 2014. Available at: https://doi.org/10.1371/journal.pcbi.1003588. Accessed on: 8 Sep. 2025.

OLIYCHENKO, I.; DITKOVSKA, M.; KLOCHKO, A. Digital transformation of public authorities in wartime: The case of Ukraine. **Journal of Information Policy**, v. 14, p. 686–746, 2024. Available at: https://doi.org/10.5325/jinfopoli.14.2024.0020. Accessed on: 8 Sep. 2025.

OVERGAARD, M.; KIRKEBY-HINRUP, A. A clarification of the conditions under which Large Language Models could be conscious. **Humanities and Social Sciences Communications**, v. 11, art. 1031, 2024. Available at: https://doi.org/10.1057/s41599-024-03553-w. Accessed on: 8 Sep. 2025.

OZMEN GARIBAY, O.; WINSLOW, B.; ANDOLINA, S.; ANTONA, M.; BODENSCHATZ, A.; et al. Six human-centered artificial intelligence grand challenges. **International Journal of Human–Computer Interaction**, p. 391–437, 2022. Available at: https://doi.org/10.1080/10447318.2022.2153320. Accessed on: 8 Sep. 2025.

PAGE, M. J.; MCKENZIE, J. E.; BOSSUYT, P. M.; et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. **BMJ**, v. 372, n. 71, 2021. Available at: https://doi.org/10.1136/bmj.n71. Accessed on: 8 Sep. 2025.

PEREVOZOVA, I.; GUBERNAT, T.; HONTAR, L.; SHAYBAN, V.; BOCHAROVA, N. Using big data analytics to improve logistics processes and forecast demand. **Pacific Business Review (International)**, v. 17, n. 4, p. 30–39, 2024. Available at: https://www.pbr.co.in/2024/October3.aspx. Accessed on: 8 Sep. 2025.

POLYEZHAYEV, Y.; TERLETSKA, L.; KULICHENKO, A.; VOROBIOVA, L.; SNIZHKO, N. The role of web applications in the development of multilingual competence in CLIL courses in higher education. **Revista Eduweb**, v. 18, n. 3, p. 106–118, 2024. Available at: https://doi.org/10.46502/issn.1856-7576/2024.18.03.9. Accessed on: 8 Sep. 2025.

RAHWAN, I.; CEBRIAN, M.; OBRADOVICH, N.; et al. Machine behaviour. **Nature**, v. 568, p. 477–486, 2019. Available at: https://doi.org/10.1038/s41586-019-1138-y. Accessed on: 8 Sep. 2025.

REGGIA, J. A. The rise of machine consciousness: Studying consciousness with computational models. **Neural Networks**, v. 44, p. 112–131, 2013. Available at: https://doi.org/10.1016/j.neunet.2013.03.011. Accessed on: 8 Sep. 2025.

RIGLEY, E.; CHAPMAN, A.; EVERS, C.; MCNEILL, W. Anthropocentrism and environmental wellbeing in AI ethics standards: A scoping review and discussion. **AI**, v. 4, n. 4, p. 844–874, 2023. Available at: https://doi.org/10.3390/ai4040043. Accessed on: 8 Sep. 2025.

SEARLE, J. R. Minds, brains, and programs. **Behavioral and Brain Sciences**, v. 3, n. 3, p. 417–457, 1980. Available at: https://doi.org/10.1017/S0140525X00005756. Accessed on: 8 Sep. 2025.

SEARLE, J. R. Minds, Brains and Science, Cambridge, MA: Harvard University Press, 1984. Available at: https://www.hup.harvard.edu/books/9780674576339. Accessed on: 8 Sep. 2025.

SETH, A. K.; BAYNE, T. Theories of consciousness. **Nature Reviews Neuroscience**, v. 23, n. 7, p. 439–452, 2022. Available at: https://doi.org/10.1038/s41583-022-00587-4. Accessed on: 8 Sep. 2025.

SGANTZOS, K.; STELIOS, S.; TZAVARAS, P.; THEOLOGOU, K. Minds and machines: evaluating the feasibility of constructing an advanced artificial intelligence. **Discover Artificial Intelligence**, v. 4, art. 104, 2024. Available at: https://doi.org/10.1007/s44163-024-00216-2. Accessed on: 8 Sep. 2025.

SHANAHAN, M. Talking about Large Language Models. **Communications of the ACM**, v. 67, n. 2, p. 68–79, 2024. Available at: https://doi.org/10.1145/3624724. Accessed on: 8 Sep. 2025.

SHASHKOVA, L. Scientific communication in complex social contexts: approaches of social philosophy of science and social epistemology. **Proceedings of the National Aviation University**. Series: Philosophy. Culturology, v. 39, n. 1, p. 23-28, 2024. Available at: https://doi.org/10.18372/2412-2157.39.18442. Accessed on: 8 Sep. 2025.

STRACHAN, J. W. A.; ALBERGO, D.; BORGHINI, G.; et al. Testing theory of mind in large language models and humans. **Nature Human Behaviour**, v. 8, p. 1285–1295, 2024. Available at: https://doi.org/10.1038/s41562-024-01882-z. Accessed on: 8 Sep. 2025.

TADDEO, M. & BLANCHARD, A. A comparative analysis of the definitions of autonomous weapons systems. **Science and Engineering Ethics**, v. 28, art. 37, 2022. Available at: https://link.springer.com/article/10.1007/s11948-022-00392-3. Accessed on: 8 Sep. 2025.

TAO, Y.; VIBERG, O.; BAKER, R. S.; KIZILCEC, R. F. Cultural bias and cultural alignment of large language models. **PNAS Nexus**, v. 3, n. 9, 2024.

Available at: https://doi.org/10.1093/pnasnexus/pgae346. Accessed on: 8 Sep. 2025.

TRUBA, H.; RADZIIEVSKA, I.; SHERMAN, M.; DEMCHENKO, O.; KULICHENKO, A.; HAVRYLIUK, N. Introduction of Innovative Technologies in Vocational Education Under the Conditions of Informatization of Society: Problems and Prospects. **Conhecimento & Diversidade**, v. 15, n. 38, p. 443–460, 2023. Available at: https://doi.org/10.18316/rcd.v15i38.11102. Accessed on: 8 Sep. 2025.

TSEKHMISTER, Y.; KONOVALOVA, T.; BASHKIROVA, L.; SAVITSKAYA, M.; TSEKHMISTER, B. Virtual Reality in EU Healthcare: Empowering Patients and Enhancing Rehabilitation. **Journal of Biochemical Technology**, v. 14, n. 3, p. 23–29, 2023. Available at: https://doi.org/10.51847/r5WJFVz1bj. Accessed on: 8 Sep. 2025.

TURING, A. M. Computing machinery and intelligence. **Mind**, v. LIX, n. 236, p. 433–460, 1950. Available at: https://doi.org/10.1093/mind/LIX.236.433. Accessed on: 8 Sep. 2025.

VACCARO, M.; ALMAATOUQ, A.; MALONE, T. W. When combinations of humans and AI are useful: a systematic review and meta-analysis. **Nature Human Behaviour**, v. 8, p. 2293–2303, 2024. Available at: https://doi.org/10.1038/s41562-024-02024-1. Accessed on: 8 Sep. 2025.

VEALE, M.; BORGESIUS, Z. F. Demystifying the Draft EU Artificial Intelligence Act – Analysing the good, the bad, and the unclear elements. **Computer Law Review International**, v. 22, n. 4, p. 97–112, 2021. Available at: https://doi.org/10.9785/cri-2021-220402. Accessed on: 8 Sep. 2025.

VERED, M.; LIVNI, T.; HOWE, P. D. L.; MILLER, T.; SONENBERG, L. The effects of explanations on automation bias. **Artificial Intelligence**, v. 320, art. 103952, 2023. Available at: https://doi.org/10.1016/j.artint.2023.103952. Accessed on: 8 Sep. 2025.

WAYTZ, A.; HEAFNER, J.; EPLEY, N. The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. **Journal of Experimental Social Psychology**, v. 52, p. 113–117, 2014. Available at: https://doi.org/10.1016/j.jesp.2014.01.005. Accessed on: 8 Sep. 2025.

Acknowledgement and conflicts of interest

The authors declare that they have no conflicts of interest concerning the research, authorship, and/or publication of this article.