Category Error and Epistemic Artifacts in AI-Assisted Symbolic Cognitive Systems

A methodological analysis of epistemic misclassification

Edervaldo José de Souza Melo
edersouzamelo@gmail.com
ORCID: 0009-0003-6835-135X

December 2025

Abstract

This paper examines the epistemic status of AI-assisted symbolic cognitive systems and argues that such systems are frequently subject to a persistent category error. Rather than being assessed as empirical theories of cognition, complete computational architectures, or psychological models, these systems are more appropriately understood as epistemic artifacts with an intermediate methodological function. Drawing on classical contributions from the philosophy of science and cognitive science—including research programmes, heuristic reasoning, and distributed cognition—the paper clarifies the type of epistemic work these systems perform. It proposes criteria for their proper evaluation, emphasizing internal coherence, organizational power, and conceptual clarity, while explicitly rejecting demands for direct empirical validation or computational performance as methodologically misplaced. By repositioning AI-assisted symbolic systems within the landscape of scientific inquiry, the paper aims to enable more accurate criticism, responsible development, and cumulative theoretical progress.

1 Introduction — The Problem of Epistemic Misclassification

Contemporary discussions surrounding AI-assisted symbolic cognitive systems are frequently marked by a recurring pattern of epistemic misclassification. Such systems are often evaluated according to criteria appropriate to empirical theories, computational architectures, or psychological models, leading to judgments framed almost exclusively in terms of effectiveness, performance, or empirical validation. When assessed under these standards, they are commonly dismissed as incomplete, untestable, or methodologically weak.

This paper argues that these judgments rest on a persistent category error. The problem does not primarily concern whether such systems *work* in an empirical or computational sense, but whether they are being evaluated as the kind of epistemic objects they actually are. Treating all cognitive systems as if they were intended to function as theories of mind or predictive models imposes inappropriate standards of assessment and obscures their genuine methodological role.

The history of science offers numerous examples of epistemic instruments whose primary function was not to deliver immediate empirical confirmation, but to organize inquiry, guide conceptual exploration, and structure problem spaces prior to formal testing. Frameworks, research programmes, heuristics, and representational tools have repeatedly played decisive roles in scientific development without themselves constituting empirical theories in the strict sense. To judge such instruments by criteria reserved for fully articulated theories is to misunderstand their place within scientific practice.

AI-assisted symbolic cognitive systems, this paper suggests, are best understood within this intermediate epistemic category. Their contribution lies not in producing direct empirical claims about cognition, nor in functioning as complete computational models, but in shaping how cognitive phenomena are described, explored, and rendered intelligible in hybrid human—machine contexts. When evaluated according to inappropriate standards, their epistemic function is systematically obscured.

The objective of this paper is therefore classificatory and methodological rather than empirical or technical. By clarifying the epistemic status of AI-assisted symbolic cognitive systems, it seeks to establish appropriate criteria for their evaluation and to delineate the limits of legitimate criticism. In doing so, the paper aims to reduce misplaced objections, enable more precise scientific dialogue, and support cumulative and responsible development across philosophy, cognitive science, and engineering.

2 Lessons from the History of Science: Intermediate Epistemic Instruments

Historical analyses of scientific practice reveal that not all epistemically valuable instruments take the form of fully articulated empirical theories. The development of science has repeatedly relied on conceptual frameworks, research programmes, and heuristic structures whose primary role was to organize inquiry rather than to generate immediately testable predictions. Confusing such instruments with theories in the strict sense has long been recognized as a source of methodological error.

A particularly instructive account is provided by Lakatos's distinction between isolated theories and scientific research programmes. For Lakatos, the rational appraisal of scientific work cannot be reduced to the success or failure of individual hypotheses; instead, it requires attention to the broader methodological structures that guide problem selection, theoretical modification, and long-term inquiry. Research programmes, in this sense, function as organizing devices that structure scientific activity over time, even when particular theoretical components remain underdeveloped or empirically inconclusive (Lakatos, 1978).

Crucially, Lakatos does not treat research programmes as empirical theories themselves. They do not directly compete at the level of prediction or falsification, but rather provide a framework within which such competition becomes meaningful. To evaluate a research programme solely by the immediate empirical success of its constituent theories is therefore to misapply criteria and to overlook its epistemic function. The significance of this distinction lies not in elevating programmes above criticism, but in recognizing that different epistemic objects demand different standards of assessment.

Beyond Lakatos, the history of science offers numerous examples of intermediate epistemic instruments that preceded or accompanied empirical consolidation. Conceptual schemata, classificatory frameworks, and representational devices have often enabled scientific progress by stabilizing vocabularies, delineating problem spaces, and coordinating inquiry across domains. As Kuhn famously emphasized, periods of normal science depend on shared conceptual and methodological frameworks that guide research prior to, and independently of, decisive empirical breakthroughs (Kuhn, 1962).

These historical considerations suggest that epistemic value cannot be exhaustively measured by empirical performance alone. Instruments that organize, guide, or constrain inquiry occupy a legitimate methodological role even in the absence of direct validation. Recognizing this role is a prerequisite for avoiding category errors in the evaluation of contemporary cognitive systems that similarly operate at an intermediate epistemic level.

3 Heuristics Without Truth: Methodological Utility

The tendency to demand direct empirical validation from all cognitive instruments reflects a narrow conception of epistemic value. Within both philosophy of science and cognitive science, it has long been recognized that usefulness and truth do not coincide in a simple or universal manner. Heuristic devices, in particular, can play a decisive epistemic role without purporting to offer accurate or exhaustive representations of the phenomena they address.

Gigerenzer's work on heuristic reasoning provides a clear articulation of this point. Against the assumption that rationality consists primarily in approximation to normative models of optimal inference, Gigerenzer emphasizes the concept of ecological rationality: the adequacy of a cognitive strategy depends on how well it performs within a given environment, not on its conformity to abstract standards of correctness. Heuristics, on this view, are not defective theories awaiting completion, but context-sensitive tools whose value lies in their capacity to guide action and judgment under conditions of uncertainty (Gigerenzer, 2000, 2014).

This perspective has important methodological implications. If heuristics are evaluated according to criteria appropriate to theories—such as completeness, generality, or representational accuracy—they will inevitably appear deficient. Yet such evaluations miss the point of heuristic instruments altogether. Their epistemic contribution consists in simplifying problem spaces, directing attention, and enabling tractable forms of reasoning where exhaustive modeling is neither possible nor desirable.

AI-assisted symbolic cognitive systems can be understood in a similar manner. Their function is not to provide true descriptions of cognitive mechanisms, nor to deliver predictive models subject to straightforward empirical testing. Instead, they operate as heuristic and organizational devices that structure inquiry, scaffold reasoning, and render complex cognitive domains more intelligible. To require direct empirical validation or computational performance from such systems is therefore to commit a methodological error analogous to judging heuristics by the standards of theories.

Recognizing the heuristic character of these systems does not exempt them from evaluation, but it shifts the criteria by which they should be assessed. Coherence, transparency, and their capacity to orient inquiry become central, while demands for truth, prediction, or optimization are revealed as misplaced. This shift prepares the ground for understanding AI-assisted symbolic systems as epistemic artifacts situated between theory and application, rather than as failed attempts at empirical science.

4 Hybrid Cognitive Systems and Distributed Cognition

Accounts of cognition that locate cognitive processes exclusively within individual minds have been increasingly challenged over the past decades. Work on distributed and extended cognition has emphasized that cognitive activity often emerges from configurations involving agents, artifacts, representations, and environments. From this perspective, cognition is not confined to an internal locus, but is enacted through structured interactions between humans and material or symbolic supports.

Hutchins's analyses of navigation and coordinated activity illustrate how cognitive processes can be distributed across individuals and artifacts without being reducible to any single component. Rather than locating cognition within an isolated agent, this approach emphasizes how representational structures, material tools, and social coordination jointly realize cognitive functions (Hutchins, 1995). Similarly, extended cognition accounts emphasize that tools, symbols, and representational systems can become integral parts of cognitive activity when they reliably participate in reasoning, memory, or problem-solving. On this view, cognitive processes may extend beyond the biological boundaries of the individual when external structures are appropriately integrated into cognitive routines; what matters epistemically is the functional role such structures play within coordinated cognitive activity (Clark, 2008).

AI-assisted symbolic cognitive systems are naturally situated within this hybrid landscape. Their epistemic role cannot be adequately characterized by asking whether cognition occurs "in the human" or "in the machine." Instead, what matters is how symbolic structures, computational support, and human interpretation jointly contribute to the organization and execution of cognitive tasks. The relevant unit of analysis is therefore the configuration itself, not its individual components taken in isolation.

This configurational perspective has important implications for evaluation. When a system is treated as if it were a self-contained computational model, its reliance on human interpretation may appear as a defect. Conversely, when it is judged as a purely human cognitive practice, the presence of artificial support may be dismissed as auxiliary or superficial. Both assessments rest on a failure to recognize the hybrid nature of the system and the distributed character of the cognitive activity it enables.

Understanding AI-assisted symbolic systems as hybrid cognitive configurations aligns them with familiar epistemic tools such as maps, diagrams, formal notations, and modeling languages. These instruments do not think or reason independently, yet they profoundly shape cognitive processes by constraining representations, stabilizing inferences, and coordinating action. Their epistemic significance lies not in autonomy or internal processing power, but in their capacity to structure cognition across human–artifact assemblages. This perspective provides a conceptual bridge between heuristic utility and the classificatory analysis developed in the following section.

5 The Epistemic Status of AI-Assisted Symbolic Cognitive Systems

Clarifying the epistemic status of AI-assisted symbolic cognitive systems requires a negative and positive characterization. Rather than asking what such systems are capable of achieving empirically or computationally, the appropriate task at this stage is to determine the category of epistemic object to which they belong. This section therefore proceeds by distinguishing what these systems are not from what they are.

What They Are Not

AI-assisted symbolic cognitive systems should not be construed as empirical theories of cognition. They do not advance testable hypotheses about cognitive mechanisms, nor do they aim to explain cognitive phenomena through predictive or causal models. Evaluating them according to standards of empirical adequacy or falsifiability therefore misrepresents their intended epistemic role.

Nor should these systems be understood as complete computational architectures. They do not constitute self-contained models capable of autonomous cognitive processing, nor do they aspire to algorithmic completeness or optimization. Their reliance on human interpretation and interaction is not a limitation to be overcome, but a defining feature of their design.

Finally, AI-assisted symbolic cognitive systems are not clinical, diagnostic, or predictive instruments. They are not intended to deliver validated assessments of psychological states, behavioral outcomes, or cognitive performance. Treating them as such risks both methodological confusion and inappropriate application.

What They Are

Properly understood, AI-assisted symbolic cognitive systems function as epistemic artifacts. Their primary contribution lies in organizing, stabilizing, and rendering intelligible complex cognitive domains through symbolic representation and structured interaction. They shape how phenomena are described and explored rather than asserting claims about how those phenomena fundamentally operate.

More specifically, these systems operate as descriptive and organizational schemes. They provide vocabularies, classificatory structures, and representational constraints that guide inquiry and facilitate communication across hybrid human–machine contexts. Their value emerges from their capacity to coordinate reasoning, not from their ability to generate empirical results.

As epistemic artifacts, AI-assisted symbolic cognitive systems occupy an intermediate methodological position. They mediate between conceptual exploration and empirical investigation, enabling the formulation of questions, distinctions, and problem spaces that may later be subjected to empirical or technical development. Recognizing this status is essential for applying appropriate standards of evaluation and for avoiding persistent category errors in their assessment.

6 Implications for Scientific Evaluation

Once the epistemic status of AI-assisted symbolic cognitive systems is properly clarified, the question of evaluation can be reformulated in a methodologically coherent manner. If such systems are understood as epistemic artifacts occupying an intermediate position between conceptual exploration and empirical investigation, then the criteria by which they are assessed must reflect this status.

Appropriate standards of evaluation should focus on internal coherence, organizational power, and conceptual clarity. Internal coherence concerns the consistency of the symbolic structures and distinctions the system employs. Organizational power refers to the system's capacity to structure problem spaces, coordinate inquiry, and support meaningful exploration within hybrid human–machine contexts. Conceptual clarity concerns the transparency of the categories, assumptions, and representational commitments that the system introduces.

By contrast, several commonly applied criteria are methodologically misplaced when directed at systems of this kind. Demands for direct empirical validation, predictive accuracy, or computational performance presuppose that the system is intended to function as an empirical theory, a technical architecture, or an autonomous model. When such presuppositions are absent, evaluations based on these criteria amount to category errors rather than substantive critiques.

This distinction does not imply that AI-assisted symbolic cognitive systems are insulated from criticism. On the contrary, recognizing their epistemic role enables more precise and productive forms of critique. Questions concerning conceptual adequacy, heuristic usefulness, and the potential for misrepresentation or overextension remain fully legitimate. What is excluded are criticisms that rest on inappropriate expectations regarding empirical output or technical implementation.

Clarifying evaluative standards in this way contributes to a more constructive relationship between philosophy, cognitive science, and engineering. It allows different forms of inquiry to coexist without collapsing their respective aims into a single metric of success, thereby supporting cumulative and responsible development across disciplinary boundaries.

7 Conclusion — Why Category Errors Matter

This paper has argued that AI-assisted symbolic cognitive systems are frequently subject to a persistent category error in contemporary scientific and philosophical discussions. When such systems are evaluated as if they were empirical theories, complete computational architectures, or psychological models, their epistemic role is systematically misunderstood. The resulting critiques, while often presented as substantive, rest on misplaced standards of assessment.

By drawing on insights from the history and philosophy of science, heuristic reasoning, and distributed cognition, the paper has proposed an alternative classification of these systems as epistemic artifacts occupying an intermediate methodological position. Their primary function is not to generate empirical claims or predictive models, but to organize inquiry, structure conceptual exploration, and support intelligibility within hybrid human–machine contexts.

Clarifying this epistemic status has practical consequences. It enables more precise forms of criticism, grounded in coherence, organizational capacity, and conceptual clarity, while excluding demands that presuppose aims the systems do not possess. This distinction does not weaken scientific rigor; rather, it preserves it by aligning evaluative criteria with the nature of the epistemic objects under consideration.

More broadly, avoiding category errors is essential for productive interaction between philosophy, empirical science, and engineering. When different forms of inquiry are assessed according to inappropriate standards, dialogue collapses into misunderstanding or dismissal. As Velmans has argued in the context of consciousness studies, methodological confusion often arises when explanatory levels are conflated or when theoretical frameworks are evaluated outside the domains for which they were designed (Velmans, 2009). By contrast, recognizing the diversity of epistemic instruments and their respective roles allows for cumulative and responsible development across disciplinary boundaries. In this sense, the clarification offered here is not merely classificatory, but a prerequisite for more coherent scientific practice in increasingly hybrid cognitive landscapes.

References

Lakatos, I. (1978). *The Methodology of Scientific Research Programmes.* Cambridge: Cambridge University Press.

Gigerenzer, G. (2000). *Adaptive Thinking: Rationality in the Real World.* Oxford: Oxford University Press.

Gigerenzer, G. (2014). *Risk Savvy: How to Make Good Decisions.* New York: Viking.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

Hutchins, E. (1995). *Cognition in the Wild.* Cambridge, MA: MIT Press.

Clark, A. (2008). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension.* Oxford: Oxford University Press.

Velmans, M. (2009). *Understanding Consciousness.* London: Routledge.