Where Does Epistemic Capacity Reside? Functional Externalization in Generative AI Systems

Jose-Manuel Vega-Baez
Tecnologico de Monterrey, Campus Laguna, School of Business, Torreon, Mexico

jmvegabaez@tec.mx

Abstract: Generative AI systems are increasingly used to retrieve, synthesize, evaluate, and communicate information in knowledge-intensive tasks. Yet observed epistemic performance is often attributed either to the human user or to the AI model, even when it depends on a wider configuration of models, retrieval mechanisms, external sources, tools, interfaces, permissions, and human oversight. This paper asks under what conditions epistemic capacity should be attributed to a configured system rather than to either the human or the model alone, and what follows from locating that capacity. Drawing on distributed cognition, extended and social epistemology, and work on epistemic technologies, the paper develops a task-relative attribution framework grounded in functional externalization. For a specified task class, operating conditions, and acceptable performance criterion, it identifies a minimal adequate capacity-bearing configuration as an organized set of functionally integrated components sufficient to realize the relevant capacity, such that no proper subconfiguration remains sufficient under the same specification. Minimality does not imply uniqueness: different configurations may be equally minimal and adequate. Through comparative functional reconstruction of context-bounded synthesis, retrieval-augmented generation, and tool-augmented systems, the paper distinguishes operational components from epistemic resources and background conditions and shows why individual component necessity is not required when redundant routes realize the same functional role. The paper advances two claims. First, epistemic capacity should be attributed at the minimal organized level required to explain its realization across the relevant class of tasks, rather than inferred from the most visible component. Second, locating that capacity does not determine the loci of understanding, epistemic authority, purpose, control, authorship, or responsibility. These properties may coincide with, overlap, or diverge from the capacity-bearing configuration, but only for independent reasons. The contribution is therefore not to rediscover distributed cognition or sociotechnical epistemic practice, but to provide a deliberately minimal framework for disciplined attribution without presupposing a theory of extended mind, artificial understanding, or epistemic subjecthood.

Keywords: Epistemic capacity, Functional externalization, Generative AI, Distributed cognition, Social epistemology

1. Introduction: The Attribution Problem

A researcher asks a generative AI system to synthesize evidence about a current question. A fluent answer appears, perhaps accompanied by retrieved documents, citations, or the result of an external calculation. At the interface, the interaction appears simple: a person asks; an AI answers. Functionally, however, the trajectory may involve a language model, retrieval mechanisms, external information resources, specialized tools, access permissions, system instructions, and human decisions concerning acceptable evidence or verification.

This creates an attribution problem. When we say that "the AI" found the evidence, evaluated the information, or answered the question, are we identifying the bearer of the relevant capacity or merely naming the most visible point of interaction? Interface unity does not establish functional unity.

The problem is not that epistemic activity has suddenly become distributed. Hutchins' analysis of ship navigation showed that a cognitive capacity may depend on coordinated transformations across people, representations, artifacts, and procedures (Hutchins, 1995). Extended cognition and extended epistemology subsequently examined when external resources may be constitutive rather than merely supportive of cognition and knowledge (Clark and Chalmers, 1998; Palermos, 2014). Social epistemology has likewise demonstrated the depth of epistemic dependence on testimony, specialization, and other people's competence (Hardwig, 1985).

Recent work on generative AI brings this problem into specifically contemporary epistemic practice. Bareither and Griessl (2026) describe "hybrid epistemic practices" in which GenAI becomes entangled with academic reading, interpretation, and writing. Cabitza and Colombo (2026) similarly argue that LLM-mediated knowledge practices cannot be adequately understood by isolating the model from the wider circuits of interpretation, verification, use, and responsibility in which outputs acquire epistemic significance.

This paper begins one step later. Once a task-relative epistemic capacity depends on a wider configuration, how should its attribution boundary be drawn? Under what conditions should capacity be attributed to a configured system rather than to the human user or AI model alone?

A second problem follows. Suppose that the capacity has been correctly located. Have we thereby identified who understands the result, establishes the purpose, warrants epistemic deference, controls the process, qualifies for authorship, or bears responsibility? I argue that we have not.

The paper proposes a two-stage framework. First, it identifies a minimal adequate capacity-bearing configuration relative to a specified task class, operating conditions, and acceptable performance criterion. Second, it treats understanding, epistemic authority, purpose, control, authorship, and responsibility as distinct attribution problems. Their loci may coincide with the capacity-bearing configuration, overlap with it, or lie elsewhere—but coincidence requires independent justification.

2. From Distributed Epistemic Activity to Attribution

Distributed cognition provides the clearest precedent for expanding the explanatory unit beyond an isolated individual. In Cognition in the Wild, Hutchins (1995) analyses navigation as a system in which information is transformed across people, representational artifacts, instruments, and procedures. A capacity of the organized whole may therefore be inadequately explained by assigning it to any participant taken alone.

Extended cognition poses a stronger question. Clark and Chalmers (1998) argue that, under appropriate conditions, environmental resources may actively participate in cognitive processes rather than function merely as aids to an internally bounded mind. Palermos (2014) brings this issue into epistemology by connecting cognitive integration with knowledge-producing processes. These approaches expose a boundary problem: causal participation does not by itself determine whether an external element is supportive, functionally integrated, or constitutive of cognition.

The present paper does not attempt to resolve that metaphysical question. Its concern is narrower: explanatory attribution of a specified epistemic capacity. Functional relevance may therefore be recognized without deciding whether a technical component literally belongs to a mind.

Social epistemology provides a parallel lesson. Hardwig (1985) shows that justified belief and scientific knowledge can depend deeply on the testimony and expertise of others. Palermos (2022a) further argues that sufficiently integrated epistemic collaborations can warrant genuinely collective epistemic attribution. This matters because the present thesis is not that different properties must always have different bearers. Their loci can coincide; the point is that such coincidence cannot be inferred merely from the existence of a shared functional capacity.

AI adds another layer. Alvarado (2023) characterizes AI systems as epistemic technologies because they perform operations over epistemically relevant content within inquiry. Gouveia and Wang (2026) nevertheless distinguish significant cognitive contribution from the stronger status of cognitive subjecthood, while Hauswald (2025) separately examines whether artificial systems may warrant epistemic authority.

The framework proposed here differs from two nearby approaches in a specific way. Kaplan (2012) addresses the demarcation of components within cognitive mechanisms and therefore engages a constitutive question: what belongs to the mechanism itself? Cabitza and Colombo (2026), by contrast, broaden the unit of analysis from the isolated LLM toward the wider epistemic practice in which outputs are produced and evaluated.

The present approach is deliberately more limited than either. It asks neither what constitutes cognition nor what fully constitutes an epistemic practice. Instead, it asks which subset of a sociotechnical environment must be included to explain a specified capacity under specified conditions. The gain is analytical parsimony and applicability to deployed systems without first resolving disputed questions about mind, agency, or subjecthood. The corresponding loss is deliberate: the framework cannot, by itself, determine whether the resulting configuration constitutes a cognitive system, an epistemic agent, or a complete unit of sociotechnical practice.

3. Functional Externalization and Levels of Attribution

3.1 Epistemic capacity, Performance, and Specification

For present purposes, epistemic capacity is a task-relative disposition to perform operations over epistemically relevant information in accordance with standards pertinent to inquiry, such as relevance, accuracy, evidential adequacy, consistency, or appropriate source integration.

Such operations may include retrieving evidence, comparing claims, transforming representations, integrating sources, assessing information, or producing a warranted synthesis. The definition is deliberately non-factive.

Possessing a capacity does not imply succeeding on every occasion, just as human investigative competence does not disappear whenever a researcher makes an error.

Capacity must therefore be distinguished from performance. A particular output is an exercise of a capacity under particular conditions. Altering one document and obtaining a different answer does not establish that the underlying capacity has changed. What matters is whether a change systematically affects the performance profile across the relevant class of tasks.

Attribution consequently requires prior specification of three elements:

1. Task class: what kind of epistemic task is being attributed?
2. Operating conditions: what resources, interfaces, permissions, and environmental constraints are assumed?
3. Acceptable performance criterion: what level or profile of performance counts as possession of the relevant capacity?

Without these elements, "where does the capacity reside?" is under-specified.

3.2 Functional Externalization

Functional externalization occurs when part of the task-relevant load required to represent, preserve, retrieve, transform, assess, integrate, or coordinate epistemically relevant information is borne by an extra-individual substrate or organized process, thereby reducing dependence on the memory, presence, or immediate intervention of particular individuals.

The notion of functional externalization used here was developed in broader form in Fuera de Nosotros, which examines how representational, operational, and coordinative loads can shift to extra-individual supports across historical and technological settings (Vega-Baez, 2026). The present paper narrows and formalizes that notion for the attribution of epistemic capacity in generative AI systems.

The concept makes a deliberately minimal claim. It does not determine where a mind ends, whether a technical component understands what it processes, or whether a sociotechnical configuration qualifies as an epistemic agent. It asks a narrower question: where is part of the work required for the task being borne?

Different configurations externalize different operations. A model may bear part of the load of comparing evidence already supplied in context. A retrieval subsystem may bear part of the work of locating representations relevant to a query. A calculator may bear a numerical operation required in a larger inquiry. An orchestration layer may bear part of the work of selecting among specialized processes.

Functional externalization therefore identifies a change of support. The attribution framework then asks which of those externally supported operations must be included in the bearer of the overall capacity.

3.3 Components, Resources, and Background Conditions

Every deployed AI system depends causally on a broader sociotechnical environment: hardware, networks, software libraries, organizations, developers, institutions, and users all contribute to its existence. Yet causal relevance alone cannot define the bearer of a capacity. Otherwise, every attribution boundary would expand indefinitely.

Three roles should therefore be distinguished.

- A component performs or mediates a task-relevant operation that is organizationally integrated into realization of the specified capacity.
- A resource supplies content or state on which such operations depend, without necessarily performing the relevant operation itself.
- A background condition enables the activity but does not enter the organized functional trajectory through which the specified capacity is realized.

These are role-relative categories, not permanent properties of objects. A corpus may function as an epistemic resource in one configuration. A dynamically updating corpus-management process may become an operational component in another. A permission may be a background precondition when granted once before a task, but a task-relevant control operation when dynamically evaluated during execution.

This distinction prevents difference-making from doing too much work. Changes to resources and background conditions can influence performance without converting those elements into components of the capacity-bearing configuration.

3.4 Minimal adequate Capacity-bearing Configurations

For a specified task class, operating conditions, and acceptable performance criterion, a minimal adequate capacity-bearing configuration is an organized set of functionally integrated components sufficient to realize the relevant capacity, such that no proper subconfiguration remains sufficient under the same specification.

Minimality is therefore defined by inclusion, not by the numerical count of components or by physical size.

It also does not imply uniqueness. Two non-equivalent configurations can each be minimal and adequate if each realizes the specified capacity and neither contains a dispensable proper subconfiguration. The framework therefore seeks a minimal adequate configuration, not necessarily the unique one.

Four considerations guide attribution:

- Task-relevant functional role: the candidate performs or mediates an operation required by the specified capacity.
- Organized incorporation: its output, state, or constraint enters in a structured way into realization of the capacity.
- Intervention evidence: altering or removing the relevant functional role or subsystem systematically affects the capacity profile under the specified task class and conditions.
- Boundary parsimony: elements unnecessary to explain that capacity are excluded even when they remain causally important to deployment.

Intervention evidence is thus neither necessary at the level of every individual component nor sufficient on its own. Redundant components may be individually dispensable while the functional role they jointly realize remains necessary.

Finally, the sociotechnical deployment environment and the capacity-bearing configuration need not have identical boundaries. An organization may set goals or control permissions without belonging to the minimal configuration that performs retrieval and synthesis once a query has been supplied. Conversely, human verification may become part of the capacity-bearing configuration if meeting the defined performance criterion systematically depends upon it.

4. Comparative Functional Reconstruction

The framework can be examined through three configurations chosen analytically because they progressively relocate operations required for epistemic tasks: from processing evidence already supplied, to retrieving evidence externally, to executing specialized external operations.

4.1 Context-bounded Synthesis

Consider a researcher who supplies several passages to a language model and asks it to compare their evidence and summarize agreements and conflicts. The relevant capacity is not generic linguistic generation but context-bounded synthesis of supplied evidence.

Functional externalization concerns part of the work of transforming and relating those representations. If all relevant evidence is provided directly and no external operation is required, the capacity-bearing configuration can remain comparatively narrow: the deployed model operating over the supplied context.

The texts themselves are resources. They influence the quality of a particular performance, but they do not thereby become operational components of the synthesis mechanism.

4.2 Retrieval-augmented Generation

Retrieval-augmented generation changes the functional distribution. Lewis et al. (2020) combine parametric generation with retrieval from external non-parametric memory for knowledge-intensive NLP tasks.

The relevant capacity may be specified as retrieving and synthesizing relevant evidence from an externally maintained information resource. Functional externalization now includes retrieval itself: part of the load of locating evidence is borne by a retrieval subsystem.

The distinction among component, resource, and background condition becomes important. The retriever performs an operation and therefore belongs to the operational configuration. The corpus normally functions as a resource whose coverage and quality condition the attainable performance profile. Particular retrieved documents are inputs to individual executions. Access permissions may remain background conditions when statically granted, but could become operational components if permissions are dynamically evaluated as part of execution.

Changing a document can change an answer without changing the system's retrieval capacity. By contrast, removing or disabling the retrieval role should systematically reduce performance on a task class requiring external information. For that capacity, model-centric attribution is therefore incomplete.

4.3 Tool-augmented Inquiry

Toolformer trains a model to decide which APIs to invoke, when to invoke them, what arguments to supply, and how to incorporate the results (Schick et al., 2023). ReAct similarly interleaves language-model processing with actions that generate external observations used in subsequent steps (Yao et al., 2023).

Suppose an evidential task requires a precise numerical operation. The model may interpret the request and decide to invoke a calculator; the calculator performs the numerical transformation; the result is then returned to the model for integration.

Functional externalization occurs at several points. The model bears part of the interpretive and integrative load, while the calculator bears the mathematical operation. If reliable performance on the defined task class depends on this division of labour, neither "the model calculated the answer" nor "the calculator answered the question" adequately characterizes the capacity.

The relevant configuration includes the functionally integrated model-mediated orchestration, tool operation, and interface required for the result to re-enter the trajectory.

4.4 Hard Cases: Redundancy and Contingent Verification

Straightforward decomposition is not enough to test the attribution rule. Consider a system with two interchangeable retrieval routes, R1 and R2, either of which can perform the same required retrieval function.

Removing R1 does not materially reduce performance because R2 takes over. Removing R2 likewise leaves the capacity intact. If component membership required individual counterfactual necessity, neither retriever would count as part of the bearer—an implausible result.

The correct unit of intervention is therefore sometimes the functional role or functionally equivalent subsystem, not each redundant component taken separately. Removing retrieval as a role reduces performance below the required threshold even though no individual retrieval route is indispensable.

Now add a human verifier, H, who intervenes only when the system's confidence or evidence quality falls below a defined threshold. Whether H belongs to the capacity-bearing configuration depends on the capacity specification.

If the task class concerns routine cases that meet the required performance criterion without human intervention, H need not belong to the minimal configuration for that capacity.

If, however, the specified capacity requires reliable performance across ambiguous cases and attainment of the threshold systematically depends on H's verification, then H may belong to a minimal adequate configuration.

If H instead performs only an independent governance or approval function after the epistemic task has already satisfied its performance criterion, H may remain outside the capacity-bearing configuration while still being central to control or responsibility.

This example demonstrates why attribution cannot be reduced either to simple component removal or to inclusion of the whole surrounding practice.

4.5 Orchestrated Systems

Kaplunovich (2025) provides a contemporary example from ICAIR. His question-answering architecture combines keyword search, vector similarity, knowledge-graph querying, summarization, LangGraph orchestration, multiple LLMs, and ensemble verification.

The relevance of this architecture is not that modularity itself proves system-level epistemic capacity. It is that several competing attribution boundaries become possible. A model-centred description omits retrieval, graph operations, orchestration, and verification. A description of the entire deployment ecosystem includes much more than is needed to explain the QA capability.

The proposed framework instead asks which organized configuration realizes the specified QA capacity under its defined conditions. Redundant routes and verification mechanisms can then be assessed by functional role rather than by visibility or individual indispensability.

5. Different Properties, Different Loci

Assume that an appropriate capacity-bearing configuration has been identified. What exactly has been located? A capacity.

That conclusion does not by itself locate understanding, epistemic authority, purpose, control, authorship, or responsibility. Their loci may coincide with, overlap with, or diverge from the capacity-bearing configuration. What matters is that each attribution requires grounds appropriate to the property in question.

Table 1: Distinct attribution questions and their relevant grounds

Property	Attribution question	Relevant grounds
Epistemic capacity	What organized configuration can perform the specified task across the relevant conditions?	Functional realization and integration
Understanding	Who possesses the relevant understanding of content, reasons, or implications?	Cognitive and epistemic criteria
Epistemic authority	To whom is epistemic deference justified?	Competence, reliability, evidence, trust
Purpose	Who sets or structures what the activity is for?	Goal-setting and institutional direction
Control	Who can authorize, modify, constrain, suspend, or revoke?	Effective governance powers
Authorship	Who merits intellectual attribution?	Norms of contribution and epistemic achievement
Responsibility	Who ought to answer for the process or outcome?	Duties, knowledge, authority, control, and participation

5.1 Understanding and Epistemic Authority

Capacity is weaker than understanding. Gouveia and Wang (2026) distinguish significant cognitive contribution by generative AI from the stronger status of cognitive subjecthood. Nothing in the present framework settles that controversy. A retrieval-and-generation configuration may exhibit a stable capacity profile without thereby possessing understanding in whatever philosophically relevant sense is ultimately defended.

Epistemic authority requires a different argument. Hauswald (2025) examines whether artificial systems may warrant epistemic deference despite lacking some properties normally associated with human testimony. The issue is not simply whether a system causally contributes to a correct result, but whether reliance upon it is justified.

Norkin and Kirchner (2025) illustrate this distinction in organizational settings. Their interviews with managers in GenAI-augmented knowledge-intensive organizations associate greater trust with employees who use GenAI critically and can question, understand, and explain its outputs. Their findings do not locate epistemic authority philosophically, but they show that AI-supported productive capacity does not eliminate the epistemic relevance of human assessment.

5.2 Purpose and Control

Purpose follows another map. A configured system can retrieve and integrate evidence without determining why an inquiry should occur, which objective should receive priority, or which outcome counts as institutionally acceptable.

The user entering a prompt need not exhaust the locus of purpose. Research protocols, organizational policies, professional standards, or managerial decisions may structure the task before interaction begins.

Control is also distinct from proximal use. The person receiving an answer may not choose the indexed sources, change retrieval rules, grant credentials, modify system instructions, update the model, or suspend deployment.

Kabir's (2025) distinction among Model Stewardship, Operational Alignment, and Strategic Guardrails makes such organizational layering visible. The configuration that produces the answer need not coincide with the actors or structures that establish objectives, permissions, constraints, or powers of termination.

5.3 Authorship and Responsibility

Authorship cannot be inferred from causal production alone. Rodrigues (2026) argues that AI-assisted authorship may involve higher-order epistemic competences such as design, refinement, verification, and synthesis rather than direct composition of every sentence. Whether one accepts his full account, the example demonstrates that authorial attribution depends on norms of intellectual contribution rather than merely on the mechanism that generated the text.

Responsibility requires still another analysis. Palermos (2022b) asks whether responsibility within epistemic collaborations may belong to individuals, groups, or both. This reinforces a central point: even collective responsibility, where justified, requires criteria beyond collective functional capacity.

In AI-mediated settings, responsibility may depend on obligations, knowledge, authority, foreseeable consequences, effective control, and reasonable opportunities to intervene. A causal or functional map can inform such judgments, but it cannot replace them.

The paper therefore avoids two opposite reductions. Individualistic reduction assigns a system-dependent capacity entirely to the person who prompts or signs the output. System personification assigns understanding, purpose, authority, authorship, or responsibility to a configured system simply because it realizes a capacity.

The alternative is not to insist that every property has a different bearer. It is to require separate justification for each attribution.

6. Discussion and Conclusion

Generative AI makes an old epistemological problem newly visible. Human inquiry has long depended on people, representations, instruments, institutions, and technologies. What generative interfaces increasingly do is present heterogeneous operations through a unified conversational surface.

The appropriate response is neither model reduction nor unlimited system expansion. Before drawing an attribution boundary, the analyst must specify the relevant task class, operating conditions, and acceptable performance criterion. A capacity should then be attributed to a minimal adequate organized configuration: one sufficient to realize the specified capacity and containing no dispensable proper subconfiguration. Several such configurations may exist.

Functional externalization helps identify which parts of epistemically relevant work have changed support. Comparative functional reconstruction then distinguishes integrated components from resources and background conditions and tests relevant functional roles rather than assuming that each individual component must be counterfactually indispensable.

This has practical implications for AI evaluation. Model evaluation is not necessarily evaluation of deployed epistemic capacity. When performance depends systematically on retrieval, tool use, orchestration, verification, or other organized operations, those relations belong to the explanatory object appropriate to that capacity.

Yet system-level functional attribution is not system-level personification. Once capacity has been located, understanding, epistemic authority, purpose, control, authorship, and responsibility remain separate attribution questions. Their loci may coincide, overlap, or diverge.

The framework deliberately stops short of deciding whether a capacity-bearing configuration constitutes a mind, an epistemic agent, or a complete sociotechnical practice. That is its limitation, but also its purpose: to discipline attribution before stronger ontological and normative conclusions are drawn.

Locating epistemic capacity is the beginning of attribution, not its conclusion.

Ethics Declaration: This study is conceptual and does not involve human participants, personal data, animal subjects, or experimental interventions. Ethical clearance was therefore not required.

AI Declaration: ChatGPT was used to support literature discovery and synthesis, conceptual development and drafting during the preparation and revision of this paper. As a non-native speaker of English, the author also used ChatGPT and Grammarly for translation, language refinement, and stylistic editing. The author critically reviewed all AI-assisted material and assumes full responsibility for the accuracy, interpretation, originality, and final content of the manuscript.

References

Alvarado, R. (2023) 'AI as an Epistemic Technology', Science and Engineering Ethics, 29(5), Article 32. DOI: 10.1007/s11948-023-00451-3.

Bareither, C. and Griessl, L. (2026) 'Hybrid epistemic practices and the transformation of academic assemblages: generative AI and epistemic messiness in the qualitative social sciences and humanities', AI & Society, 41, pp. 6459–6475. DOI: 10.1007/s00146-026-02988-w.

Cabitza, F. and Colombo, G. (2026) 'Beyond Epistemia: Epistemic Schizologia and Large Language Models as Techno-Semiotic Machines', arXiv preprint, arXiv:2607.25620. DOI: 10.48550/arXiv.2607.25620.

Clark, A. and Chalmers, D. (1998) 'The Extended Mind', Analysis, 58(1), pp. 7–19. DOI: 10.1093/analys/58.1.7.

Gouveia, S.S. and Wang, Y. (2026) 'Can generative artificial intelligence be considered a cognitive subject? An analytic analysis', AI & Society, 41, pp. 4859–4868. DOI: 10.1007/s00146-026-02924-y.

Hardwig, J. (1985) 'Epistemic Dependence', The Journal of Philosophy, 82(7), pp. 335–349. DOI: 10.2307/2026523.

Hauswald, R. (2025) 'Artificial Epistemic Authorities', Social Epistemology, 39(6), pp. 716–725. DOI: 10.1080/02691728.2025.2449602.

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

Kabir, M.N. (2025) 'Governing Generative AI at Scale: Institutionalizing Alignment for Organizational Purpose', in Proceedings of the 5th International Conference on AI Research (ICAIR 2025), pp. 172–180. DOI: 10.34190/icair.5.1.4374.

Kaplan, D.M. (2012) 'How to demarcate the boundaries of cognition', Biology & Philosophy, 27(4), pp. 545–570. DOI: 10.1007/s10539-012-9308-4.

Kaplunovich, A. (2025) 'LangGraph-Orchestrated LLM Agents for Scalable Movie Knowledge Graphs and Question Answering', in Proceedings of the 5th International Conference on AI Research (ICAIR 2025), pp. 181–191. DOI: 10.34190/icair.5.1.4142.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S. and Kiela, D. (2020) 'Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks', Advances in Neural Information Processing Systems, 33, pp. 9459–9474.

Norkin, S. and Kirchner, K. (2025) 'Interpersonal Trust Development in GenAI-augmented Organisations', in Proceedings of the 5th International Conference on AI Research (ICAIR 2025), pp. 358–365. DOI: 10.34190/icair.5.1.4310.

Palermos, S.O. (2014) 'Knowledge and Cognitive Integration', Synthese, 191(8), pp. 1931–1951. DOI: 10.1007/s11229-013-0383-0.

Palermos, S.O. (2022a) 'Epistemic Collaborations: Distributed Cognition and Virtue Reliabilism', Erkenntnis, 87, pp. 1481–1500. DOI: 10.1007/s10670-020-00258-9.

Palermos, S.O. (2022b) 'Responsibility in epistemic collaborations: Is it me, is it the group or are we all to blame?', Philosophical Issues, 32(1), pp. 335–350. DOI: 10.1111/phis.12230.

Rodrigues, T.V. (2026) 'The Apt Curation Model: An Epistemic Virtue Theory of AI-Assisted Authorship', Philosophy & Technology, 39, Article 39. DOI: 10.1007/s13347-026-01038-z.

Schick, T., Dwivedi-Yu, J., Dessì, R., Raileanu, R., Lomeli, M., Hambro, E., Zettlemoyer, L., Cancedda, N. and Scialom, T. (2023) 'Toolformer: Language Models Can Teach Themselves to Use Tools', Advances in Neural Information Processing Systems, 36, pp. 68539–68551. DOI: 10.52202/075280-2997.

Vega-Baez, J.-M. (2026) 'Fuera de Nosotros'. SERIE CIMA. ISBN: 979-8193428503.

Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K.R. and Cao, Y. (2023) 'ReAct: Synergizing Reasoning and Acting in Language Models', International Conference on Learning Representations.