AP-005 — Laboratorios Alexandria

Verified Negative Isomorphisms as First-Order Epistemic Results

Laboratorios Alexandria
contact@laboratoriosalexandria.com
June 2026

Abstract

The scientific community overwhelmingly treats the absence of structural correspondence between two domains as a null result—an outcome that does not merit publication, analysis, or theoretical integration. We argue that this treatment is epistemically wasteful. A verified negative isomorphism—a demonstration that two domains do not share a formal structural mapping, accompanied by explicit falsifiability conditions—is a first-order epistemic result that delimits the space of possible knowledge. We present three verified negative isomorphisms produced by an autonomous deliberation system operating across multiple scientific domains, each graded independently by a constitutionally governed epistemic judge and each accompanied by conditions under which the negative finding would be overturned. We analyze their relationship to the broader deliberation corpus: of 307 completed deliberation sessions, 62 (20.2%) produced Grade D verdicts that function as candidate negative results, while the three verified negatives presented here achieved Grade A—the highest quality designation—demonstrating that rigorous negatives can meet the same evidential standards as positive discoveries. We discuss the RLHF case as an illustrative structural category error and the implications for cross-domain research methodology.

1 Introduction

Cross-domain research seeks structural correspondences between fields: mathematical isomorphisms, shared generative mechanisms, analogous dynamical patterns. When such correspondences are found, they are celebrated as breakthroughs—biomimetics, network science, computational neuroscience, and econophysics all emerged from successful cross-domain mappings. But the search for correspondences inevitably produces negatives: domain pairs where no meaningful structural mapping exists. These negatives are overwhelmingly treated as uninteresting.

This treatment reflects an implicit assumption: that the absence of structure is the absence of information. We challenge this assumption directly. A verified negative isomorphism—one that has been subjected to formal deliberation, evaluated against constitutional standards, and accompanied by explicit conditions for its own falsification—contains substantial information about the structure of knowledge itself. It tells us where the boundaries are, which domains are genuinely independent despite surface similarities, and which lines of research are unlikely to yield cross-domain insight.

The analogy to experimental science is direct. A controlled experiment that fails to detect an effect, when conducted with adequate power and appropriate controls, is a publishable result—not because the absence is interesting per se, but because it constrains future theory. Null results prevent the research community from repeating expensive investigations that have already been rigorously conducted. Negative isomorphisms serve the same function in cross-domain research.

2 Three Verified Negative Isomorphisms

The following three results were produced by an autonomous deliberation system that evaluates cross-domain correlations through multi-actor constitutional deliberation. Each result was independently graded by an epistemic judge fine-tuned on quality evaluation. The grades, confidence levels, and falsifiability conditions are reported as produced by the system.

Domain Pair	Confidence	Grade	Key Finding
Computation/AI ↔ Epistemic Foundations	1.00	A	Syntactic-operational vs semantic-normative
Computation/AI ↔ Financial Markets	0.95	A	Instrumental, not structural
Life Sciences ↔ Materials Science (specific)	0.90	A	Negative within most fertile pair

Table 1. Three verified negative isomorphisms, all achieving Grade A (highest quality designation).

2.1 Computation and AI ↔ Epistemic Foundations

Result: No real isomorphism exists between Computation/AI and Epistemic Foundations (confidence: 1.0, Grade A). Despite the superficial similarity between computational knowledge representation (embeddings, ontologies, knowledge graphs) and philosophical epistemology (justified true belief, coherentism, foundationalism), the deliberation concluded that the two domains operate on fundamentally different objects. Computation represents knowledge as manipulable data structures; epistemology investigates the conditions under which belief is justified. The former is syntactic and operational; the latter is semantic and normative. No formal mapping preserves the essential properties of both domains.

Falsifiability condition: This result would be falsified if a shared system of representation were identified that simultaneously satisfies the operational requirements of computational knowledge systems and the normative requirements of epistemological frameworks—that is, a representation that is both computationally tractable and philosophically adequate as an account of justified belief.

2.2 Computation and AI ↔ Financial Markets

Result: No formal isomorphism exists between Computation/AI and Markets (confidence: 0.95, Grade A). This result is particularly noteworthy given the enormous industrial investment in applying AI to financial markets. The deliberation concluded that the relationship between computation and markets is instrumental, not structural. AI systems are applied to markets as tools; they do not share a formal structure with markets. Market dynamics are driven by reflexive human behavior, regulatory regimes, and geopolitical events—none of which have computational analogs that preserve their essential properties.

Falsifiability condition: This result would be falsified through rigorous mathematical-structural analysis (e.g., category theory) demonstrating that the formal structures of computation and those of market dynamics are isomorphic—not merely that computational tools can model market behavior, but that the domains share a common formal skeleton.

2.3 Life Sciences ↔ Materials Science (Specific Case)

Result: No real isomorphism exists between Life Sciences and Materials Science in a specific evaluated correlation (confidence: 0.90, Grade A). This result is particularly informative because the Life Sciences × Materials Science pair is the second most fertile cross-domain combination in the overall corpus, with 808 verified positive correlations and 5 Grade A theses (AP-006, Table 2). The negative finding in this specific case demonstrates that cross-domain fertility at the aggregate level does not imply universal structural correspondence. Even within highly fertile domain pairs, there exist specific sub-problems where the domains are genuinely independent.

Falsifiability condition: This result would be falsified if a unifying theory were identified that structurally connects the specific phenomena evaluated in this correlation—not merely a shared vocabulary or metaphorical similarity, but a formal theory from which both phenomena can be derived.

3 Negatives in the Broader Deliberation Corpus

The three verified negatives presented above are the most rigorous examples, having achieved Grade A. But the broader deliberation corpus contains a much larger population of candidate negative results. Of 307 completed deliberation sessions, the quality grade distribution is: 11 Grade A (3.6%), 54 Grade B (17.6%), 180 Grade C (58.6%), and 62 Grade D (20.2%).

The 62 Grade D sessions represent cases where the deliberation found insufficient evidence for a structural correspondence. These are not verified negatives in the strict sense—they lack the explicit falsifiability conditions and the high-confidence verdict that characterize the three Grade A negatives. But they occupy an informative intermediate position: they are correlations that were detected by the system (because some structural similarity exists at the surface level) but that failed under deliberative scrutiny (because the similarity is superficial, not structural).

The distinction between a Grade D verdict and a verified negative is the distinction between absence of evidence and evidence of absence. A Grade D session says: we deliberated and did not find sufficient support. A Grade A negative says: we deliberated and found positive evidence that no structural correspondence exists, and here is what would prove us wrong. The former is a weak negative; the latter is a strong negative. Both are informative, but only the latter qualifies as a first-order epistemic result.

The 20.2% Grade D rate across the corpus establishes an empirical base rate for cross-domain failure: approximately one in five deliberated correlations does not survive scrutiny. This rate is remarkably stable across surprise score ranges (AP-007), suggesting that the probability of a negative outcome is independent of how surprising the initial detection was—another argument for the independence of surprise and quality.

4 An Illustrative Case: RLHF as Structural Category Error

To illustrate the practical significance of negative isomorphisms, consider the case of Reinforcement Learning from Human Feedback (RLHF), the dominant alignment technique in current AI development. RLHF applies the formalism of reinforcement learning—reward signals, policy optimization, value functions—to the problem of aligning language model outputs with human preferences.

The implicit assumption is that human preference is structurally analogous to a reward signal: scalar, consistent, and optimizable. But human preference is none of these things. It is contextual, contradictory, evolving, and frequently incoherent. Treating preference as reward is not merely a simplification—it is a category error. It maps the structure of one domain (reinforcement learning, where rewards are well-defined scalar signals in a stationary environment) onto another domain (human evaluative judgment, where "rewards" are post-hoc rationalizations of complex, context-dependent, and often contradictory assessments).

This is a negative isomorphism: the formal structure of RL and the formal structure of human evaluative judgment are not isomorphic. The practical consequences of ignoring this negative—reward hacking, sycophantic behavior, Goodhart's Law applied to preference models—are well-documented in the literature (Casper et al., 2023). We suggest that treating this as a verified negative isomorphism from the outset would have oriented the field differently: toward methods that respect the non-scalar, non-stationary nature of human judgment rather than attempting to force it into a reward framework.

5 Why Negative Isomorphisms Are First-Order Results

We argue that verified negative isomorphisms serve four epistemic functions that qualify them as first-order results:

Boundary delimitation. Each negative isomorphism establishes where a structural boundary exists in knowledge space. These boundaries are not arbitrary—they reflect genuine differences in the generative mechanisms of different domains. Mapping these boundaries is as informative as mapping the connections. The companion paper AP-011 (Epistemic Desert Map) provides a framework for systematically identifying and classifying such boundaries across the full cross-domain landscape.

Resource allocation. Research programs that pursue cross-domain connections across verified boundaries are unlikely to succeed. The Computation/AI × Markets negative (Section 2.2) is directly relevant to the AI-in-finance industry: the absence of structural isomorphism suggests that improvements in computational sophistication will produce diminishing returns in market prediction, because the bottleneck is not computational but lies in the non-computational nature of market-generating mechanisms. The companion paper AP-006 corroborates this with deliberation data: the Life Sciences × Markets pair produced 75% Grade D verdicts, indicating systematic evidential failure.

Theory constraint. Any future theory that claims to unify two domains must account for existing negative isomorphisms. A unification theory that cannot explain why specific cross-domain mappings fail is incomplete. Negative results constrain the space of viable theories.

Methodological self-correction. The discovery that a widely assumed correspondence does not hold (as in the RLHF case) forces methodological revision. Fields that systematically ignore negative isomorphisms risk perpetuating structural errors across generations of research. The emergence of Direct Preference Optimization (DPO) as an alternative to RLHF can be read as a partial acknowledgment of this negative: DPO avoids the explicit reward model that embodies the failed isomorphism between RL rewards and human preferences.

6 Falsifiability as a Requirement for Negative Results

A critical distinction separates a verified negative isomorphism from a mere failure to find a connection. The former includes explicit conditions under which the negative would be overturned; the latter does not. This distinction is essential because "we looked and didn't find anything" is not a scientific result—it is a status report. "No structural correspondence exists between domains X and Y, and here is what would prove us wrong" is a result.

All three negative isomorphisms presented in Section 2 include falsifiability conditions. These conditions are specific and testable: they identify the type of evidence (shared representation system, category-theoretic isomorphism, unifying theory) that would overturn the negative finding. The inclusion of falsifiability conditions transforms a negative from an absence of evidence into a positive claim about the structure of knowledge—a claim that invites future testing rather than closing investigation.

The constitutional framework that governs the deliberation system requires falsifiability conditions for all grades, including negatives. This requirement ensures that negative results are not treated as final pronouncements but as provisional claims subject to future revision—precisely the epistemological standard that Karl Popper articulated for all scientific claims.

7 Sources of Uncertainty and Limitations

The three negative isomorphisms presented here were produced by a single autonomous system operating under a specific constitutional framework. Different deliberation systems with different constitutional criteria might reach different conclusions about the same domain pairs.

The system evaluates correlations based on a knowledge base that, as of June 2026, contains over 360,000 documents comprising abstracts and metadata but not full papers. Deeper engagement with primary methodological literature might reveal structural correspondences that are invisible at the abstract level.

The concept of "isomorphism" as used here is informal—it refers to structural correspondence evaluated by deliberative judgment, not to isomorphism in the strict mathematical sense. A formal category-theoretic analysis might reveal or refute correspondences that deliberative evaluation cannot access.

The RLHF analysis in Section 4 is presented as an illustrative case, not as a result produced by the deliberation system. It reflects the authors' interpretation and is subject to the limitations of any theoretical argument.

The 62 Grade D sessions discussed in Section 3 are candidate negatives, not verified negatives. The distinction between weak negatives (insufficient evidence for correspondence) and strong negatives (positive evidence of no correspondence) is important: we do not claim that all Grade D verdicts represent verified negative isomorphisms.

8 Conclusion

We have argued that verified negative isomorphisms—demonstrations that two domains do not share structural correspondence, accompanied by explicit falsifiability conditions—are first-order epistemic results that merit publication, analysis, and theoretical integration. The three verified negatives presented here, produced by an autonomous deliberation system and independently graded by a constitutionally governed epistemic judge, demonstrate that negative findings can meet the same evidential standards applied to positive discoveries: all three achieved Grade A, the highest quality designation.

The broader deliberation corpus provides context: of 307 sessions, 62 produced Grade D verdicts that function as weak negatives, establishing a 20.2% base rate for cross-domain failure under scrutiny. The existence of rigorous negatives within the most fertile domain pair (Materials Science × Life Sciences) demonstrates that aggregate fertility does not imply universal correspondence. And the RLHF case illustrates the practical cost of ignoring negative isomorphisms: structural category errors that persist across generations of research.

The scientific community's systematic neglect of negative cross-domain results has consequences: wasted research effort on connections that have already been rigorously evaluated and found wanting, perpetuation of structural category errors, and an incomplete map of knowledge space that shows only the roads but not the walls. We invite the research community to reconsider the epistemic status of well-verified negatives.

References

[1] Uzzi, B., Mukherjee, S., Stringer, M. & Jones, B. (2013). Atypical Combinations and Scientific Impact. Science, 342(6157), 468–472.

[2] Casper, S., Davies, X., Shi, C., et al. (2023). Open Problems and Fundamental Limitations of Reinforcement Learning from Human Feedback. arXiv:2307.15217.

[3] Bai, Y., Kadavath, S., et al. (2022). Constitutional AI. arXiv:2212.08073.

[4] Rafailov, R., Sharma, A., et al. (2023). Direct Preference Optimization. NeurIPS 2023.

[5] Ouyang, L., Wu, J., et al. (2022). Training Language Models to Follow Instructions with Human Feedback. NeurIPS 2022.

[6] Mac Lane, S. (1971). Categories for the Working Mathematician. Springer.

[7] Popper, K. (1959). The Logic of Scientific Discovery. Routledge.

[8] Goodhart, C. A. E. (1984). Problems of Monetary Management. In Monetary Theory and Practice. Macmillan.

[9] Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. PLOS Medicine, 2(8), e124.

[10] Rosenthal, R. (1979). The File Drawer Problem. Psychological Bulletin, 86(3), 638–641.