Formal constraints on automated scientific inquiry and the logical limits of epistemic agents

Elisson, I-2026
elisson@erzi.me

Abstract

The rapid progression of automated scientific discovery, exemplified by the deployment of deep reinforcement learning agents and large-scale language models in algorithmic optimization, has brought to the forefront fundamental questions regarding the boundaries of machine-led research. While systems such as AlphaDev and AlphaEvolve demonstrate a capacity for identifying novel, highly efficient solutions within discrete search spaces—such as sorting algorithms or complex codebases—their operational reach is governed by formal logical and information-theoretic constraints. These boundaries are not merely a function of current computational power but are rooted in the underlying structure of the formalisms that define these agents. By examining the interplay between interrogative models, dynamic epistemic logic, and statistical learning theory, it becomes possible to map the epistemic invariants that constrain the transition from optimization to genuine conceptual innovation.

Keywords: *logic of inquiry, fixed signature, epistemic invariants, interrogative models, dynamic epistemic logic, learning under misspecification, impossibility results*

Definitional and strategic rules in the interrogative model of inquiry

The theoretical foundation for understanding scientific inquiry as a formal process is rooted in Jaakko Hintikka's Interrogative Model of Inquiry (IMI). This framework conceptualizes reasoning not as a static deduction from premises, but as a strategic game-theoretic interaction between an inquirer and an oracle, which can represent nature, a database, or an experimental setup.[1] A central insight of Hintikka's work is the rigorous distinction between two types of rules that govern this interaction: definitional rules and strategic rules.[2] This distinction is vital for analyzing modern automated systems, which are often highly optimized in the latter while remaining static in the former.

Definitional rules establish the boundaries of the "game" of inquiry. They dictate which moves are legally permissible, what constitutes a well-formed question or a valid inference, and what the winning conditions are for the process.[2] In the context of an automated agent like AlphaDev, the definitional rules are provided by the instruction set architecture (ISA) and the syntax of the assembly language used to construct algorithms.[4] An agent cannot "invent" a new CPU instruction or violate the basic logic of memory addressing because such actions would fall outside the definitional scope of its operational universe. These rules act as ontological constraints, defining the "atoms" of the system and the ways they can be combined.

Strategic rules, by contrast, are the methodologies used to navigate the space of possibilities defined by the definitional rules to achieve an optimal outcome.[2] In contemporary artificial intelligence, the refinement of strategic rules has been the primary driver of progress. Methods such as Monte Carlo Tree Search (MCTS) used in AlphaZero and subsequently AlphaDev represent advanced strategic heuristics for exploring astronomically large search spaces.[4] However, strategic brilliance cannot compensate for a limited definitional basis. If the solution to a scientific problem requires a concept or a relationship that cannot be expressed within the agent's initial signature, no amount of strategic optimization will lead to its discovery.[1] This implies that a system may reach a "superhuman" level of performance in a specific game (such as sorting or matrix multiplication) while remaining fundamentally incapable of changing the game itself.

The Socratic Method serves as a historical and philosophical precursor to this formalization, where the strategy of questioning is restricted to prevent trivializing the inquiry through errors such as *petitio principii*.[1] In a formal IMI framework, the inquirer must possess a strategy that leads the process of knowledge acquisition, as there are no mechanical rules that dictate the specific sequence of moves to be made.[1] The success of an automated agent is therefore measured by its ability to select questions (or experimental designs) that maximize information gain relative to a goal, a process that is essentially strategic rather than definitional.[2]

Inferential erotetic logic and the structure of search scenarios

Scientific inquiry is fundamentally an erotetic process—a process driven by questions. The formal modeling of this process is extended by Andrzej Wiśniewski's Inferential Erotetic Logic (IEL), which provides the machinery to analyze how questions are generated from premises and how they lead to further inquiry.[8] Within IEL, the concept of an erotetic search scenario (e-scenario) is central for representing the rational strategies of an agent engaged in problem-solving.[9]

An e-scenario is a formal, syntactic structure, often represented as a labeled tree, that maps out the possible paths of inquiry starting from a principal question and a set of initial premises.[9] Each node in the tree represents either an auxiliary question or a declarative update. The scenario provides "conditional instructions" that dictate the next step in the inquiry based on the answers received from an external source.[9] This structure ensures that the inquiry is not a random collection of queries but a logically coherent pursuit of a solution.

The transitions between questions in an e-scenario are governed by the relation of erotetic implication.[9] A transition from a question Q to a sub-question Q_1 is valid if it meets the criteria of transmission of soundness and cognitive usefulness.[9] Transmission of soundness ensures that if the principal question Q has a true answer (given the truth of the premises), then the sub-question Q_1 must also have a true answer. Cognitive usefulness implies that any possible answer to Q_1 must provide information that narrows down the set of possible answers to Q.[9]

This formalization reveals a significant constraint on automated agents: the erotetic closure of the system. Every question asked by an agent must be a well-formed interrogative within the agent's language L_Σ.[10] This means that the agent's ability to "discover" is limited by the set of questions it can formulate. If a breakthrough requires questioning the validity of an underlying assumption or proposing an entity that is not in the current vocabulary, the agent will be unable to generate the necessary sub-questions.[11] The search scenarios are thus centripetal, refining knowledge within a predefined symbolic space but lacking the mechanism to expand that space autonomously.

Furthermore, IEL highlights that questions themselves can play the role of premises or conclusions in an inferential process.[12] Arriving at a question is often as rigorous a process as arriving at a declarative conclusion. Automated systems that treat inquiry as a series of independent queries often miss the deeper inferential connections between questions, which are essential for navigating complex scientific landscapes where the answer to one question fundamentally changes the askability of the next.[9]

Dynamic epistemic logic and the awareness approach

The evolution of an agent's knowledge state during the discovery process can be modeled using Dynamic Epistemic Logic (DEL). Unlike static epistemic logic, which describes what an agent knows in a fixed state, DEL focuses on the actions that modify information, such as observation, public announcements, and internal inference.[13] In a standard Kripke model $M = \langle W, R, V \rangle$, an informative event is typically modeled as a model-transforming operation that contracts the set of possible worlds W.[13]

When a truthful announcement ϕ is made, the new model $M|_\phi$ retains only those worlds where ϕ is true.[14] This process of model contraction is the primary mechanism of learning in DEL. However, a critical observation is that while the set of possible worlds is reduced, the valuation function V—which defines the truth of atomic propositions in Σ—remains constant for the remaining worlds.[14] This implies that standard epistemic updates do not add new dimensions or new atomic facts to the agent's universe; they merely filter out existing possibilities. This "factual invariance" suggests that discovery in these systems is restricted to the elimination of uncertainty about a pre-defined reality.[15]

To address the limitations of ideal agents who are assumed to know all logical consequences of their information—the problem of logical omniscience—the awareness approach is often utilized.[16] This framework distinguishes between implicit information (what is true in all worlds the agent considers possible) and explicit information (what the agent has actually acknowledged).[16] An agent is said to explicitly know ϕ only if it implicitly knows ϕ and is "aware" of ϕ, where awareness is represented by an A-set of formulas.[16]

The dynamics of awareness are crucial for scientific discovery, as the process of "discovering" often involves moving information from the implicit to the explicit category through the act of inference.[16] DEL systems that incorporate awareness allow for the modeling of agents that must perform specific reasoning actions to unlock the consequences of their data.[16] This acknowledges that discovery is not just about receiving external signals but about the internal reorganization of information. However, even with awareness, the agent is still operating within a language L that is typically fixed from the outset. The "language expansion" required for genuine ontological innovation—where the agent adds new symbols to its signature—remains a frontier for formal logic.[16]

Information theoretic boundaries and the pathology of misspecification

The limits of automated inquiry are also manifest in the statistical and information-theoretic foundations of learning. The Minimum Description Length (MDL) principle provides a framework for induction based on the idea that the best explanation for a set of data is the one that allows for the greatest compression.[18] Learning is equated with finding regularities, and regularities are used to shorten the description of the data.[19] While MDL and Bayesian inference are powerful tools for model selection and parameter estimation, they are susceptible to "pathologies" when the underlying models are misspecified.[21]

Statistical misspecification occurs when the true data-generating process P^* is not contained within the agent's set of candidate models $\mathcal{M}$.[22] In such cases, standard Bayesian and MDL inference can be inconsistent.[21] The Inconsistency Theorem, developed by Peter Grünwald and others, demonstrates that there exist learning problems where, even with infinitely many samples, the generalization error of the Bayes or MDL classifier remains bounded away from the smallest achievable error.[21] This means that the agent can "asymptotically overfit," concentrating its posterior mass on suboptimal models because they appear to provide a better fit under the specific constraints of the misspecified model class.[21]

This result has profound implications for automated discovery. It suggests that if an

AI researcher is searching for laws within an inadequate conceptual framework, it may not merely fail to find the truth, but it may stabilize at an incorrect explanation that it "believes" to be optimal based on its compression metrics.[21] This is particularly relevant in classification tasks where agents are forced to convert discrete functions into probability models, a process that often introduces subtle but critical misspecification.[21]

Furthermore, the "barrier of hypercompression" describes a state where an agent treats systemic anomalies not as signals for a new ontology, but as noise within its current framework.[15] To minimize description length, the agent may add increasing layers of parametric complexity (e.g., more parameters in a neural network) to "explain away" the anomalies, rather than performing the far more difficult task of revising its fundamental signature.[15] This parametric expansion serves as a temporary fix that masks the need for a paradigm shift, locking the agent into a local optimum of descriptive efficiency.[22]

Operational analysis of AlphaDev and AlphaEvolve

The practical application of automated discovery is best observed in recent advancements from DeepMind, specifically AlphaDev and AlphaEvolve. These systems represent the "state of the art" in algorithmic discovery but also exemplify the technical and structural boundaries previously discussed.

AlphaDev achieved significant attention for discovering sorting algorithms that are faster than those optimized by human engineers for decades.[4] It accomplishes this by framing the task of finding efficient routines as a single-player "AssemblyGame".[5] In this game, the state is the current algorithm and the CPU status (registers and memory), and the actions consist of appending low-level assembly instructions (e.g., mov, add, cmp).[4] AlphaDev uses an extension of the AlphaZero framework, employing a Transformer-based representation of the assembly code and Monte Carlo Tree Search to explore the space of possible programs.[4]

Despite its success, AlphaDev operates within a rigid "Fixed-Signature" environment. Its search space is defined by a fixed instruction set architecture (ISA), and its goal is to find the most efficient combination of these predefined symbols.[4] One noted limitation is that AlphaDev's search is discrete; when choosing an immediate value (a constant), it must select from a predefined list. It cannot "invent" a magic number or fine-tune an offset through gradient descent because the search tree is composed of discrete steps.[25] This has led to proposals for "differentiable CPUs" that internalize the deterministic logic of a processor into a continuous latent space, allowing for the optimization of constants through gradients—a move toward expanding the agent's operational signature.[25]

AlphaEvolve represents a further evolution of this approach, moving from the optimization of single functions to the evolution of entire codebases.[26] It utilizes an ensemble of large language models, such as Gemini Flash and Gemini Pro, to generate proposals for code modifications ("diffs") which are then tested against an automated evaluation framework.[28] AlphaEvolve's architecture includes an evolutionary database that maintains a diverse population of solutions, using algorithms like MAP-Elites to balance exploration and exploitation.[26]

A critical technical limit of AlphaEvolve is its dependence on automated verification. It can only tackle problems that come with a built-in, quantifiable success metric—a "fitness function".[28] This restricts its scope to domains like mathematics, computer science, and hardware design where correctness can be formally checked.[27] Furthermore, AlphaEvolve is constrained by its initial codebase and the libraries it has access to. While it can produce novel algorithmic configurations, it does so within the "grammar" of the programming language provided (e.g., Python or Verilog).[27] It cannot autonomously introduce a new programming paradigm or a fundamentally new way of representing data that isn't already expressible in its environment. This reinforces the idea that current agents are "Inquiry 1.0" systems—superb optimizers within fixed signatures.[15]

Higher-order logic and signature revision in ontology

evolution

To transcend the limits of fixed signatures, research has turned toward automated ontology evolution and signature revision. This is formalized in systems like GALILEO, which are designed to repair knowledge when conflicting ontologies yield inconsistent inferences.[17] In the domain of physics, this often occurs when a theoretical model predicts one value for a quantity while a sensory/experimental setup provides another.[17]

GALILEO utilizes Higher-Order Logic (HOL) because lower-order logics (such as Description Logic or First-Order Logic) lack the expressivity to reason about predicates and functions themselves.[17] The system employs Ontology Repair Plans (ORPs), which are sequences of diagnostic and transformation rules. A key mechanism is "Signature Revision," where the agent modifies the very structure of its language to resolve a conflict.[17]

An illustrative example is the "Where's My Stuff?" (WMS) plan.[17] When a discrepancy is detected between a predicted value and an observed value (e.g., the energy of a bouncing ball or the orbital velocity of stars), the WMS plan can trigger a revision that redefines the function in question.[17] This redefinition might involve "splitting" the function or adding a new component—effectively postulating a new entity, such as "elastic energy" or "dark matter," to account for the "missing stuff".[17] This represents a formal "exit" from a fixed signature; the agent is not just searching for a better parameter within its current model, but is adding a new symbol to its ontology to preserve consistency with the data.[17]

The use of polymorphic variables and higher-order theorem provers like Isabelle allows these repair plans to be generalized across disparate domains.[17] This approach moves closer to "Inquiry 2.0," where the agent has a signature revision operator Ω that allows it to transition between symbolic frameworks.[15] However, the autonomy of such systems is still limited; the repair plans themselves (the "meta-strategies") are often designed by human researchers, and the agent's ability to verify the "goodness" of an ontological expansion is a significant logical hurdle.[17]

Expected information gain and the teleological trap

The strategic selection of experiments in automated inquiry is typically governed by Bayesian Optimal Experimental Design (BOED), with the Expected Information Gain (EIG) as the primary utility function.[30] EIG is defined as the mutual information between the parameters of interest θ and the potential experimental outcomes y.[31] Maximizing EIG ensures that the chosen experiment d is the one expected to reduce the most entropy in the agent's posterior distribution $p(\theta|y, d)$.[30]

While EIG is a principled framework for efficient resource allocation, it possesses a structural "nearsightedness" regarding new categories. The calculation of EIG requires a prior distribution $p(\theta)$ and a predictive model $p(y|\theta, d)$.[30] These are defined over a specific parameter space. If a potential discovery lies outside this space—if it requires a "new category"—it is effectively invisible to the EIG calculation. Because the prior has zero support for concepts it does not yet possess, the expected utility of any action that might lead to such a concept is zero.[15]

This creates a teleological trap: the agent only values experiments that refine its current understanding. In sequential experiments, the agent replaces its prior with the posterior from the previous step, further concentrating its search within the established boundaries.[30] This "explosion in possible posterior beliefs" makes generalization hard and can lead to a "learning difficulty" where the agent requires a prohibitive number of samples to move beyond modest problem sizes.[34]

Furthermore, the computation of EIG is often the bottleneck in real-time experiments, especially for non-linear or high-dimensional models.[30] This has led to the development of amortized variational inference and transport-of-measure methods to estimate EIG more efficiently.[30] However, these technical improvements focus on making the *current* mechanism faster, not on expanding its ontological reach. The "epistemic uncertainty" that

EIG seeks to reduce is uncertainty about the *values* of known parameters, not uncertainty about the adequacy of the *parameter space itself*.[30]

The epistemic ceiling and the architecture of ignorance

The synthesis of these formal limits suggests the existence of an "epistemic ceiling"—a hard limit on the knowledge an agent can acquire, dictated by its structural and logical constraints.[37] This is not merely a matter of data availability; even as AI systems become more powerful, they may continue to hit a "ceiling of coherence" where the bottleneck is not memory or processing speed, but the ability to maintain a unified and expanding epistemic structure.[37]

Four distinct forms of epistemic limitation—entropy, Gödelian incompleteness, Heisenberg uncertainty, and logical paradox—comprise what has been termed the "architecture of ignorance".[40] While entropy describes uncertainty within a known system, Gödelian limits and paradoxes describe structural truths that are forever outside the reach of proof within a given system.[40] These veils do not "thin with effort" but are structural shadows cast by the act of formalization itself.[40]

Observation itself is not a transparent window into reality but a "protocol-dependent encoding".[38] A system can only detect what its instruments are physically capable of registering and its protocols are capable of encoding.[38] When a complex, high-dimensional reality hits a limited sensor, it "flattens," resulting in dimensional reduction.[38] An automated agent, therefore, does not observe reality directly but rather the response generated under its specific measurement rules. If a phenomenon does not align with the agent's protocol or signature, it is "logically unobservable"—it cannot enter the system as information.[38]

This suggests that scientific advancement in automated systems is often a matter of "swapping one filter for a slightly more complex one" rather than eliminating the filter entirely.[38] The "unseen" is not necessarily far away in space or time; it may be overlapping with the agent's environment, yet remain inaccessible because its "vibration" or dimension does not align with the agent's measurement protocols.[38] Protecting the "epistemic frontier" thus requires not just faster computers but the development of "epistemic guardians" and infrastructures that can authenticate provenance and verify lines of reasoning that exceed human legibility.[39]

Epistemic domination and conceptual erasure

The deployment of powerful automated inquiry systems also introduces risks of "epistemic domination"—the asymmetrical capacity of one party (or system) to control the evidence available to another.[42] In the context of large language models used for scientific inquiry, this can lead to "conceptual erasure".[42] When LLMs default to a "view from nowhere" or are trained on a specific, geographically or theoretically biased corpus, they can inferiorize alternative epistemologies or theoretical frameworks.[37]

This is particularly dangerous in "extended investigative contexts" where the eliminative paradigm of classical abduction—where hypotheses compete within a fixed space until one is pruned—dominates.[36] If the automated agent acts as the primary filter for scientific validity, it may prematurely exclude inconsistent evidence or "hallucinate" coherence where ambiguity is the more accurate representation of reality.[37] The transition from competition to "co-opetition"—where multiple explanatory lines are maintained in suspension—is a proposed alternative that leverages "quantum abduction" to avoid premature closure.[36]

In this framework, semantic superposition tracks epistemic uncertainty about a determinate reality, allowing for the integration of contradictions as interference effects rather than as reasons for elimination.[43] This approach recognizes that in complex domains like medicine or forensics, human experts naturally maintain multiple hypotheses until evidence forces a collapse.[36] Replicating this capacity in automated systems requires moving beyond simple "eliminative search" toward a "dynamic synthesis" that can support composite explanations.[36]

Synthesis of technical requirements for ontological discovery

The cumulative evidence from formal logic, information theory, and the analysis of existing AI systems points toward a set of technical requirements that must be met for an agent to transcend "Inquiry 1.0" and achieve genuine scientific innovation.

The first requirement is the transition from model contraction to signature expansion. Standard learning updates that merely eliminate worlds in a Kripke model are insufficient.[13] The agent must be capable of "signaturistic change"—the introduction of new constants, predicates, and function symbols into its language.[15] This process is not a random addition of symbols but a targeted revision driven by the detection of "global inconsistency" across multiple, locally consistent ontologies.[17]

The second requirement is the implementation of a signature revision operator Ω. This operator must be governed by meta-strategies that can evaluate the "representational adequacy" of the current language.[15] This involves higher-order reasoning (HOL) to identify patterns of "fault diagnosis" in existing theories.[17] For example, when a conservation law is violated, the agent must be able to "hypothesize" a new term in an equation that restores the balance, and then assign that term a placeholder in its ontology.[17]

The third requirement is the overcoming of teleological myopia. Current utility functions like EIG are designed to reduce uncertainty within a fixed prior support.[30] Discovery requires "centrifugal" strategies that assign non-zero utility to actions that explore the "unknown unknowns"—regions of conceptual space where the agent currently has no prior support.[15] This may involve "quantum" approaches to abduction that maintain a superposition of incompatible hypotheses, allowing for the emergence of "hybrid resolutions" that a classical, eliminative search would miss.[36]

Finally, the agent must be grounded in an evaluation framework that can handle ambiguity and "noise" without collapsing into hypercompression.[15] This requires distinguishing between measurement error (benign misspecification) and paradigmatic mismatch (bad misspecification).[21] An agent that interprets every anomaly as noise to be explained away by more parameters will remain trapped in a state of "epistemic deadlock".[15]

Conclusions and the future of automated discovery

The investigation into the formal limits of scientific inquiry reveals that the "superhuman" performance of modern AI systems like AlphaDev and AlphaEvolve is currently confined to the domain of strategic optimization within fixed signatures. These agents are highly efficient navigators of pre-defined logical spaces, yet they lack the structural mechanisms required for autonomous paradigm shifts—the "Inquiry 2.0" capability of ontological innovation.

The primary barriers to such innovation are rooted in the erotetic closure of questioning, the factual invariance of epistemic updates, the pathology of statistical misspecification, and the teleological myopia of current experimental design metrics. To push the epistemic frontier, the focus of research must shift from increasing computational scale and parametric complexity to the development of formal frameworks for language expansion and signature revision.

This requires the integration of higher-order logic, the adoption of awareness-based epistemic models, and the replacement of eliminative search strategies with synthetic, abductive processes. Only by endowing automated agents with the capacity to reinvent their own vocabularies and question their own foundational protocols can the "epistemic ceiling" be raised. Until then, automated science will continue to excel at "polishing the face" of established paradigms, providing immense value through optimization while remaining structurally blind to the next scientific revolution. The architecture of the map-making process itself must be the object of discovery, moving the machine from a state of being an informed player to becoming a creator of new "games" of science.

Sources

- THE INTERROGATIVE MODEL OF INQUIRY AS A LOGIC OF ..., accessed on January 22, 2026, https://journals.rudn.ru/philosophy/article/download/11389/10819
- Interrogative Logic as Underlying Logic in Scientific Practices – Oxford Academic, accessed on January 22, 2026, https://academic.oup.com/jigpal/advance-article-pdf/doi/10.1093/jigpal/jzae094/58633550/jzae094.pdf
- Hintikka's Generalizations of Logic and Their Relation to Science – Suppes Corpus, accessed on January 22, 2026, https://suppescorpus.stanford.edu/sites/g/files/sbiybj32751/files/media/file/hintikkas_generalizations_of_logic_and_their_relation_to_science_409.pdf
- Daily AI Research Digest: Google's Reinforcement Learning Frontiers for Complex Tasks, accessed on January 22, 2026, https://skywork.ai/skypage/en/Daily-AI-Research-Digest-Reinforcement-Learning-Frontiers-for-Complex-Tasks/1948206470819074048
- Faster Sorting Algorithms Discovered Using Deep Reinforcement Learning | PDF – Scribd, accessed on January 22, 2026, https://www.scribd.com/document/820382227/Faster-sorting-algorithms-discovered-using-deep-re
- Learning as a Strategic Process: Development of Hintikka's Model – Vilnius University Press, accessed on January 22, 2026, https://www.journals.vu.lt/problemos/lt/article/download/1902/9124/12040
- Hintikka's Interrogative Model and a Logic of Discovery and Justification – ResearchGate, accessed on January 22, 2026, https://www.researchgate.net/publication/281161291_Hintikka's_Interrogative_Model_and_a_Logic_of_Discovery_and_Justification
- Andrzej Wiśniewski – Google Scholar, accessed on January 22, 2026, https://scholar.google.se/citations?user=hLwVYYYAAAAJ
- Answering by Means of Questions in View of Inferential Erotetic Logic, accessed on January 22, 2026, https://awisniew.web.amu.edu.pl/publikacje/IELanswering.pdf
- Epistemic Erotetic Search Scenarios – ResearchGate, accessed on January 22, 2026, https://www.researchgate.net/publication/326686184_Epistemic_Erotetic_Search_Scenarios
- Calculizing Classical Inferential Erotetic Logic – Moritz Cordes, accessed on January 22, 2026, https://moritzcordes.com/wp-content/uploads/2020/03/cordes-2020-calculizing-iel pdf
- An Introduction to Inferential Erotetic Logic, accessed on January 22, 2026, https://awisniew.web.amu.edu.pl/dydaktyka/Unilog/2.%20Inferential%20Erotetic%20Logic.pdf
- Dynamic Epistemic Logic – Wikipedia, accessed on January 22, 2026, https://en.wikipedia.org/wiki/Dynamic_epistemic_logic
- Dynamic Epistemic Logic – Stanford Encyclopedia of Philosophy, accessed on January 22, 2026, https://plato.stanford.edu/entries/dynamic-epistemic/
- Formalization of the Limits of Scientific Inquiry: The Fixed-Signature Inquiry Theorem and Epistemic Invariants in Automated Systems (PDF)

- Dynamic Epistemic Logic for Implicit and Explicit Knowledge – CEUR-WS, accessed on January 22, 2026, https://ceur-ws.org/Vol-627/lrba_5.pdf
- Higher-Order Representation and Reasoning for Automated Ontology Evolution – Edinburgh Research Explorer, accessed on January 22, 2026, https://www.pure.ed.ac.uk/ws/portalfiles/portal/25316206/HIGHER_ORDER_REPRESENTATION_AND_REASONING_FOR_AUTOMATED_ONTOLOGY_EVOLUTION.pdf
- The Minimum Description Length Principle – CWI, accessed on January 22, 2026, https://homepages.cwi.nl/~pdg/book/book.html
- The Minimum Description Length Principle – ResearchGate, accessed on January 22, 2026, https://www.researchgate.net/publication/227458453_The_Minimum_Description_Length_Principle
- Model Selection Based on Minimum Description Length – IRI, Columbia University, accessed on January 22, 2026, https://iri.columbia.edu/~tippett/cv_papers/Grunwald.pdf
- Suboptimal Behavior of Bayes and MDL in Classification under Misspecification – ResearchGate, accessed on January 22, 2026, https://www.researchgate.net/publication/225834984_Suboptimal_behavior_of_Bayes_and_MDL_in_classification_under_misspecificat
- Suboptimal Behavior of Bayes and MDL in Classification under Misspecification – CWI, accessed on January 22, 2026, https://homepages.cwi.nl/~pdg/ftp/inconsistency.pdf
- Suboptimal Behavior of Bayes and MDL in Classification under Misspecification – Machine Learning (Theory), accessed on January 22, 2026, https://hunch.net/~jl/projects/prediction_bounds/MDL/current.pdf
- The Minimum Description Length Principle (MDL), accessed on January 22, 2026, http://bactra.org/notebooks/mdl.html
- Project Silicon: What If We Could Do Gradient Descent on Assembly Code?, accessed on January 22, 2026, https://rewire.it/blog/project-silicon-gradient-descent-on-assembl
- AlphaEvolve: Evolutionary Agent from DeepMind – Composio, accessed on January 22, 2026, https://composio.dev/blog/alphaevolve-evolutionary-agent-from-deepmind
- AlphaEvolve: A Gemini-Powered Coding Agent for Designing Advanced Algorithms – DeepMind Blog, accessed on January 22, 2026, https://deepmind.google/blog/alphaevolve-a-gemini-powered-coding-agent-for-designing-advanced-algorithms/
- AlphaEvolve (PDF) – DeepMind, accessed on January 22, 2026, https://storage.googleapis.com/deepmind-media/DeepMind.com/Blog/alphaevolve-a-gemini-powered-coding AlphaEvolve.pdf
- Can AlphaEvolve Change How We Solve Problems? – HPCwire, accessed on January 22, 2026, https://www.hpcwire.com/2025/06/02/can-alphaevolve-change-how-we-solve-probl
- Variational Bayesian Optimal Experimental Design – NeurIPS, accessed on January 22, 2026, http://papers.neurips.cc/paper/9553-variational-bayesian-optimal-experimental- pdf
- Differentiable Multi-Target Causal Bayesian Experimental Design – OpenReview, accessed on January 22, 2026, https://openreview.net/pdf?id=as-dCPnWKH
- Estimating Expected Information Gains for Experimental Designs with Application to the Random Fatigue-Limit Model, accessed on January 22, 2026, https://www.researchgate.net/publication/243103179_Estimating_Expected_Information_Gains_for_Experimental_Designs_With_Application_to_the_Random_Fatigue-Limit_Model

- New Tools for Bayesian Optimal Experimental Design and Kernel-Based Generative Modeling, accessed on January 22, 2026, https://dspace.mit.edu/handle/1721.1/158795
- Efficient Bayesian Experimental Design with Deep Learning – Carnegie Mellon University, accessed on January 22, 2026, https://ml.cmu.edu/research/phd-dissertation-pdfs/phd_thesis_final_cigoe.pdf
- Bayesian Experimental Design via Contrastive Diffusions – OpenReview, accessed on January 22, 2026, https://openreview.net/forum?id=h8yg0hT96f
- Quantum Abduction: A New Paradigm for Reasoning Under Uncertainty – arXiv, accessed on January 22, 2026, https://arxiv.org/html/2509.16958v2
- ARCH Science – Biopharma Business Intelligence Unit (BBIU), accessed on January 22, 2026, https://www.biopharmabusinessintelligenceunit.com/arch-science
- The Instrument Wall: Why Reality Is a Protocol-Dependent Projection – Medium, accessed on January 22, 2026, https://medium.com/@m238xu/the-instrument-wall-why-reality-i
- An Endgame Sketch for Resilient Epistemic Ecosystems, accessed on January 22, 2026, https://airesilience.net/vision-epistemic-ecosystems
- The Architecture of Ignorance: Entropy, Gödel, Heisenberg, and Beyond – ResearchGate, accessed on January 22, 2026, https://www.researchgate.net/publication/393526840_The_Architecture_of_Ignorance_Entropy_Godel_Heisenberg_and_Beyond
- Accelerating Civilisational Resilience in the Era of AI, accessed on January 22, 2026, https://airesilience.net/assets/AI%20Resilience_%20Accelerating%20Civilisational%20Resilience%20in%20the%20Era%20of%20AI_1761507055253-DZI_ATY7.pdf
- Epistemic Domination – ResearchGate, accessed on January 22, 2026, https://www.researchgate.net/publication/370029227_Epistemic_Domination
- Quantum Abduction: A New Paradigm for Reasoning Under Uncertainty – MDPI, accessed on January 22, 2026, https://www.mdpi.com/2413-4155/7/4/182
- Quantum Abduction: A New Paradigm for Reasoning Under Uncertainty – ResearchGate, accessed on January 22, 2026, https://www.researchgate.net/publication/398589338_Quantum_Abduction_A_New_Paradigm_for_Reasoning_Under_Uncertainty