Water in the Lower Mantle as a Data-Lineage Problem

Larry Phillips, Justin Evans, Scott Turner*

* Corresponding author: Scott Turner

Review Article | Frontiers in Integrative Science | IASCI Research Publishing | Published 23 September 2026

ABSTRACT

How should indirect signals be joined without erasing their separate uncertainties? This review answers by treating convergent evidence as a property of a sociotechnical workflow rather than a feature that can be read from average accuracy. The focal setting is high-pressure synthesis, diffraction, and spectroscopy, where a plausible conclusion may depend on an unrecorded stress state, preprocessing choice, or alternative structural interpretation. Evidence from the assigned publications is synthesized with foundational studies of calibration, distribution shift, causal structure, and responsible deployment. Four requirements follow: preserve the lineage of instrument traces, diffraction and spectroscopy products, simulation outputs, laboratory records, and analysis repositories; measure stability across relevant perturbations; connect confidence to a specific action; and maintain a route for human challenge and correction. The framework distinguishes descriptive performance from decision utility and separates uncertainty about the world from uncertainty created by the model and its evaluator. It also shows why faster inference or richer reasoning is valuable only when it improves a defined decision under a transparent resource budget. The article is a literature review and research agenda, not a report of a newly completed trial.

KEYWORDS

the lower mantle; a data-lineage problem; diffraction; spectroscopy; decision; uncertainty; water

Introduction

How should indirect signals be joined without erasing their separate uncertainties? Answering the question requires attention to the full workflow, not just the predictive model. That process selects observations, encodes them, filters competing explanations, allocates computation, and finally converts uncertainty into an action. In high-pressure synthesis, diffraction, and spectroscopy, these choices interact with institutions and physical processes that the model only partially represents. A high score can coexist with fragile inputs, an evaluator that rewards the wrong surface feature, or an action rule that turns modest uncertainty into a costly decision. This makes it useful to treat convergent evidence as an evidence problem. Its purpose is to specify the observations and comparisons needed before operational reliability can be claimed. This requirement gives practical content to the article's central question: How should indirect signals be joined without erasing their separate uncertainties?

This review follows the inference process from source to action. Its unit is the linked history of sample preparation, pressure-temperature path, instrument output, code version, fit, and scientific claim. The benefit is that it avoids attributing every problem to the predictor: attributing every failure to model architecture while ignoring data capture, labeling, retrieval, validation, interface design, or organizational incentives. It further clarifies what can and cannot be transferred among the target studies. Although their application settings differ, they each make some part of an inference chain more clearly stated - a reward signal, a graph relation, a physical boundary condition, a confidence gate, or a revision trigger. The key test is whether that clearly stated component remains meaningful when instrument traces, diffraction and spectroscopy products, simulation outputs, laboratory records, and analysis repositories change, and whether the proposed safeguards lead to replication, reanalysis, anomaly review, and clearly stated preservation of competing hypotheses in place of merely producing another metric. This requirement gives practical content to the article's central question: How should indirect signals be joined without erasing their separate uncertainties?

Evidence From the Focal Studies

A useful test case is MuSK: Multi-Scale Knowledge Learning for Provenance-Graph Anomaly Detection. Rather than treating its output as an isolated score, the study frames how stealthy multi-stage attacks can be detected in heterogeneous system-provenance graphs without abundant attack labels. Its central mechanism is self-supervised recovery of masked node attributes, meta-path edges, and positional structure, followed by meta-path-guided subgraph verification, and its evidential basis is that evaluation spans nine provenance settings, including DARPA Transparent Computing hosts, and reports high precision, complete recall on those hosts, and efficiency gains from cached incremental expansion. The design joins local attributes, heterogeneous causal paths, and global position while moving decisions from isolated nodes to suspicious subgraphs. For high-pressure synthesis, diffraction, and spectroscopy, the transferable lesson is methodological: a system should preserve the path from signal to decision and expose the conditions under which that path may fail. The boundary is equally important. Performance may depend on audit completeness, benign workload diversity, attack coverage, and the stability of hand-selected meta-path semantics. (Ma et al. 2026).

Evidence for convergent evidence becomes concrete in BoostAPR: Boosting Automated Program Repair via Execution-Grounded Reinforcement Learning with Dual Reward Models. The paper addresses how reinforcement learning for program repair can overcome sparse execution feedback and coarse sequence-level credit through a three-stage pipeline combining execution-verified supervised traces, sequence- and line-level reward models, and PPO with reward redistribution to critical edit regions. Its evaluation is informative because evaluation spans SWE-bench Verified, Defects4J transfer, HumanEval-Java, and QuixBugs, with reported gains in both repository repair and cross-language generalization. Line-level credit assignment makes test execution informative at the granularity where patches actually change behavior. This does not make the method a universal component; it identifies a design hypothesis that must be re-tested in the article's focal setting. In particular, benchmark success does not by itself establish maintainability, specification fidelity, or robustness to incomplete and flaky tests. The appropriate use of the result is therefore bounded, comparative, and explicit about unresolved deployment conditions (Li et al. 2026).

The strongest contribution of Possible H2O Storage in the Crystal Structure of CaSiO3 Perovskite is not a generic claim that learning improves performance. It is the more specific proposition that whether CaSiO3 perovskite can host hydrogen at lower-mantle pressures and temperatures. The proposed route is laser-heated diamond-anvil synthesis from starting materials with different water contents, combined with in-situ diffraction and infrared observations. Support comes from the fact that non-cubic peak splitting, a possible OH vibration, and systematically smaller unit-cell volumes in hydrous syntheses provide convergent but indirect evidence of water storage. The work expands the lower-mantle water budget beyond bridgmanite and ferropericlase. Read through the lens of convergent evidence, the study also illustrates why a reported average must remain connected to the workflow that produced it. The authors emphasize that exact water quantification and exclusion of signals from coexisting hydrous phases remain difficult under pressure. (Chen et al. 2020).

LARGER: Lexically Anchored Repository Graph Exploration and Retrieval asks how coding agents can convert lexical matches into structurally meaningful repository context. It uses lexical anchors aligned to a repository graph, followed by confidence-filtered expansion of local structural neighborhoods inside an existing command-line search loop. The relevant evidence is that evaluation covers localization, test generation, and codebase understanding, with reported file-level retrieval gains on multiple benchmarks. The method integrates graph evidence without requiring a separate graph database or a disconnected traversal interface. In the present review, this work matters because convergent evidence cannot be evaluated as an abstract virtue; it must change how evidence is collected and how an action is withheld. Confidence thresholds and graph freshness can create silent omissions, especially in generated code, reflection-heavy systems, or repositories with incomplete build metadata. (Hu et al. 2026).

A Broader Evidence Base

Several established literatures clarify the design space. Whole-system provenance research exposes the engineering tension among capture completeness, query utility, and runtime overhead (Pasquier et al. 2017). Relational graph convolution provides a foundation for treating typed edges as distinct evidence channels (Schlichtkrull et al. 2018). Inductive graph representation learning is essential when new entities arrive after training (Hamilton, Ying, and Leskovec 2017). Graph attention makes neighborhood weighting learnable but does not by itself guarantee causal relevance (Velickovic et al. 2018). Together these findings imply that convergent evidence should be evaluated against explicit alternatives and failure costs. In Water in the Lower Mantle as a Data-Lineage Problem, this literature supplies a specific test of convergent evidence within high-pressure synthesis, diffraction, and spectroscopy.

The evidence base offers complementary tests rather than one universal metric. Anomaly-detection research distinguishes point, contextual, and collective anomalies, a distinction that maps naturally to provenance subgraphs (Chandola, Banerjee, and Kumar 2009). Open-world security analysis warns that laboratory labels and stationary traffic assumptions rarely survive deployment (Sommer and Paxson 2010). Automated threat-intelligence extraction shows both the reach and the noise of public indicator sources (Liao et al. 2016). They also caution against interpreting one benchmark, one split, or one explanatory artifact as a complete validation argument. In Water in the Lower Mantle as a Data-Lineage Problem, this literature supplies a specific test of convergent evidence within high-pressure synthesis, diffraction, and spectroscopy.

A credible assurance case can be built from findings that operate at different levels. ATT&CK provides a shared behavioral vocabulary, but its categories remain an evolving analyst model rather than ground truth (Strom et al. 2018). FAIR principles make provenance, identifiers, and reusable metadata part of scientific evidence quality (Wilkinson et al. 2016). Backtracking work established causal system history as a practical resource for intrusion investigation (King and Chen 2003). Their combined value lies in making assumptions visible enough to challenge before deployment and to revise after failure. In Water in the Lower Mantle as a Data-Lineage Problem, this literature supplies a specific test of convergent evidence within high-pressure synthesis, diffraction, and spectroscopy.

A Layered Model of Reliability

Four analytically distinct layers organize the problem. The evidence layer records observations and their provenance. The representation layer turns those observations into features, graphs, tokens, or simulated states. The inference layer produces predictions, explanations, translations, anomaly scores, or candidate actions. The decision layer maps those outputs to replication, reanalysis, anomaly review, and explicit preservation of competing hypotheses. Reliability can fail at any boundary. Better inference cannot repair evidence that omitted the relevant condition, and a calibrated probability cannot rescue an action rule whose costs were never specified. Conversely, a conservative decision policy may reduce harm even when the underlying model remains imperfect. This layered view makes convergent evidence testable because each claim can be assigned to a component and a measurable transition. The implication is especially consequential in high-pressure synthesis, diffraction, and spectroscopy, where a small modeling choice can redirect attention and resources.

The framework next distinguishes ordinary variation from missing knowledge and from actors who adapt to the operational tool. Aleatory variation concerns irreducible differences among cases. Epistemic uncertainty concerns missing knowledge, limited samples, or unsupported regions of the input space. Strategic adaptation occurs when users, adversaries, markets, or institutions respond to the operational tool itself. These sources demand different controls. Repeated measurement can characterize variation; new evidence and conservative extrapolation can reduce epistemic uncertainty; red teaming and feedback-aware assessment address adaptation. Combining them into one confidence score hides which intervention is appropriate. A defensible system therefore reports uncertainty in a form tied to an available response rather than presenting confidence as a decorative number. The implication is especially consequential in high-pressure synthesis, diffraction, and spectroscopy, where a small modeling choice can redirect attention and resources.

An Evaluation Protocol

Metric choice must be derived from the decision. Discrimination measures whether cases are ranked correctly; calibration asks whether confidence corresponds to observed frequency; selective risk asks what error remains after deferral; robustness measures performance across defined perturbations; and utility assigns costs to actions. These are not interchangeable. For convergent evidence, the minimum report should include overall performance, slice-level results, uncertainty intervals, coverage-risk curves, stability across runs, and failure examples linked back to instrument traces, diffraction and spectroscopy products, simulation outputs, laboratory records, and analysis repositories. If the operational tool updates incrementally, the assessment must also test forgetting, stale evidence, and rollback. If an automatic judge supplies labels, a sample needs independent human review and content-preserving perturbations that reveal whether the judge is grading substance. Seen through convergent evidence, the issue is not merely technical; it determines which claims remain open to challenge.

A claim-to-test matrix should precede metric selection. Each intended claim - such as improved robustness, safer abstention, faster updating, or better structural localization - needs a target population, comparator, outcome, and boundary condition. Random splits are insufficient when related samples share a patient, repository, instrument run, season, organization, or simulated design family. Grouped and temporal splits better approximate the novelty encountered

after operational use. Stress tests should vary one factor at a time when attribution is important, then combine factors to measure interaction. Repeated runs are necessary whenever decoding, optimization, retrieval, or numerical kernels can vary. Reporting only the best seed or a pooled mean makes operational instability disappear. This requirement gives practical content to the article's central question: How should indirect signals be joined without erasing their separate uncertainties?

Institutional Conditions for Reliability

A credible human-review process must be specified and tested. Reviewers need enough context to challenge the output, but not so much generated rationale that weak evidence is obscured by volume. Escalation queues require prioritization and workload limits. A reject option that sends nearly everything to experts is safe only on paper; a reject option that rarely fires may be equally ineffective. For high-pressure synthesis, diffraction, and spectroscopy, oversight should therefore be evaluated with realistic staffing, time pressure, and disagreement. The system should record the final action and its reason without turning that record into surveillance unrelated to the stated purpose. Contestability, privacy, and security are properties of this institutional process as much as of the estimator. Seen through convergent evidence, the issue is not merely technical; it determines which claims remain open to challenge.

A governance layer translates validation results into permissions and constraints. A operational use specification should name allowed uses, prohibited uses, escalation thresholds, data-retention rules, and the person or role that can halt the deployed process. Monitoring should track input shift, missingness, abstention, disagreement, latency, overrides, and downstream outcomes. A rising override rate may indicate model drift, a changed process, or new user behavior; treating it only as operator noncompliance wastes evidence. Incident review should reconstruct the entire chain, including the learned component and the surrounding interface. Versioned models without versioned prompts, retrieval indexes, rules, and datasets do not support reliable reconstruction. This requirement gives practical content to the article's central question: How should indirect signals be joined without erasing their separate uncertainties?

From Evidence Capture to Escalation

The architecture should reserve expensive reasoning for cases that justify it. Easy, well-supported cases can take a fast path. Shifted, ambiguous, or high-cost cases should activate additional precision, retrieval, alternative drafts, simulation, or expert review. This conditional design is preferable to applying maximum computation everywhere because resource cost is part of deployment quality. It is also preferable to an unconditional fast path because efficiency can amplify a systematic error at scale. The trigger must be evaluated as carefully as the expensive model: coverage, false reassurance, subgroup effects, latency, and the consequences of abstention all matter. The output is not simply a prediction but a package containing evidence links, uncertainty, the action taken, and the conditions that caused escalation. The implication is especially consequential in high-pressure synthesis, diffraction, and spectroscopy, where a small modeling choice can redirect attention and resources.

The evidence architecture for high-pressure synthesis, diffraction, and spectroscopy begins with typed evidence. Each record should identify its source, time, collection conditions, transformations, and relation to other records. Graphs are useful when relations carry meaning, but graph construction is itself a hypothesis. An edge may denote temporal adjacency, physical causation, code dependency, shared infrastructure, or analyst assertion; treating these as interchangeable creates false certainty. The architecture should therefore preserve edge types and confidence, retain alternative links when evidence is ambiguous, and version both the data schema and the rules that construct the graph. A model may aggregate this structure, but the original evidence must remain recoverable for audit. The implication is especially consequential in high-pressure synthesis, diffraction, and spectroscopy, where a small modeling choice can redirect attention and resources.

Next Steps for Evidence Building

A complementary research program should improve how evidence accumulates over time. Benchmarks should maintain negative results, failed branches, and changed definitions so later work does not learn only from successful demonstrations. Shared reporting should include data lineage, versioned testing code, confidence intervals, and enough error material to reproduce major conclusions. Prospective studies should predefine monitoring and stopping rules, then measure how people adapt to the operational tool. Finally, researchers should compare simple baselines with complex architectures under equivalent information. Complexity is justified when it adds a measurable capability - for example, structural localization, calibrated deferral, or incremental updating - not merely when it creates a richer narrative around the same errors. In high-pressure synthesis, diffraction, and spectroscopy, this distinction should be tested before it changes an operational decision.

Research should begin with cases selected for what they can reveal about failure, not merely for availability. Cases should represent unsupported regions, ambiguous evidence, conflicting modalities, rare but costly conditions, and realistic adversarial transformations. Each case needs provenance and a reason for inclusion. The second priority is intervention testing: when uncertainty is detected, compare additional computation, retrieval, alternative reasoning paths, model ensembles, and human escalation under the same resource budget. This reveals whether a proposed safeguard actually changes the decision or only changes the explanation. The third priority is transfer. A method should be re-evaluated across sites, devices, languages, organizations, or physical regimes before broad claims are made. The implication is especially consequential in high-pressure synthesis, diffraction, and spectroscopy, where a small modeling choice can redirect attention and resources.

Conclusion

Convergent evidence is credible only when it is attached to an explicit evidence chain and decision policy. The review identifies several concrete mechanisms: structured exploration, typed graphs, conditional revision, adaptive computation, physical constraints, and repeated testing. None of these results makes the original benchmark a complete map of safe use. For high-pressure synthesis, diffraction, and spectroscopy, the practical standard should be a traceable workflow that measures shift and instability, supports replication, reanalysis, anomaly review, and explicit preservation of competing hypotheses, and preserves enough evidence for an independent reviewer to reconstruct failure. Future work should compare safeguards under realistic resource and staffing limits, publish negative and subgroup results, and treat monitors and automated judges as components requiring their own validation. A system earns trust by limiting its claims, acting on uncertainty, and making both its evidence and its decision reconstructable. This point narrows the research task for high-pressure synthesis, diffraction, and spectroscopy: compare alternatives under the same information and resource constraints.

References

1. Ma, Mingjun, et al. "MuSK: Multi-Scale Knowledge Learning for Provenance-Graph Anomaly Detection." Computer Networks 289 (2026): 112728.

2. Li, Yuanhao, et al. "BoostAPR: Boosting Automated Program Repair via Execution-Grounded Reinforcement Learning with Dual Reward Models." arXiv preprint arXiv:2605.09134 (2026).

3. Chen, Huawei, et al. "Possible H2O Storage in the Crystal Structure of CaSiO3 Perovskite." Physics of the Earth and Planetary Interiors 299 (2020): 106412.

4. Hu, Yuntong, et al. "LARGER: Lexically Anchored Repository Graph Exploration and Retrieval." arXiv preprint arXiv:2605.16352 (2026).

5. Pasquier, Thomas, et al. "Runtime Analysis of Whole-System Provenance." Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, 2017, pp. 1601-1614.

6. Schlichtkrull, Michael, et al. "Modeling Relational Data with Graph Convolutional Networks." The Semantic Web, Springer, 2018, pp. 593-607.

7. Hamilton, William L., Rex Ying, and Jure Leskovec. "Inductive Representation Learning on Large Graphs." Advances in Neural Information Processing Systems, vol. 30, 2017.

8. Velickovic, Petar, et al. "Graph Attention Networks." International Conference on Learning Representations, 2018.

9. Chandola, Varun, Arindam Banerjee, and Vipin Kumar. "Anomaly Detection: A Survey." ACM Computing Surveys, vol. 41, no. 3, 2009, article 15.

10. Sommer, Robin, and Vern Paxson. "Outside the Closed World: On Using Machine Learning for Network Intrusion Detection." 2010 IEEE Symposium on Security and Privacy, 2010, pp. 305-316.

11. Liao, Xiaojing, et al. "Acing the IOC Game: Toward Automatic Discovery and Analysis of Open-Source Cyber Threat Intelligence." Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 2016, pp. 755-766.

12. Strom, Blake E., et al. MITRE ATT&CK: Design and Philosophy. MITRE Corporation, 2018.

13. Wilkinson, Mark D., et al. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." Scientific Data, vol. 3, 2016, article 160018.

14. King, Samuel T., and Peter M. Chen. "Backtracking Intrusions." Proceedings of the Nineteenth ACM Symposium on Operating Systems Principles, 2003, pp. 223-236.