ORIGINAL RESEARCH **Open access**

The Problem of Negative Results in Materials AI: A Conceptual Proposal for Failure-Aware Learning Systems

Fatima Zahra Amrani[1*], Youssef Benali[1]

Abstract

The integration of artificial intelligence (AI) into materials science has significantly accelerated discovery and optimization processes. Yet, it simultaneously amplifies long-standing epistemic vulnerabilities rooted in the systematic underrepresentation of negative results. Failed experiments, unstable material phases, and inaccurate predictions are often excluded from the published record, resulting in datasets that are skewed and shape AI model training and inference. This conceptual paper examines how epistemic gaps distort the dynamics of data generation, model development, and experimental validation in materials AI. By synthesizing literature on publication bias, model robustness, and uncertainty-aware learning, the study demonstrates how positive-only knowledge bases foster overconfident predictions, limit generalization, and obscure material boundary conditions. To address these challenges, the paper proposes a failure-aware epistemic learning framework that structurally integrates negative results into AI-driven materials discovery through recursive feedback structures, uncertainty modulation, and inclusive steering logics. Ethical reasoning situates this framework within principles of epistemic accountability, sustainability, and responsible innovation. By reinterpreting negative results as indispensable sources of information rather than peripheral artifacts, the paper advances a conceptual foundation for more resilient, transparent, and reliable AI applications in materials science.

Keywords Materials AI, Interaction dynamics, Epistemic reasoning, Negative results, Publication bias, Failure-aware systems

*Correspondence:
Fatima Zahra Amrani
fatima.amrani@gmail.com
[1] Department of Computational Materials Science, Faculty of Engineering, Mohammed V University, Rabat, Morocco

Introduction

The integration of artificial intelligence (AI) into materials science marks a paradigmatic transformation in how materials are designed, evaluated, and optimized. By enabling the rapid processing of high-dimensional datasets and the extraction of non-linear relationships beyond human cognitive capacity, AI has fundamentally altered the tempo and scope of materials research [1, 2]. Conventional materials discovery has historically relied on iterative empirical experimentation supported by theoretical modeling, a process often characterized by long development cycles, high financial cost, and limited exploration of vast compositional spaces. In contrast, machine learning (ML) techniques—ranging from supervised property prediction to generative models and reinforcement learning—offer scalable alternatives that can accelerate discovery across diverse domains, including energy storage materials, structural alloys, catalysts, and biomedical devices [3, 4].

Despite these advances, the epistemic foundations upon which AI systems operate remain deeply entangled with the structure and quality of the scientific knowledge they ingest. AI does not generate knowledge ex nihilo; rather, it

extrapolates from existing datasets, inheriting both their strengths and their limitations. In materials science, this dependence exposes AI systems to long-standing structural biases embedded within research practices, particularly the systematic underreporting of negative results. These omissions pose critical challenges for the reliability, interpretability, and generalizability of AI-driven materials discovery.

Negative results—defined here as experimental or computational outcomes that fail to meet anticipated performance metrics, exhibit instability, or contradict prevailing hypotheses—are ubiquitous in materials research. Failed syntheses, non-reproducible phases, degradation under operational conditions, or predictive inaccuracies are integral to the field's exploratory nature. However, such outcomes are frequently excluded from formal dissemination, as publication cultures in both academic and industrial contexts privilege novelty, success, and apparent impact [5, 6]. This selective reporting produces datasets skewed toward positive outcomes, creating an incomplete and idealized representation of material behavior.

The analytical consequences of this imbalance are particularly pronounced in AI applications. Machine learning models trained predominantly on successful cases are prone to optimistic bias, often extrapolating beyond validated domains without adequate awareness of failure boundaries [7]. As a result, predictions may appear statistically robust while masking underlying fragilities, such as sensitivity to minor compositional variations or unmodeled thermodynamic constraints. In materials informatics, where predictive accuracy directly informs experimental investment, such distortions can propagate inefficiencies, misallocate resources, and delay genuine innovation.

These challenges are further compounded by the dynamics of interaction between human researchers and AI systems. As materials scientists increasingly delegate tasks such as phase stability prediction, property optimization, and high-throughput screening to AI tools, the opacity of many algorithmic architectures limits critical interrogation of their outputs [8]. Black-box models, particularly deep neural networks, often obscure the influence of biased or incomplete training data, fostering undue confidence in predictions that align with established positive trends. At a systems level, this can generate epistemic feedback loops in which AI models reinforce prevailing assumptions while systematically marginalizing anomalous or negative evidence.

Concrete implications emerge in practical settings such as alloy design, polymer discovery, and catalyst optimization. High-throughput computational screening may rapidly identify promising candidates but may fail to account for synthesis failures or instability observed in undocumented experimental trials [9]. Subsequent experimental validation, guided by these incomplete predictions, risks repeating known failures, undermining the very efficiency gains AI is intended to deliver. Thus, the promise of AI-accelerated discovery is inseparable from the integrity and inclusivity of the data ecosystems that sustain it.

From an ethical and epistemic perspective, the marginalization of negative results raises concerns that extend beyond methodological rigor. The pursuit of sustainable and environmentally responsible materials—central to addressing global challenges such as climate change, the energy transition, and resource scarcity—demands comprehensive knowledge systems capable of learning from both success and failure [10]. Ignoring negative outcomes inflates perceived innovation rates, obscures material lifecycles, and contributes to avoidable waste, including the environmental costs associated with unsuccessful synthesis and testing pathways [11]. Ethical stewardship of AI in materials science, therefore, requires a recalibration of research values, recognizing negative results as essential contributors to collective understanding.

Designing AI systems that meaningfully engage with negative data introduces inherent trade-offs. Emphasizing speed and computational efficiency may incentivize selective data inclusion, whereas comprehensive datasets increase computational complexity and curation demands [12]. Navigating these tensions necessitates adaptive feedback structures, such as iterative model updating informed by real-time experimental data and failure annotations [13]. Within such frameworks, negative results are reframed not as obstacles but as signals that delineate boundary conditions, refine search spaces, and enhance model resilience.

The underrepresentation of negative results in materials AI reflects broader patterns of publication bias across scientific disciplines. Meta-analytical studies consistently demonstrate that selective reporting leads to inflated effect sizes and distorted knowledge landscapes [14, 15]. In the context of materials informatics, this bias manifests as overfitted models optimized for narrow compositional regimes, leaving vast, potentially critical regions of compositional space underexplored [16]. Addressing this challenge requires systems-level interventions, including interdisciplinary collaboration among materials scientists, data scientists, and AI ethicists to develop inclusive data curation standards and reporting protocols [17].

Equity considerations further complicate this landscape. Materials challenges are global in scope, yet AI training datasets are disproportionately derived from well-resourced institutions and industrial laboratories. This imbalance risks reinforcing epistemic hierarchies, marginalizing insights from underrepresented regions, and constraining innovation diversity [18]. Ethical reasoning grounded in epistemic justice calls for deliberate inclusion of heterogeneous data sources and perspectives, ensuring that AI-driven materials discovery does not replicate existing structural inequities [19].

Across the full lifecycle of materials innovation—from initial design and synthesis to deployment and degradation—negative outcomes provide critical information about durability, safety, and performance under real-world conditions [20]. Effective interaction between AI predictions and experimental verification thus requires hybrid epistemologies, wherein human judgment, domain expertise, and algorithmic efficiency operate in complementary alignment rather than hierarchical substitution [21]. Negative results in materials science occur across multiple stages of AI-enabled discovery pipelines and vary in epistemic function **(Table 1)**.

Table 1. Typology of negative results in materials research and AI pipelines

Type of negative result	Stage of pipeline	Typical cause	Epistemic information provided	Risk if excluded from AI training
Failed synthesis	Experimental	Thermodynamic instability, processing mismatch	Reveals infeasible compositional or processing regimes	Repeated experimental failure, wasted resources
Non-reproducible phase	Experimental/Validation	Metastability, scale effects	Identifies fragile or context-dependent structures	Overestimation of material robustness
Poor property performance	Experimental/Testing	Trade-offs between structure and function	Defines performance ceilings and trade-off boundaries	Inflated performance expectations
Model prediction failure	Computational/AI	Dataset bias, extrapolation beyond the training domain	Signals model uncertainty and blind spots	Overconfident deployment decisions
Simulation–experiment mismatch	Hybrid	Incomplete physical assumptions	Reveals missing physics or scale-dependent effects	Persistent model–reality divergence

In summary, the problem of negative results in AI-driven materials science represents a pivotal challenge with methodological, ethical, and epistemic dimensions. By reconceptualizing failure as a source of insight rather than deficiency, the field can move toward more resilient and transparent discovery paradigms. This paper advances a failure-aware framework that integrates negative data into AI workflows, positioning comprehensive reporting as a cornerstone of sustainable, reliable, and ethically grounded materials innovation.

Theoretical Background and Literature Synthesis

Publication biases and epistemic gaps in scientific knowledge

The production of scientific knowledge is systematically shaped by publication norms that privilege positive, novel, and confirmatory findings, often marginalizing inconclusive or negative outcomes [22]. In materials science, this asymmetry produces a persistent epistemic imbalance in which successful syntheses, optimized properties, and performance peaks dominate the literature. At the same time, failed experiments, unstable phases, or irreproducible pathways remain largely undocumented [23]. The analytical consequence of this selective visibility is a distorted collective knowledge base, in which reported success rates are inflated, and the feasibility boundaries are poorly articulated. Such distortions generate self-reinforcing feedback loops, as subsequent research builds upon incomplete representations of prior work, thereby amplifying optimistic narratives while undermining epistemic robustness [24].

These gaps are further intensified by interaction dynamics within academic and industrial research ecosystems. Publication incentives, career pressures, and funding mechanisms frequently frame negative outcomes as individual shortcomings rather than as structurally informative signals, discouraging their dissemination [25]. At a systems level, this selective reporting propagates downstream effects: meta-analyses of materials properties often exhibit inflated effect sizes, and predictive reviews—particularly in energy and functional materials—tend to converge on idealized performance envelopes that are weakly supported by the full experimental record [26]. From an ethical standpoint, such epistemic incompleteness has material consequences. Misallocated resources, prolonged development cycles, and delayed deployment of sustainable technologies can be traced to knowledge infrastructures that systematically obscure failure, uncertainty, and constraint [27]. The exclusion of negative outcomes produces cascading epistemic distortions across materials AI systems **(Table 2)**.

Table 2. Epistemic and systems-level consequences of positive-only training data in materials AI

AI system level	Excluded information	Epistemic distortion	Systems-level consequence
Training data	Failed experiments, unstable compounds	Optimistic bias in learned representations	Reduced generalization
Model inference	Infeasible regions of chemical space	Artificial confidence in predictions	Misguided candidate selection
Validation	Prior known failures	Redundant experimentation	Inefficient resource allocation
Deployment	Degradation and failure modes	Underestimated operational risk	Environmental and safety concerns
Knowledge accumulation	Boundary conditions	Inflated innovation narratives	Slower long-term progress

Negative results in materials research and AI integration

Negative results in materials research constitute critical epistemic signals, delineating feasibility boundaries, exposing latent instabilities, and informing iterative refinement of compositional and processing parameters [28]. Their exclusion from dominant literature syntheses, however, produces hermeneutical voids in which the absence of failure data is misinterpreted as implicit success [29]. Conceptually, negative outcomes should be understood not as null results but as active contributors to scientific understanding, revealing trade-offs among performance, stability, scalability, and environmental tolerance that remain invisible in success-only narratives [30].

The integration of artificial intelligence into materials research amplifies the consequences of these omissions. Machine learning systems inherit the epistemic structure of their training data; when negative results are underrepresented, models internalize a systematically biased view of material viability [31]. Prevailing steering logics in model development—often optimized around benchmark accuracy and positive prediction rates—further marginalize failure modes and anomalous behaviors [32]. Systems-level syntheses indicate that AI-driven materials discovery pipelines are particularly vulnerable to such gaps, frequently overestimating stability and performance in sparsely explored or extrapolative chemical spaces [23]. Ethical concerns arise when these biases propagate into deployment contexts, where misguided confidence in material feasibility can lead to environmental risk, resource waste, or unsafe applications [14].

Robustness and uncertainty in AI models for materials

Robustness in AI-enabled materials science extends beyond predictive accuracy to encompass resilience against epistemic incompleteness, data sparsity, and structural uncertainty [15]. The literature increasingly emphasizes uncertainty quantification strategies—such as ensemble learning, Bayesian inference, and probabilistic surrogates—as mechanisms for exposing model limitations and rendering predictions more interpretable [16]. Analytically, the explicit representation of epistemic uncertainty enhances interaction dynamics between computational outputs and experimental decision-making, enabling more informed validation, rejection, or refinement of predicted hypotheses [17].

Nonetheless, trade-offs persist. Increasing model complexity to capture uncertainty can compromise computational efficiency and exacerbate overfitting to already biased datasets, thereby reducing generalizability [18]. Feedback structures, particularly active learning and iterative human-in-the-loop protocols, offer a partial resolution by reintegrating negative signals and anomalous observations into model retraining cycles [19]. Systems-level analyses converge on the view that robust AI architectures—those that explicitly acknowledge uncertainty and failure—are essential for mitigating epistemic risk and sustaining trust in AI-mediated materials prediction, especially in high-stakes domains such as energy, infrastructure, and sustainability-critical technologies.

Proposed conceptual framework: Failure-aware epistemic learning in materials AI

The conceptual framework proposed here advances a structural reinterpretation of negative results as indispensable epistemic components of AI-enabled materials discovery systems. Rather than treating failures as peripheral artifacts or data noise, the framework positions unsuccessful experiments, instabilities, and violated assumptions as active signals that shape representational confidence, hypothesis navigation, and decision reliability across the materials AI pipeline [1, 2]. At its core, the framework introduces a failure-aware feedback architecture in which negative outcomes are recursively reintegrated into model development, enabling adaptive learning under conditions of epistemic incompleteness [3].

Interaction dynamics within this architecture emphasize symmetry between positive and negative evidentiary streams. Positive results support performance validation and local optimization, while negative results delineate feasibility boundaries and constrain overconfident extrapolation into sparsely sampled chemical spaces [4]. By preserving this dual evidentiary structure, AI systems can discriminate between robust material regularities and artifacts arising from selective reporting or data sparsity [5]. Conceptually, failures are reframed not as endpoints but as interpretive inflection points that expose latent material behaviors—such as phase fragility, processing sensitivity, or non-linear trade-offs—thereby strengthening systems-level resilience [6].

Steering logics embedded in the framework prioritize epistemic inclusivity through modular, uncertainty-aware architectures. Negative outcomes influence internal representations via dedicated uncertainty channels, failure registries, and adaptive weighting mechanisms that modulate model confidence in response to observed breakdowns [7]. Iterative refinement is operationalized through feedback loops coupling computational inference with experimental interrogation, ensuring that discrepancies between predicted feasibility and empirical failure trigger systematic recalibration rather than silent exclusion [8]. Ethical reasoning is embedded structurally rather than appended post hoc, guiding how uncertainty,

limitations, and risk are communicated and acted upon in downstream decision-making contexts [9]. Key distinctions between conventional materials, AI approaches, and the proposed failure-aware framework are summarized in **Table 3**.

Table 3. Comparison between conventional and failure-aware materials AI systems

Dimension	Conventional materials AI	Failure-aware framework (this paper)	Epistemic implication
Data inclusion	Predominantly positive results	Balanced positive and negative results	Reduced epistemic bias
Treatment of failure	Excluded or ignored	Explicitly encoded and learned from	Boundary-aware inference
Uncertainty handling	Implicit or secondary	Structurally integrated	Improved interpretability
Model confidence	Performance-driven	Constraint- and failure-modulated	Reduced overconfidence
Learning dynamics	Static optimization	Recursive feedback adaptation	Enhanced resilience
Ethical orientation	Efficiency-focused	Accountability- and sustainability-driven	Responsible innovation

Figure 1 illustrates the proposed failure-aware epistemic learning framework as a cyclical system. The AI core is positioned centrally, surrounded by interacting loops representing dual input streams (positive and negative data), uncertainty-aware processing modules, and bifurcated output pathways for validation and feedback. Dynamic arrows depict recursive refinement driven by failure signals, while dashed epistemic boundaries mark zones of human oversight and interpretive intervention.

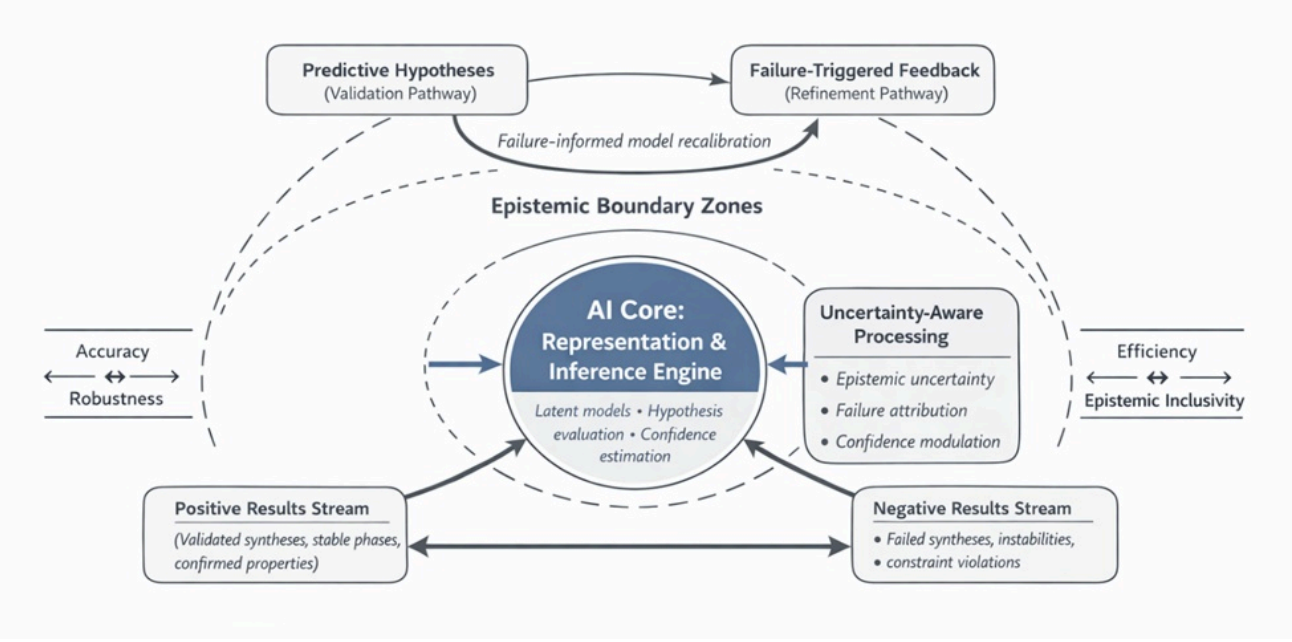

Figure 1. Failure-aware epistemic learning framework: Cyclical AI system with dual data streams and recursive refinement.

Systems-level insights suggest that this approach transforms epistemic vulnerabilities into strengths, promoting adaptive learning in complex environments. Interaction dynamics reveal how incorporating failures enhances model generalization, addressing gaps in traditional AI paradigms.

Epistemic reorientation: From success-driven learning to boundary-aware inference

A central implication of the proposed framework is an epistemic shift from success-driven learning toward boundary-aware inference. Traditional materials AI systems, trained predominantly on positive results, internalize a distorted epistemic landscape in which unexplored or failure-prone regions appear artificially viable [10]. By contrast, failure-aware learning explicitly encodes infeasible regimes, enabling models to represent not only where materials succeed, but also where and why they break down [11]. This reorientation transforms epistemic uncertainty—from a residual limitation into an active steering signal that guides exploration, validation, and hypothesis pruning [12].

At the level of interaction dynamics, this shift alters how data curation, model inference, and experimental design co-evolve. Negative outcomes reshape training distributions, constrain surrogate model confidence, and inform acquisition strategies in active learning loops, reducing redundant exploration of implausible candidates [13]. Systems-level analyses suggest that such boundary-aware inference mitigates epistemic overconfidence, a recurrent artifact in positive-biased learning regimes, particularly in high-dimensional materials spaces [14].

Trade-Off structures: Inclusivity, complexity, and generalization

The explicit integration of negative results introduces identifiable trade-offs in model design and computational governance. Expanding training scopes to include failure data increases representational complexity and computational demands, particularly when negative outcomes are heterogeneous or weakly structured [15]. However, this added complexity yields epistemic dividends by enhancing model robustness against extrapolation errors and premature convergence in unexplored chemical spaces [16]. The principal trade-offs introduced by failure-aware learning and corresponding mitigation strategies are outlined in **Table 4**.

Table 4. Trade-off structures and mitigation strategies in failure-aware materials AI

Trade-off	Introduced risk	Failure-aware design response	Net epistemic benefit
Data inclusivity vs efficiency	Increased dataset size	Selective failure weighting	Improved generalization
Model complexity	Computational burden	Modular uncertainty channels	Targeted robustness
Noise vs signal	Heterogeneous failures	Failure classification schemes	Meaningful boundary detection
Speed vs reliability	Slower convergence	Feedback-triggered recalibration	Reduced downstream waste
Exploration vs exploitation	Conservative search	Boundary-informed acquisition	Sustainable discovery

The framework resolves this tension by combining expressive models with uncertainty-aware mechanisms. Ensemble methods, probabilistic surrogates, and modular architectures enable selective emphasis on informative failures while avoiding indiscriminate complexity inflation [17]. Feedback structures further stabilize these trade-offs by allowing failure-induced discrepancies to trigger localized recalibration rather than global retraining, preserving efficiency while maintaining epistemic sensitivity [18].

Systems-Level adaptation and generalization dynamics

At the systems level, failure-aware architectures promote adaptive generalization by redistributing interpretive weight across success and failure regimes. Negative results illuminate latent material variabilities—such as synthesis bottlenecks, metastable transitions, or scale-dependent instabilities—that are systematically underrepresented in success-only datasets [19]. By incorporating these signals, AI systems develop representations that generalize across broader operational envelopes rather than optimizing narrowly around reported performance peaks [20].

Interaction dynamics evolve from unidirectional prediction pipelines to bidirectional epistemic dialogue. AI outputs inform boundary-setting in experiments, while experimental failures actively reshape model assumptions, reducing cycles of redundant positive pursuit and accelerating convergence toward robust design frontiers [21]. In large-scale discovery settings, this dynamic prevents premature collapse onto suboptimal solutions and sustains exploratory diversity over extended optimization horizons [22].

Ethical and governance implications of failure-aware design

Ethically, the framework addresses a structural blind spot in contemporary materials AI: the silent externalization of failure costs. When negative data are excluded, AI systems guide experimentation toward over-optimistic candidates, amplifying resource waste, environmental burden, and downstream deployment risk [23]. Failure-aware learning counters this trajectory by promoting transparent acknowledgment of limitations and uncertainty, aligning AI-mediated decision-making with principles of responsible innovation and sustainability governance [24].

From a governance perspective, incorporating negative outcomes necessitates standardized reporting and interoperable data protocols to ensure traceability and reuse of failure knowledge [25]. While this investment introduces short-term coordination costs, systems-level analyses indicate that it yields long-term epistemic gains by stabilizing knowledge accumulation and reducing systemic inefficiencies across materials innovation pipelines [26].

Results and Discussion

The conceptual proposal interprets negative results as transformative elements within materials AI ecosystems. Interaction dynamics between positive and negative inputs reveal emergent properties: models trained inclusively exhibit tempered confidence, interpreting discrepancies as opportunities for refinement rather than anomalies [13]. Systems-level insights demonstrate how feedback structures integrate epistemic gaps, creating self-correcting loops that evolve as evidence accumulates [14].

Steering logics prioritize resilience, balancing trade-offs between predictive precision on known successes and robustness to real-world failures [15]. Ethical reasoning frames this as an epistemic imperative, countering publication biases that skew collective knowledge and impede the advancement of sustainable materials [16]. Conceptual interpretations position failures as hermeneutic tools, unveiling constraints that positive narratives obscure [17].

Analytical implications for interdisciplinary collaboration emerge: materials experts and AI developers co-design protocols that value negative data, fostering hybrid intelligence where human insight complements algorithmic patterns [18]. Trade-offs in deployment—such as increased validation needs versus reduced experimental waste—highlight pragmatic pathways forward [19].

The framework's integrative nature suggests broader epistemic shifts in materials science, where AI evolves from an accelerator of known successes to a navigator of uncertainty landscapes [20]. Interaction dynamics thus promote humility in predictions, aligning technological promise with scientific rigor [21].

Conclusion

This conceptual exploration of negative results in materials AI culminates in a failure-aware paradigm that reinterprets epistemic shortcomings as pathways to robustness. By emphasizing interaction dynamics, feedback structures, and steering logics, the framework integrates negative outcomes into core learning processes, addressing trade-offs inherent in biased knowledge bases. Ethical and epistemic reasoning affirms the value of inclusivity, ensuring AI contributes to equitable, sustainable materials innovation.

Ultimately, transforming negative results from overlooked artifacts to foundational interpretive elements fosters resilient systems capable of navigating complex material realities. This interpretive shift invites a reorientation toward comprehensive knowledge systems, where apparent failures underpin enduring progress in applied artificial intelligence for materials science.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 27 Jan 2025 Revised: 23 Feb 2025 Accepted: 21 Mar 2025
Published online: 18 July 2025

References

Kozlowski MC. Negative data in data sets for machine learning training. J Org Chem. 2023;88(9):5678-80. https://doi.org/10.1021/acs.joc.3c00423.

Tran K, Ulissi ZW. Uncertainty prediction for machine learning models of material properties. ACS Omega. 2021;6(47):31457-65. https://doi.org/10.1021/acsomega.1c03752.

Northcutt C, Jiang L, Chuang I. Confident learning: Estimating uncertainty in dataset labels. J Artif Intell Res. 2021;70:1373-411.

Wang AYT, Murdock RJ, Kauwe SK, Oliynyk AO, Gurlo A, Brgoch J, et al. Machine learning for materials scientists: An introductory guide toward best practices. Chem Mater. 2020;32(12):4954-65. https://doi.org/10.1021/acs.chemmater.0c01907.

Dogan A, Birant D. Machine learning and data mining in manufacturing. Expert Syst Appl. 2021;166:114060.

Tao Q, Xu P, Li M, Lu W. Machine learning for perovskite materials design and discovery. npj Comput Mater. 2021;7:23. https://doi.org/10.1038/s41524-021-00493-1.

Xu P, Ji X, Li M, Lu W. Small data machine learning in materials science. npj Comput Mater. 2023;9:42. https://doi.org/10.1038/s41524-023-01000-z.

Reiser P, Neubert M, Eberhard A, Torresi Z, Zhou C, Shao C, et al. Graph neural networks for materials science and chemistry. Commun Mater. 2022;3:93. https://doi.org/10.1038/s43246-022-00315-6.

Oviedo F, Ferres JL, Buonassisi T, Carbune V, Cai E, Chen X, et al. Interpretable and explainable machine learning for materials science and chemistry. Acc Mater Res. 2022;3(6):597-607. https://doi.org/10.1021/accountsmr.1c00221.

Morgan D, Jacobs R. Opportunities and challenges for machine learning in materials science. Annu Rev Mater Res. 2020;50:71-103. https://doi.org/10.1146/annurev-matsci-070218-010015.

Himanen L, Jäger MOJ, Morooka EV, Canova FF, Ranawat YS, Gao DZ, et al. Dscribe: Library of descriptors for machine learning in materials science. Comput Phys Commun. 2020;247:106949. https://doi.org/10.1016/j.cpc.2019.106949.

Lu Z, Li X, Zhang H, Liu Y, Li H, Liu L, et al. Interpretable machine-learning strategy for soft-magnetic property and thermal stability in fe-based metallic glasses. npj Comput Mater. 2020;6:187. https://doi.org/10.1038/s41524-020-00460-4.

Amsler M, Sutherland DR, Chang MJ, Guevarra D, Connolly AB, Gregoire JM, et al. Autonomous materials synthesis via hierarchical active learning of nonequilibrium phase diagrams. Sci Adv. 2021;7(51):eabg4930. https://doi.org/10.1126/sciadv.abg4930.

Gomes CP, Fink D, van Dover RB, Gregoire JM. Computational sustainability meets materials science. Nat Rev Mater. 2021;6(7):524-36. https://doi.org/10.1038/s41578-021-00348-2.

Moosavi SM, Jablonka KM, Smit B. The role of machine learning in the understanding and design of materials. J Am Chem Soc. 2020;142(47):20273-87. https://doi.org/10.1021/jacs.0c09105.

Wang K, Shi H, Li T, Zhao L, Zhai H, Korani D, et al. Computational and data-driven modelling of solid polymer electrolytes. Digit Discov. 2023;2(6):1660-82.

Strieth-Kalthoff F, Sandfort F, Küpper C, Paßlick D, Glorius F. Machine learning for chemical synthesis. Chem. 2022;8(8):2116-52. https://doi.org/10.1016/j.chempr.2022.07.010.

Fanelli D. Pressures to publish: What effects do we see. Gaming the metrics. 2020:111.

Curry S. Ending publication bias: A values-based approach to surface null and negative results. PLoS Biol. 2024;22(9):e3003368. https://doi.org/10.1371/journal.pbio.3003368.

Schweinfurth MK, Frommen JG. Beyond the null: Recognizing and reporting true negative findings. iScience. 2024;27(1):108676. https://doi.org/10.1016/j.isci.2023.108676.

Abdar M, Pourpanah F, Hussain S, Rezazadegan D, Liu L, Ghavamzadeh M, et al. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. Inf Fusion. 2021;76:243-97. https://doi.org/10.1016/j.inffus.2021.05.008.

Merchant A, Batzner S, Schoenholz SS, Aykol M, Cheon G, Cubuk ED. Scaling deep learning for materials discovery. Nature. 2023;624(7990):80-5. https://doi.org/10.1038/s41586-023-06735-9.

Chen C, Ong SP. A universal graph deep learning interatomic potential for the periodic table. Nat Comput Sci. 2022;2:718-28. https://doi.org/10.1038/s43588-022-00364-5.

Strieth-Kalthoff F, Glorius F. Illuminating ‘the ugly side of science’: Fresh incentives for reporting negative results. Nature. 2024. https://doi.org/10.1038/d41586-024-01389-7.

Suarez A, Jimenez J, Loperena A, Rodriguez C, Molina J, Hicke I. Negative chemical data boosts language models in reaction outcome prediction. Sci Adv. 2024;10(45):eadt5578. https://doi.org/10.1126/sciadv.adt5578.

Hirschfeld L, Swanson K, Yang K, Barzilay R, Coley CW. Uncertainty quantification using neural networks for molecular property prediction. J Chem Inf Model. 2020;60(8):3770-80.

Peterson AA, Christensen R, Khorshidi A. Addressing uncertainty in atomistics at the machine learning-accelerated materials discovery stage. Phys Rev Mater. 2021;5:033802. https://doi.org/10.1103/PhysRevMaterials.5.033802.

Schwaller P, Vaucher AC, Laino T, Reymond JL. Prediction of chemical reaction yields using deep learning. Mach Learn Sci Technol. 2021;2:015016. https://doi.org/10.1088/2632-2153/abc81d.

Li L, Chang J, Vakanski A, Wang Y, Yao T, Xian M. Uncertainty quantification in multivariable regression for material property prediction with bayesian neural networks. Sci Rep. 2024;14:61189. https://doi.org/10.1038/s41598-024-61189-x.

Lanini J, Huynh MTD, Scebba G, Schneider N, Rodríguez-Pérez R. Unique: A framework for uncertainty quantification benchmarking. J Chem Inf Model. 2024;64(22):8379-86. https://doi.org/10.1021/acs.jcim.4c01578.

Bilbrey JA, Firoz JS, Lee MS, Choudhury S. Uncertainty quantification for neural network potential foundation models. npj Comput Mater. 2024;10:1572. https://doi.org/10.1038/s41524-025-01572-y.

Perez D, Subramanyam APA, Maliyov I, Swinburne TD. Uncertainty quantification for misspecified machine learned interatomic potentials. npj Comput Mater. 2024;11:758. https://doi.org/10.1038/s41524-025-01758-4.