REVIEW Open access

Failure, Uncertainty, and Risk in Materials AI — How Negative Outcomes Are Handled Across the Literature: A Review Study

Ravi Menon[1*], Arjun Nair[1], Meera Pillai[2], Suresh Varma[1]

Abstract

The integration of artificial intelligence (AI) and machine learning (ML) in materials science has accelerated discovery and design processes, yet it introduces challenges related to failure, uncertainty, and risk. This narrative review examines how the materials AI literature addresses negative outcomes, including model uncertainties, predictive failures, and associated risks in application. Drawing on peer-reviewed studies, we explore uncertainty quantification techniques, robustness evaluations, and risk mitigation strategies. Key themes include Bayesian methods for uncertainty estimation, benchmark studies on prediction reliability, and strategies to handle data scarcity and extrapolation errors. The review highlights gaps in handling adversarial conditions and real-world failures, proposing future directions for more resilient AI frameworks in materials research. By synthesizing these insights, we aim to foster a more cautious and effective use of AI in advancing materials innovation.

Keywords Materials AI, Uncertainty quantification, Risk assessment, Predictive failure, Machine learning robustness, Negative outcomes

*Correspondence:
Ravi Menon
ravi.menon@outlook.com
[1] Department of Materials Engineering and AI Modeling, Faculty of Engineering, University of Kerala, Thiruvananthapuram, India
[2] Department of Intelligent Materials Analytics, Faculty of Engineering, Manipal University, Manipal, India

Introduction

Over the past decade, artificial intelligence (AI) and machine learning (ML) have emerged as transformative tools in materials science, fundamentally reshaping how materials are discovered, characterized, and optimized. By leveraging large datasets generated from experiments, simulations, and materials databases, AI-driven approaches enable rapid prediction of material properties, inverse design of compounds with targeted functionalities, and high-throughput virtual screening at unprecedented scales [1, 2]. These capabilities have accelerated progress across a wide range of applications, including energy storage and conversion materials [1], optoelectronic and fluorescent compounds [3], catalysts, and structural materials, significantly shortening development cycles that previously relied on labor-intensive experimentation and computationally intensive first-principles calculations [4].

The rapid adoption of AI in materials science has been further fueled by advances in model architectures and data representations. Techniques such as graph neural networks, which naturally encode atomic structures and bonding information, have become central to state-of-the-art materials property prediction. In contrast, probabilistic frameworks such as Bayesian neural networks provide a means to incorporate uncertainty into predictions [5, 6]. Large, curated databases—most notably the Materials Project, Open Quantum Materials Database (OQMD), and AFLOW—have

provided the data foundation necessary for training increasingly sophisticated models. As a result, AI-based materials modeling has shifted from proof-of-concept demonstrations to widespread use in both academic research and industrial workflows.

Despite these successes, the growing reliance on AI systems has exposed a critical gap between model performance metrics and real-world reliability. Many ML models achieve impressive accuracy on benchmark datasets yet exhibit fragile behavior when applied beyond their training distributions, such as to chemically novel compositions, extreme operating conditions, or sparse-data regimes [7]. Overconfident predictions, hidden biases in training data, and inadequate treatment of uncertainty can lead to negative outcomes, including misleading scientific conclusions, inefficient allocation of experimental resources, and, in safety-critical domains such as energy storage, aerospace, or structural materials, potentially hazardous design decisions [8, 9]. These concerns underscore the need to move beyond accuracy-centric evaluations toward a more holistic understanding of model failure modes and associated risks.

In this review, negative outcomes are broadly defined to encompass predictive failures, unquantified or mischaracterized uncertainties, lack of robustness to distribution shifts, and the downstream risks that arise from these limitations. A growing body of evidence indicates that uncertainty in materials ML models arises from multiple sources. Epistemic uncertainty reflects limitations in model structure, training data coverage, or representation choices, and is particularly pronounced in extrapolative predictions. Aleatoric uncertainty, by contrast, originates from intrinsic noise in experimental measurements or variability in computational data, such as differences in density functional theory (DFT) settings or convergence criteria [10, 11]. When these uncertainty sources are not explicitly quantified or communicated, users may place unwarranted trust in model outputs, increasing the likelihood of failure.

Although uncertainty quantification (UQ) techniques—such as ensemble methods, Bayesian inference, and conformal prediction—have been proposed and applied in materials informatics, their adoption remains inconsistent, and their limitations are not always critically discussed. Moreover, many studies focus on methodological novelty or benchmark performance, while negative results, model breakdowns, and practical deployment challenges are often underreported. This imbalance has contributed to a fragmented understanding of risk in AI-driven materials research, despite the field's increasing maturity and real-world impact.

The primary aim of this review is to address this gap by providing a systematic and critical synthesis of how negative outcomes are treated in contemporary materials AI literature. Specifically, this review pursues three interrelated objectives: (i) to examine how recent studies identify, characterize, and evaluate failure modes in ML models for materials science, with particular emphasis on uncertainty quantification strategies; (ii) to analyze the risks associated with these negative outcomes, especially in high-stakes applications where erroneous predictions may have significant technical, economic, or safety consequences [12, 13]; and (iii) to summarize and compare mitigation strategies, including robust model design, benchmark-driven evaluation, active learning, and uncertainty-aware decision-making frameworks [14-16].

By focusing on peer-reviewed journal articles published, this review offers a thematic and integrative perspective on reliability, risk, and robustness in AI-enabled materials science. Rather than proposing a single methodological solution, we aim to distill practical insights and best practices that can guide researchers toward more transparent, trustworthy, and deployment-ready AI systems. In doing so, this work seeks to contribute to the ongoing transition of materials AI from a rapidly evolving research frontier to a dependable component of scientific and engineering practice.

Main Text

Uncertainty quantification in materials AI

Uncertainty quantification (UQ) plays a central role in mitigating negative outcomes in AI-driven materials research, as it provides explicit measures of confidence alongside model predictions and enables the identification of potentially unreliable results [5, 8]. In materials science, where predictions often guide costly, time-consuming experimental

validation, the absence of reliable uncertainty estimates can lead to misdirected experimental efforts, inefficient resource use, and erroneous scientific conclusions. This is particularly critical for properties such as formation energies, phase stability, mechanical strength, and electrochemical performance, where even modest prediction errors may lead to fundamentally incorrect material selection or design decisions [6, 15].

Bayesian approaches have emerged as a dominant paradigm for UQ in materials AI, offering principled probabilistic frameworks capable of capturing both epistemic uncertainty, arising from limited data and model inadequacy, and aleatoric uncertainty, stemming from intrinsic noise in experimental or computational datasets [10, 17]. Bayesian neural networks (BNNs), in particular, have been applied to a range of materials modeling tasks, including property prediction under data-scarce conditions, where conventional deterministic models tend to overfit [17]. By treating model parameters as probability distributions rather than fixed values, BNNs provide uncertainty estimates that directly reflect model confidence and knowledge gaps [5]. Benchmark studies have reported substantial improvements in predictive robustness, with Bayesian methods reducing error magnitudes by up to an order of magnitude compared to standard ML approaches in selected materials datasets [16]. Nevertheless, the practical deployment of Bayesian inference remains challenging, as exact posterior estimation is computationally intractable for large-scale datasets. Consequently, commonly used approximations—such as variational inference or Monte Carlo dropout—may introduce biases or underestimate uncertainty, particularly in extrapolative regimes [8].

Gaussian process (GP) models represent another widely used UQ technique in materials informatics, valued for their built-in probabilistic structure and analytical uncertainty estimates [18]. GPs have been successfully applied to benchmark comparisons of UQ strategies, demonstrating superior capability in flagging unreliable predictions and identifying extrapolations beyond the training domain [18]. In studies of polymer property prediction, GP-based models were shown to reduce systematic over- and underestimation of errors, thereby improving decision-making reliability [14]. However, the scalability of GPs remains a major limitation, as their computational cost grows rapidly with dataset size and feature dimensionality—both of which are characteristic of modern materials datasets. To address this, hybrid approaches that integrate GPs with neural network feature extractors have been proposed, aiming to combine expressivity with calibrated uncertainty estimates [5, 11].

Ensemble-based methods constitute a more pragmatic and widely adopted class of UQ techniques, particularly in applied materials research. By aggregating predictions from multiple independently trained models—such as random forests, boosted trees, or deep neural networks—ensembles estimate uncertainty through prediction variance [1, 9]. Reviews of deep learning–based UQ methods highlight ensemble methods as particularly effective at handling data scarcity and heterogeneity, common challenges in domains such as energy materials and alloy design [1]. Empirical studies demonstrate that ensemble models improve robustness in predicting alloy properties by mitigating single-model failures and reducing susceptibility to spurious correlations [2]. Despite their simplicity and scalability, ensemble methods often conflate epistemic and aleatoric uncertainty and may provide poorly calibrated confidence estimates unless carefully designed and validated.

Despite significant methodological progress, the literature reveals persistent gaps in the application and evaluation of UQ in materials AI. Many studies prioritize in-distribution accuracy metrics while overlooking real-world distribution shifts that frequently cause model failures [7]. For example, comparative analyses of models trained on earlier versions of the Materials Project database have shown marked performance degradation when applied to newly reported compounds, underscoring the need for dynamic and adaptive UQ frameworks that evolve with expanding datasets [7]. These findings suggest that uncertainty-aware modeling should be coupled with continual learning and dataset monitoring to remain reliable over time. A comparative overview of uncertainty quantification approaches commonly used in materials AI, including their strengths and limitations, is summarized in **Table 1**.

Table 1. Uncertainty quantification techniques in materials AI: A comparative summary

UQ technique	Description	Strengths	Limitations	Key references
Bayesian neural networks (BNNs)	Probabilistic modeling of parameters to capture epistemic and aleatoric uncertainty.	Captures uncertainty from both data and model; interpretable confidence intervals.	Computationally intensive; approximations may misestimate uncertainty.	[5, 10, 17]
Gaussian processes (GPs)	Non-parametric models with analytical uncertainty estimates.	Strong extrapolation detection; built-in uncertainty.	Scalability issues with large or high-dimensional datasets.	[14, 18]
Ensemble methods	Aggregates predictions from multiple models to estimate uncertainty.	Scalable, easy to implement, and improves robustness.	Conflates uncertainty types; may lack proper calibration.	[1, 2, 9]
Conformal prediction	Calibrates prediction intervals based on empirical error rates.	Distribution-free guarantees; interpretable intervals.	Requires well-characterized residuals; limited in extrapolation.	[8, 15]

Handling predictive failures in ML models for materials

Predictive failures in materials AI commonly arise from overfitting, dataset biases, limited chemical diversity, and extrapolation beyond the domain represented in the training data. Such failures can lead to negative outcomes, including incorrect identification of promising materials, false rejection of viable candidates, and reduced trust in AI-assisted workflows [19, 20]. Recent literature increasingly recognizes the importance of systematically identifying and characterizing these failure modes through robustness testing and targeted evaluation strategies [7, 16].

Benchmark studies have played a crucial role in exposing the disconnect between benchmark performance and real-world generalizability. Several critical assessments demonstrate that models achieving state-of-the-art accuracy on standard benchmarks often fail when evaluated on out-of-distribution samples or chemically novel systems [7]. This phenomenon is frequently attributed to database biases, where common crystal structures, compositions, or chemistries dominate training datasets, leading models to underperform on rare, metastable, or unconventional materials [15]. To address this limitation, researchers have proposed domain-aware validation protocols, such as leave-one-chemistry-out or compositionally stratified cross-validation schemes, which better approximate real-world deployment scenarios [16].

Data scarcity further exacerbates predictive failures, particularly in materials domains where experimental measurements are expensive, slow, or difficult to reproduce [21, 22]. To mitigate this challenge, strategies such as transfer learning, multitask learning, and data augmentation using generative models have gained increasing attention [21, 23]. Generative adversarial networks (GANs) and related approaches have been employed to synthesize plausible material representations and explore under-sampled regions of chemical space, thereby improving model stability and reducing failure rates in inverse design tasks [24]. While promising, these techniques introduce additional complexity, as synthetic data may inadvertently reinforce existing biases or introduce artifacts if not carefully validated.

Model interpretability has emerged as a complementary tool for diagnosing and mitigating predictive failures. Explainable AI techniques, including SHAP and feature attribution analyses, have been applied to materials ML models to identify dominant features driving predictions and to pinpoint sources of systematic error [12, 25]. Such insights enable targeted improvements in feature engineering, data curation, and model architecture. Beyond technical considerations, interpretability is increasingly linked to ethical and societal concerns. In applications such as biomedical or environmentally sensitive materials, biased or opaque models may amplify risks, prompting calls for fairness-aware, transparent AI frameworks in materials science [12, 26].

Collectively, these studies highlight that handling predictive failures requires a multifaceted approach that combines robust evaluation protocols, uncertainty-aware modeling, data-centric strategies, and interpretability tools. Rather than treating failures as isolated anomalies, the literature increasingly frames them as inherent characteristics of data-driven materials discovery—characteristics that must be explicitly managed to enable reliable and responsible AI deployment.

Risk assessment in AI-Driven materials discovery

Risks in materials AI stem from unhandled negative outcomes, including safety hazards, economic losses, and environmental impacts [9, 13]. The literature emphasizes risk assessment frameworks to quantify these, particularly in high-stakes discoveries [13, 27].

In drug-like materials or alloys, predictive risks arise from uncertainties in properties such as toxicity or stability [4, 28]. AI models for heart disease materials highlight risks from unreliable predictions and advocate for UQ to mitigate clinical failures [23]. Similarly, in self-driving labs, risks from autonomous experiments—such as synthesis failures—are addressed through Bayesian optimization with safety constraints [13].

Broader societal risks include bias amplification in models trained on incomplete data, leading to inequitable material innovations [19, 26]. Reviews on AI ethics in materials call for governance to balance risks with benefits [12, 26]. Economic risks from failed predictions are quantified in benchmark studies, where error rates translate to wasted R&D [14, 15].

Mitigation strategies include multi-objective optimization that incorporates risk terms [29]. For instance, models that optimize performance while minimizing uncertainty reduce discovery risk [5, 8].

Case studies of negative outcomes in the literature

Documented case studies of negative outcomes in materials AI provide valuable insights into the limitations of current methodologies and highlight recurring patterns of failure across application domains [7, 20]. Rather than isolated anomalies, these cases reveal systemic issues in data quality, uncertainty handling, and model assumptions, underscoring the need for a more critical evaluation of AI-driven materials workflows.

In the context of energy materials, several studies report failures arising from inadequate treatment of aleatoric uncertainty in experimental and simulation data. In one notable example, ML models trained to predict electrochemical performance exhibited high apparent accuracy yet produced overconfident predictions when exposed to noisy or heterogeneous datasets, ultimately misleading downstream materials screening efforts [1]. These findings demonstrate that neglecting data noise can mask uncertainty and inflate confidence, increasing the risk of negative outcomes in materials selection.

Similar issues have been reported in polymer property prediction, where benchmark analyses revealed systematic underestimation of prediction errors for structurally diverse polymers [14]. Although uncertainty quantification methods appeared effective on homogeneous subsets of data, their performance deteriorated when applied to broader chemical spaces, exposing limitations in UQ calibration. Such failures emphasize the importance of evaluating models under chemically and structurally diverse conditions, rather than relying solely on narrowly defined benchmarks.

Alloy design presents another prominent example of negative outcomes driven by distribution shifts. Models trained on historical alloy datasets frequently perform poorly when tasked with predicting properties of newly explored compositions or compositional extremes [7]. These failures have been attributed to imbalanced training data and implicit assumptions of compositional continuity, which break down in unexplored regions of chemical space. To address this, some studies have introduced adversarial or stress-test datasets that deliberately challenge model assumptions, revealing vulnerabilities that are otherwise hidden in conventional evaluations [11, 25].

Negative outcomes are not limited to predictive models but also extend to generative approaches used for inverse materials design. Generative models, including variational autoencoders and generative adversarial networks, have been shown to produce “hallucinated” materials that appear valid in latent space but are physically infeasible or synthetically inaccessible [24]. Case studies in inverse design report synthesis failures when generated candidates violate thermodynamic stability constraints or ignore processing limitations, highlighting the risks of decoupling generative models from physical and chemical knowledge [4, 21]. These examples illustrate that generative success metrics alone are insufficient without validation against domain-specific feasibility criteria.

Collectively, these case studies reveal that negative outcomes in materials AI often stem from a combination of unaccounted-for uncertainty, dataset bias, extrapolation beyond the training domain, and insufficient integration of physical constraints. Importantly, they demonstrate that failure analysis itself is a powerful tool for methodological improvement and should be more systematically reported in the literature. Representative failure modes and associated risk categories reported across different materials AI application domains are synthesized in **Table 2**.

Table 2. Documented failure modes and risk factors in materials AI: case study highlights

Application domain	Failure mode description	Risk type	Source of negative outcome	Reference
Energy materials	Overconfident predictions despite noisy training data	Scientific misdirection	Aleatoric uncertainty not handled	[1]
Polymer modeling	Systematic error under diverse chemical conditions	Predictive inaccuracy	UQ calibration degradation outside benchmarks	[14]
Alloy design	Poor generalization to new compositions	Material rejection/failure	Distribution shift, data imbalance	[7]
Inverse design	Generation of physically infeasible candidates	Experimental failure	Lack of physical constraints in generative models	[4, 24]
Self-driving labs	Hazardous outcomes during automated synthesis	Physical/laboratory safety risks	Lack of real-time risk monitoring and safety layers	[13]

Strategies to mitigate negative outcomes

In response to the growing recognition of negative outcomes, the literature proposes a range of strategies to enhance the robustness, reliability, and trustworthiness of AI models in materials science [16, 22]. These approaches increasingly emphasize uncertainty awareness, domain integration, and rigorous evaluation as core design principles rather than optional add-ons.

Active learning has emerged as a particularly effective strategy for mitigating failures caused by data scarcity and uneven chemical coverage. By explicitly incorporating uncertainty estimates into data-acquisition workflows, active learning prioritizes experiments or simulations expected to yield the greatest reduction in model uncertainty [5, 18]. Studies demonstrate that such approaches not only improve predictive accuracy with fewer data points but also reduce the likelihood of catastrophic failures in under-sampled regions of materials space.

Physics-informed and domain-aware modeling frameworks represent another important direction for improving generalizability and reducing negative outcomes. By embedding physical laws, thermodynamic constraints, or mechanistic insights into model architectures or loss functions, these approaches limit the solution space to physically plausible regimes [17, 26]. Evidence suggests that physics-informed models are more stable under distribution shifts and less prone to producing nonsensical predictions than purely data-driven counterparts.

Hybrid and ensemble-based approaches further enhance robustness by combining complementary modeling paradigms. For example, Bayesian–Gaussian process hybrids integrate expressive feature learning with calibrated uncertainty estimates, offering improved performance in both interpolation and extrapolation tasks [8, 10]. Ensemble strategies similarly reduce sensitivity to individual model biases and have been shown to mitigate failures arising from noisy or heterogeneous datasets.

At the evaluation level, standardizing benchmarks and reporting practices has been identified as a critical step toward reducing negative outcomes. Benchmark suites such as Matbench promote consistent, transparent, and reproducible model comparison, helping to expose weaknesses that may otherwise remain obscured by selective reporting [16]. However, several authors argue that benchmarks should evolve to include stress-testing scenarios, out-of-distribution evaluation, and uncertainty calibration metrics to reflect real-world deployment conditions better.

Looking forward, future mitigation strategies increasingly emphasize interpretability and ethical considerations. Explainable AI tools, including SHAP and related attribution methods, enable failure diagnosis by revealing the features and data patterns driving erroneous predictions [25]. At the same time, responsible AI frameworks advocate for transparency, fairness, and risk awareness, particularly in high-impact applications such as biomedical, environmental, and energy-related materials [9, 12]. Together, these developments signal a shift from performance-centric optimization toward reliability-centered AI design in materials science. **Figure 1** conceptually illustrates how data limitations and modeling assumptions propagate through predictive failures and uncertainty mischaracterization, generating downstream technical and societal risks, alongside corresponding mitigation strategies.

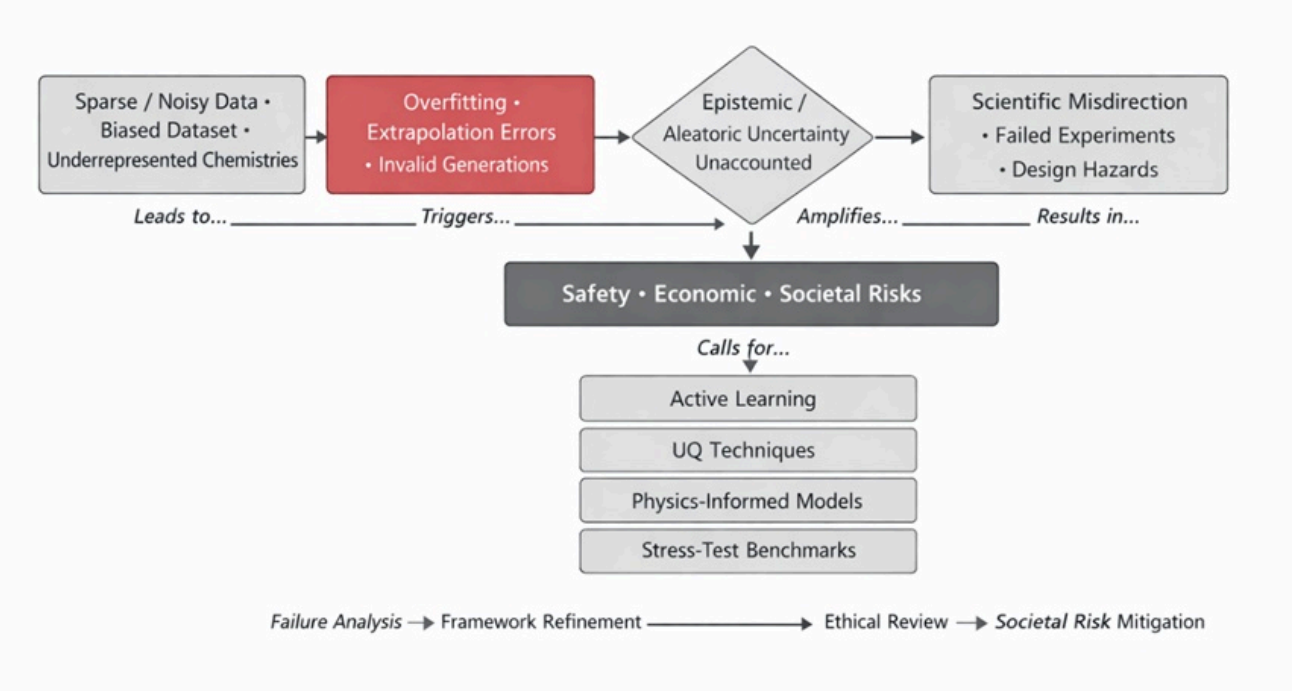

Figure 1. Failure–uncertainty–risk cascade in materials AI

Results and Discussion

The literature on materials AI reveals a maturing recognition of negative outcomes, yet significant disparities exist in how failure, uncertainty, and risk are addressed [5, 8, 26]. While uncertainty quantification (UQ) methods such as Bayesian neural networks and ensemble approaches have improved predictive reliability [5, 6, 10, 11], their adoption remains inconsistent. For instance, benchmark studies demonstrate that UQ can mitigate overconfidence in material property

predictions [14, 15, 18], but many applications overlook epistemic uncertainties arising from data limitations [7, 21]. This gap is particularly evident in generative models, where failures, such as infeasible compound generation, are common yet risk assessments are rarely integrated [4, 24].

A critical issue is the handling of predictive failures in real-world contexts. Studies show that models excel on controlled benchmarks but falter under distribution shifts, leading to negative outcomes such as erroneous material designs [7, 16]. This underscores the need for robustness evaluations that simulate adversarial conditions, as seen in engineering prognostics, where UQ tutorials emphasize propagating uncertainties to prevent cascading errors [8]. However, the literature often prioritizes accuracy over resilience, with few papers addressing ethical risks such as biased innovations that could exacerbate societal inequities [12, 19, 26]. In high-stakes areas like self-driving laboratories, unmitigated risks from AI-driven experiments—such as synthesis hazards—pose tangible dangers, calling for interdisciplinary frameworks that incorporate safety engineering principles [9, 13].

Comparatively, UQ in materials AI lags behind fields like healthcare, where machine learning risks are more rigorously quantified [23, 27, 28]. For example, while materials benchmarks focus on error metrics [15, 16], clinical AI emphasizes uncertainty-aware decisions to avoid harm [23]. This disparity suggests opportunities for cross-pollination, such as adapting Gaussian processes from engineering to materials for better extrapolation [18, 28]. Data preparation challenges further compound negative outcomes; incomplete or noisy datasets lead to failures, yet strategies like active learning are underutilized [17, 22, 25]. Reviews on machine learning in materials highlight the importance of scalable paradigms, yet practical implementations often ignore the uncertainty associated with data curation [1, 2, 17].

Thematically, the discourse on risk mitigation is fragmented. Some advocate for physics-informed models to reduce failures [3, 26], while others emphasize ensemble methods for broader coverage [1, 9]. However, a unified approach is lacking, and risks in emerging ecosystems—such as AI-driven materials intelligence—are often downplayed [2]. Bibliometric analyses reveal a surge in AI safety discussions, but materials-specific risks, such as environmental impacts arising from flawed predictions of energy materials, receive scant attention [9, 29]. This highlights a broader tension: the push for rapid discovery [13] versus the imperative for cautious deployment.

Conclusion

In summary, the reviewed literature illustrates progress in addressing failure, uncertainty, and risk in materials AI through advanced UQ techniques and robustness strategies. However, persistent gaps in handling negative outcomes—such as inadequate risk assessments and inconsistent UQ application—hinder reliable progress. By synthesizing these insights, this review underscores that while AI accelerates materials innovation, unaddressed negatives can undermine trust and efficacy.

Future directions

To enhance resilience, future research should prioritize standardized UQ benchmarks tailored to materials data challenges. Integrating ethical frameworks early in AI pipelines could mitigate societal risks. Exploring hybrid models that combine deep learning with domain-specific constraints may reduce the incidence of extrapolation failures. Additionally, self-driving labs should incorporate real-time risk monitoring and collaborative efforts across disciplines—such as adopting health prognostics—so UQ can foster more robust systems. Ultimately, emphasizing negative outcomes will pave the way for sustainable AI in materials science.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 08 Oct 2025 Revised: 08 Nov 2025 Accepted: 07 Dec 2025
Published online: 18 January 2026

References

Chen C, Zuo Y, Ye W, Li X, Deng Z, Ong SP. A critical review of machine learning of energy materials. Adv Energy Mater. 2020;10(8):1903242.
https://doi.org/10.1002/aenm.201903242.

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.

Ramprasad R, Beaudoin C, Behera S. Machine learning with limited data for materials design. MRS Bull. 2021;46:1023-31.
https://doi.org/10.1557/s43577-021-00195-3.

Batra R, Song L, Ramprasad R. Emerging materials intelligence ecosystems propelled by machine learning. Nat Rev Mater. 2021;6:655-78.
https://doi.org/10.1038/s41578-020-00255-y.

Psaros AF, Meng X, Zou Z, Guo L, Karniadakis GE. Uncertainty quantification in scientific machine learning: Methods, metrics, and comparisons. J Comput Phys. 2023;477:111902.
https://doi.org/10.1016/j.jcp.2022.111902.

Olivier A, Shields MD, Graham-Brady L. Bayesian neural networks for uncertainty quantification in data-driven materials modeling. Comput Methods Appl Mech Eng. 2021;386:114079.
https://doi.org/10.1016/j.cma.2021.114079.

Li Y, DeCost B, Choudhary K, Hattrick-Simpers J. A critical examination of robustness and generalizability of machine learning prediction of materials properties. npj Comput Mater. 2023;9:55.
https://doi.org/10.1038/s41524-023-01012-9.

Hu Y, Mahadevan S. Uncertainty quantification in machine learning for engineering design and health prognostics: A tutorial. Mech Syst Signal Process. 2023;205:110796.
https://doi.org/10.1016/j.ymssp.2023.110796.

Tamascelli N, Campari A, Parhizkar T, Paltrinieri N. Artificial intelligence for safety and reliability: A descriptive, bibliometric and interpretative review on machine learning. J Loss Prev Process Ind. 2024;90:105343.
https://doi.org/10.1016/j.jlp.2024.105343.

Dai J, Adhikari S, Wen M. Uncertainty quantification and propagation in atomistic machine learning. Rev Chem Eng. 2024;40(3):435-56.
https://doi.org/10.1515/revce-2024-0028.

Zhang H, Chen W, Iyer A, Apley DW, Chen W. Uncertainty-aware mixed-variable machine learning for materials design. Sci Rep. 2022;12:19760.
https://doi.org/10.1038/s41598-022-23431-2.

Chen Z, Chen C, Yang G, He X, Chi X, Zeng Z, et al. Research integrity in the era of artificial intelligence: Challenges and responses. Medicine. 2024;103(27):e38811.
https://doi.org/10.1097/MD.0000000000038811.

Tom G, Schmid SP, Baird SG, Cao Y, Darvish K, Hao H, et al. Self-driving laboratories for chemistry and materials science. Chem Rev. 2024;124(16):9633-732.
https://doi.org/10.1021/acs.chemrev.4c00055.

Tavazza F, DeCost B, Choudhary K. Uncertainty prediction for machine learning models of material properties. ACS Omega. 2021;6(48):32431-40.
https://doi.org/10.1021/acsomega.1c03752.

Chen S, Guo Y, Gong C, Jia Y, Rookey K, Yang S. Materials property prediction with uncertainty quantification: A benchmark study. Appl Phys Rev. 2023;10(2):021409.
https://doi.org/10.1063/5.0141920.

Jain A. Machine learning in materials research: Developments over the last decade and challenges for the future. Curr Opin Solid State Mater Sci. 2024;33:101189.
https://doi.org/10.1016/j.cossms.2024.101189.

DeCost BL, Hattrick-Simpers JR, Trautt Z. Scientific AI in materials science: A path to a sustainable and scalable paradigm. Mach Learn Sci Technol. 2020;1:033001.

Tran K, Neiswanger W, Yoon J, Zhang Q, Xing E, Ulissi ZW. Methods for comparing uncertainty quantifications for material property predictions. Mach Learn Sci Technol. 2020;1(2):025006.

Adewale MD, Azeta A, Abayomi-Alli A, Sambo-Magaji A. Impact of artificial intelligence adoption on students' academic performance in open and distance learning: A systematic literature review. Heliyon. 2024;10(22):e40025.
https://doi.org/10.1016/j.heliyon.2024.e40025.

Ghavamzadeh M, Fieguth P, Cao X, Khosravi A, Acharya UR, Liu M. 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.

Zhang Y, Ling C. A strategy to apply machine learning to small datasets in materials science. npj Comput Mater. 2020;6:1-8. https://doi.org/10.1038/s41524-020-0323-6.

Pilania G. Machine learning in materials science: From discovery to optimization. Annu Rev Chem Biomol Eng. 2021;12:499-524. https://doi.org/10.1146/annurev-chembioeng-102420-015953.

Vigneswaran J, et al. The role of artificial intelligence in improving outcomes in heart disease: A scientific statement from the American Heart Association. Circulation. 2024;149(12):e1201-21. https://doi.org/10.1161/CIR.0000000000001201.

Dan Y, Zhao Y, Li X, Li S, Hu M, Hu J. Generative adversarial networks (GAN) based efficient sampling of chemical composition space for inverse molecular design. Comput Theor Chem. 2020;1162:112503. https://doi.org/10.1016/j.comtc.2020.112503.

Lu Y, Wang H, Zhang L, Yu N, Shi S, Su H. Unleashing the power of AI in science-key considerations for materials data preparation. Sci Data. 2024;11:1039. https://doi.org/10.1038/s41597-024-03821-z.

Lopez C. Artificial intelligence and advanced materials. Adv Mater. 2023;35(20):2208683. https://doi.org/10.1002/adma.202208683.

Jering KS, Campagnari C, Claggett B, Adler E, Klein L, Ahmad FS, et al. Improving clinical trial efficiency using a machine learning-based risk score to enrich study populations. Eur J Heart Fail. 2022;24(8):1418-26. https://doi.org/10.1002/ejhf.2528.

Moon J, et al. A survey on machine learning approaches for uncertainty quantification of engineering systems. J Reliab Intell Environ. 2024;10:1-25. https://doi.org/10.1007/s44379-024-00011-x.

Tran A, Furlan JM, Pagalthivarthi V, Viswanathan K, Liu S, Wildey T, et al. eSTUNet: Embedded spatio-temporal U-Net for robust video segmentation. Mach Learn Sci Technol. 2023;4:025020.