ORIGINAL RESEARCH **Open access**

When Accuracy Is Not Enough: A Decision-Theoretic Framework for Evaluating Materials AI Models

Li Zhang[1*], Wei Chen[1]

Abstract

In the rapidly evolving field of materials science, artificial intelligence (AI) models have become integral to accelerating discovery and design processes. Yet, their evaluation often relies on simplistic accuracy measures that overlook the broader decision-making contexts. This conceptual paper develops a novel decision-theoretic framework for assessing materials AI models, integrating utility considerations, risk dynamics, and epistemic uncertainties to provide a more holistic understanding of model performance. By synthesizing recent literature on AI applications in materials science and decision theory, the framework interprets model outputs not merely as predictions but as inputs to decision processes where trade-offs between precision, computational efficiency, and real-world applicability shape outcomes. It explores analytical implications, including how utility-based evaluations reveal the interaction between model reliability and stakeholder priorities, fostering systems-level insights into AI's role in sustainable materials innovation. Ethical reasoning is woven throughout, highlighting epistemic challenges in interpreting model behaviors under uncertainty. This approach steers away from isolated metric assessments toward integrative evaluations that align AI capabilities with the multifaceted demands of materials engineering, ultimately enhancing the trustworthiness and utility of AI-driven advancements. The framework's interpretive lens offers pathways to refine evaluation practices, ensuring that AI models contribute meaningfully to decision-making in high-impact applications.

Keywords Materials AI, Epistemic uncertainty, Decision theory, Model evaluation, Utility dynamics, Risk Trade-offs

*Correspondence:
Li Zhang
li.zhang@outlook.com
[1] Department of Computational Materials Science, School of Materials Engineering, Tsinghua University, Beijing, China

Introduction

The integration of artificial intelligence (AI) into materials science has transformed the landscape of research and development, enabling unprecedented speeds in the exploration of material properties and behaviors. Over the past few years, AI models have facilitated the analysis of vast datasets, from atomic structures to macroscopic performance, allowing researchers to navigate complex parameter spaces that were previously intractable through traditional computational methods alone [1]. This shift is particularly evident in areas such as alloy design, where machine learning algorithms interpret intricate relationships between composition and functionality, or in the prediction of material stability under extreme conditions, where neural networks synthesize patterns from simulation outputs [2]. Such advancements underscore the growing reliance on AI to bridge the gap between theoretical models and practical applications, fostering innovation across sectors ranging from energy storage to biomedical implants [3].

However, as AI increasingly permeates materials science, the mechanisms for evaluating these models require scrutiny. Conventional approaches often prioritize accuracy as the primary indicator of success, measuring how closely predictions

align with known benchmarks. While this metric provides a straightforward gauge of predictive fidelity, it frequently abstracts away from the decision contexts in which these models operate [4]. In materials engineering, decisions involve not just forecasting properties but also weighing implications for manufacturing feasibility, environmental impact, and safety protocols. For instance, an AI model might accurately predict a material's tensile strength. Yet, if it underestimates variability in real-world conditions, the resulting decisions could lead to suboptimal or even hazardous outcomes [5]. This discrepancy highlights a fundamental tension: accuracy alone does not capture a model's utility for steering choices amid uncertainty, where errors carry differential costs depending on the application.

Decision theory offers a compelling lens for addressing these limitations by focusing on how choices are made under incomplete information, incorporating preferences, probabilities, and consequences [6]. Rooted in foundational concepts from economics and operations research, decision theory interprets actions in terms of expected outcomes, enabling nuanced assessments of trade-offs. In the context of AI evaluation, this perspective shifts emphasis from isolated performance scores to the model's role within broader decision systems [7]. Recent explorations in machine learning have begun to apply decision-theoretic principles to enhance model robustness, particularly in domains where decisions affect resource allocation or risk management [8]. For materials AI, this means considering how model predictions interact with user objectives, such as minimizing energy consumption in synthesis processes or maximizing durability in structural components.

The epistemic challenges inherent in materials data further complicate evaluation. Materials datasets often exhibit heterogeneity, arising from diverse experimental conditions, measurement errors, and scale dependencies—from nanoscale defects to bulk properties [9]. AI models trained on such data must navigate these uncertainties, yet accuracy metrics treat all discrepancies uniformly, ignoring the differential impacts on decision quality. A decision-theoretic approach, by contrast, integrates utility functions that assign values to outcomes based on their alignment with goals, thereby revealing dynamics in which certain prediction errors are more tolerable than others [10]. For example, overestimating a material's thermal conductivity might be preferable in conservative design scenarios, whereas underestimation could compromise safety thresholds.

Ethical dimensions also permeate this discourse, as AI-driven decisions in materials science influence societal outcomes, including sustainable resource use and equitable access to advanced technologies [11]. Epistemic reasoning prompts consideration of how model evaluations reflect underlying assumptions about knowledge generation, urging frameworks that account for biases in training data or algorithmic opacity. Systems-level insights emerge when viewing AI not as a standalone tool but as part of interconnected feedback structures, in which model outputs inform iterative refinements to experimental design [12].

This paper advances a conceptual framework that embeds decision theory into the evaluation of materials AI models, prioritizing interpretive depth over empirical validation. By examining interaction dynamics between model performance and decision utilities, it elucidates trade-offs that traditional metrics obscure. For instance, in high-stakes applications such as nuclear materials, the framework interprets accuracy through a risk-averse lens, where false negatives incur disproportionate costs [13]. Similarly, in sustainable materials development, it highlights steering logics that balance predictive precision with environmental trade-offs [14].

The need for such a framework is amplified by the accelerating pace of AI adoption in materials science. Publications on AI-assisted materials discovery have surged, reflecting investments in computational infrastructure and collaborative platforms [15]. Yet, literature syntheses reveal persistent gaps in evaluation practices, with many studies defaulting to benchmark comparisons that fail to address the complexities of real-world decision-making [16]. This oversight risks perpetuating models that excel in controlled settings but falter in dynamic environments, underscoring the imperative for integrative approaches.

Analytically, the framework proposed here interprets AI model efficacy through utility-based interactions, where performance is not a static attribute but a relational property shaped by contextual demands. This perspective fosters

epistemic humility, acknowledging that no single metric encapsulates the multifaceted nature of materials challenges [17]. Instead, it advocates for evaluations that reveal feedback structures, such as how model uncertainties propagate through decision chains, potentially amplifying or mitigating risks.

In summary, while accuracy remains a foundational element, it is insufficient for capturing the decision-theoretic nuances essential to materials AI. This introduction sets the stage for a deeper synthesis of theoretical backgrounds, leading to a proposed framework that reorients evaluation toward holistic, utility-driven insights. By doing so, it aims to enhance the alignment between AI capabilities and the interpretive demands of materials science decision-making.

Theoretical Background and Literature Synthesis

The evolution of artificial intelligence in materials discovery and design

Since approximately 2020, the role of artificial intelligence (AI) in materials science has undergone a substantive expansion, both conceptually and methodologically. Early applications were largely confined to supervised learning tasks, particularly property prediction from curated datasets, where models functioned primarily as statistical approximators of structure–property relationships. While these approaches demonstrated clear efficiency gains over traditional trial-and-error experimentation, they were epistemically narrow, emphasizing predictive accuracy without fundamentally reshaping the paradigms of materials discovery.

Recent literature documents a transition toward more integrative and generative AI frameworks, wherein models no longer merely interpolate within known chemical spaces but actively participate in the synthesis of novel material candidates. Generative design approaches—leveraging variational autoencoders, generative adversarial networks, and reinforcement learning—enable exploration of high-dimensional chemical and structural manifolds by encoding latent representations that capture underlying regularities in materials data [18]. This shift marks a conceptual departure from AI as an auxiliary tool toward AI as a co-creative agent within materials innovation pipelines.

A parallel evolution is evident in representational strategies. Graph-based learning architectures, particularly graph neural networks, have gained prominence by modeling atomic structures as relational graphs rather than fixed-length descriptors. This formalism aligns more closely with physical realities of interatomic interactions, enabling AI systems to encode symmetry, locality, and bonding constraints while supporting transferability across material classes [8]. As a result, AI models increasingly facilitate systems-level reasoning, linking microstructural configurations to emergent macroscopic behaviors, such as mechanical resilience or thermal stability in advanced composites.

Concurrently, the literature emphasizes the growing entanglement of AI with both computational simulation and experimental practice. Hybrid workflows integrate first-principles calculations with data-driven learning to accelerate materials screening and hypothesis generation, thereby reducing computational cost while maintaining acceptable fidelity [19]. However, these integrations also expose epistemic and ethical tensions. Scholars increasingly interrogate the provenance and representativeness of training data, highlighting risks of epistemic inequity when AI systems are trained on narrow, regionally biased, or historically contingent datasets that fail to reflect global materials challenges [20].

Moreover, trade-offs between model complexity, interpretability, and resource consumption have become central to theoretical debates. While deeper architectures and richer representations promise improved expressivity, they often exacerbate opacity and computational burden, raising questions about sustainable deployment and scientific accountability. The literature thus increasingly frames AI in materials science not merely as a technical advancement, but as a socio-technical system embedded within constraints of interpretability, equity, and environmental responsibility.

Traditional evaluation metrics in machine learning for materials applications

Evaluation practices in materials-focused machine learning have historically mirrored those of general predictive modeling, relying heavily on scalar accuracy metrics such as mean absolute error, root mean square error, and coefficient of determination [21]. These measures offer standardized, easily comparable indicators of predictive fidelity, particularly in regression tasks involving electronic, mechanical, or thermodynamic properties.

Within the dominant paradigm, model performance is assessed by minimizing deviation from reference values, implicitly assuming that all errors are equivalent and that predictive accuracy constitutes the primary indicator of scientific value. While this approach provides a necessary baseline for benchmarking, recent literature has criticized its conceptual limitations for guiding materials decision-making [22]. Accuracy metrics abstract away from contextual relevance, offering little insight into how prediction errors interact with experimental risk, cost, or downstream engineering consequences.

Importantly, materials data are characterized by heterogeneous sources of uncertainty, including experimental noise, synthesis variability, and model approximation error. Conventional accuracy metrics typically collapse these distinctions, obscuring epistemic uncertainty and fostering overconfidence in model outputs [23]. As a result, high benchmark performance may coexist with poor generalization to unexplored compositional regimes, particularly when models are deployed beyond their implicit domain of applicability [24].

Systems-level analyses increasingly argue that evaluation should be understood as a feedback mechanism rather than a terminal assessment. Prediction errors propagate through experimental design, resource allocation, and deployment decisions, amplifying or attenuating risk depending on context. From this perspective, traditional metrics provide an incomplete account of model quality, as they fail to capture how performance interacts with downstream material optimization pipelines.

Limitations of accuracy-centric evaluation paradigms

An accuracy-centric evaluation framework, while computationally convenient, exhibits fundamental limitations when applied to the complex decision environments of materials science. By prioritizing numerical proximity to ground truth, such approaches neglect asymmetries in consequence, where identical errors may entail radically different risks depending on the application domain. For instance, overconfident predictions in safety-critical materials—such as biomedical implants or structural alloys—may result in disproportionately severe failures relative to modest accuracy gains [25].

Recent conceptual syntheses emphasize that accuracy alone cannot account for utility dynamics inherent in materials decision-making [26]. False positives and false negatives impose unequal costs, shaped by factors such as experimental expense, environmental impact, and human safety. Yet conventional metrics treat these outcomes symmetrically, encouraging optimization strategies misaligned with real-world priorities.

From an epistemic standpoint, accuracy-based evaluation presumes the existence of a singular, stable truth. In practice, many material properties exhibit probabilistic behavior contingent on processing conditions, environmental exposure, and scale effects [27]. By ignoring this plurality, accuracy-centric models risk encoding brittle assumptions that undermine robustness and ethical accountability.

These limitations extend beyond technical performance to normative considerations. Literature increasingly links evaluation practices to broader ethical trade-offs, arguing that models optimized solely for accuracy may inadvertently privilege high-performance but environmentally unsustainable materials, thereby reinforcing narrow optimization logics [28]. The interaction between model outputs and human decision-makers further complicates evaluation, as apparent accuracy can amplify uncertainty when users misinterpret predictions as certainties. **Table 1** contrasts conventional accuracy-centric evaluation paradigms with the proposed decision-theoretic framework, highlighting differences in evaluative focus, treatment of uncertainty, ethical scope, and systems integration.

Table 1. Comparison of accuracy-centric and decision-theoretic evaluation paradigms in materials AI

Dimension	Accuracy-centric evaluation	Decision-theoretic evaluation (proposed framework)
Primary evaluative goal	Minimize prediction error relative to ground truth	Maximize decision utility under uncertainty
Core performance indicator	Scalar metrics (MAE, RMSE, R^2)	Utility-weighted outcomes across decision scenarios
Treatment of uncertainty	Implicit or collapsed into residual error	Explicitly modeled and propagated through decisions
Interpretation of errors	Symmetric and context-agnostic	Asymmetric and consequence-dependent
Relationship to decision-making	Indirect; evaluation detached from the use context	Direct evaluation embedded in decision processes
Handling of epistemic limits	Largely ignored or treated as noise	Central to interpretation and utility modulation
Ethical considerations	External or post-hoc	Embedded via value-sensitive utility assignments
Systems-level feedback	Absent or weak	Explicit feedback loops enabling iterative refinement
Suitability for high-stakes applications	Limited, prone to overconfidence	High, supports risk-aware and resilient decisions

Collectively, these critiques underscore the need to reconceptualize evaluation as a relational and context-sensitive process—one that situates predictive performance within systems of uncertainty, consequence, and value rather than treating accuracy as an isolated objective.

Foundations of decision theory in uncertain environments

Decision theory provides a robust foundation for navigating uncertainty, interpreting choices through utility assignments and probabilistic expectations [29]. In uncertain environments, it emphasizes trade-offs, where decisions balance potential gains against risks, fostering epistemic awareness of incomplete information [30]. Conceptual interpretations apply this to AI, viewing models as decision aids that integrate preferences into evaluative structures.

Recent work has synthesized decision theory with machine learning, revealing dynamics in which utility functions guide model assessments beyond mere prediction [31]. Ethical reasoning is relevant here, as it questions how utilities reflect societal values in resource-constrained settings [32]. Systems-level insights emerge from feedback loops, where iterative decisions refine utilities, steering AI toward adaptive behaviors in volatile contexts.

Applications of decision theory in AI and engineering

In AI and engineering, decision theory has been applied to optimize under constraints, such as in robotic systems, where utility-based evaluations interpret sensor data for path planning [33]. Materials engineering parallels this, with decision-theoretic models assessing trade-offs in alloy composition between performance and cost [30].

Literature syntheses highlight integrative potentials, where decision theory reveals interaction dynamics in multi-objective optimization, such as balancing strength and ductility [30]. Epistemic challenges include incorporating human judgments into utility functions and fostering ethical alignment in AI-driven engineering [31]. Steering logics from these applications suggest frameworks that interpret AI outputs as components of larger decision ecosystems.

Synthesis: Towards integrated evaluation frameworks

In summary, traditional metrics and decision theory converge toward integrated frameworks that interpret AI performance through the lenses of utility and risk. Analytical implications include enhanced system insights, with trade-offs revealing hidden dynamics in materials AI [30]. Ethical reasoning advocates for epistemic inclusivity, ensuring evaluations reflect diverse decision contexts [31]. This synthesis aims to advance conceptual understanding that amplifies AI's utility in materials science.

The proposed conceptual framework advances a decision-theoretic reinterpretation of how artificial intelligence models in materials science should be evaluated. Rather than treating predictive performance as an isolated numerical outcome, the framework situates model outputs within decision environments, emphasizing utility dynamics, risk trade-offs, and epistemic interactions that collectively shape real-world material outcomes. In this view, AI performance is not an intrinsic property of a model, but an emergent characteristic arising from interactions between predictions, uncertainties, stakeholder objectives, and downstream decisions.

At its core, the framework reconceptualizes AI predictions as decision inputs rather than terminal outputs. Model estimates of material properties—such as durability, conductivity, or degradation rates—are interpreted through utility functions that encode stakeholder-specific values, including cost efficiency, safety margins, environmental sustainability, and regulatory compliance. This shift foregrounds the insight that identical predictive accuracies may yield divergent practical value depending on how predictions are operationalized within material selection, experimental prioritization, or deployment strategies.

Utility mapping and trade-off sensitivity

A central element of the framework is the explicit incorporation of utility functions that map predictive outcomes to valued consequences. These functions enable systematic analysis of trade-offs that are otherwise obscured by accuracy-centric metrics. For example, in evaluating AI models for polymer degradation forecasting, a high-accuracy prediction may be interpreted differently under distinct utility regimes. In biomedical contexts, risk-averse utilities may assign higher value to conservative overestimations of degradation to prioritize patient safety. In contrast, industrial recycling applications may favor cost-minimizing utilities that tolerate higher uncertainty.

By formalizing these mappings, the framework enables analytical exploration of how performance metrics interact with application-specific risk profiles. Importantly, utility is not treated as static: it is context-dependent, stakeholder-mediated, and sensitive to downstream consequences. This perspective exposes the implicit value judgments embedded in model evaluation practices and renders them analytically transparent.

Epistemic uncertainty and interpretive flexibility

Epistemic reasoning plays a foundational role within the framework. Materials AI systems routinely operate under conditions of data sparsity, measurement noise, and representational bias, all of which introduce uncertainty into predictions. Rather than treating uncertainty as a nuisance to be minimized, the framework positions it as a first-class interpretive signal that modulates utility assignments and decision thresholds.

Under this lens, uncertainty estimates inform how confidently predictions should influence action. Flexible utility structures allow decision-makers to adjust tolerance levels for uncertainty depending on application criticality, fostering robustness rather than brittle optimization. This epistemic sensitivity prevents overconfident decision steering and supports adaptive strategies when models are extrapolated beyond well-characterized material regimes.

Interaction dynamics and feedback structures

The framework further emphasizes interaction dynamics through explicit feedback loops between model evaluation and system refinement. Evaluation is conceptualized as an iterative process: initial predictions are assessed within simulated or real decision scenarios, their realized utility is examined, epistemic gaps are identified, and models or decision policies are subsequently adjusted. This cyclical structure transforms evaluation from a static benchmarking exercise into a learning mechanism embedded within materials workflows.

Such feedback structures highlight how evaluation choices influence data acquisition priorities, model retraining strategies, and experimental design, thereby shaping the evolution of the entire materials discovery pipeline. Performance, in this sense, is co-produced by models and their institutional and technical environments.

Ethical integration and steering logics

Ethical considerations are not treated as external constraints but as integral components of evaluative reasoning. The framework explicitly calls for interpretations that balance technical precision with societal and environmental consequences, including equitable access to advanced materials, sustainability trade-offs, and long-term risk exposure. By embedding these considerations within utility functions and decision criteria, ethical values become operational rather than aspirational.

Finally, the framework introduces steering logics grounded in multi-objective optimization. Trade-offs between computational efficiency, predictive depth, interpretability, and environmental cost are analyzed for their implications across materials workflows. This orientation supports resilient decision-making in uncertain environments, where no single metric can adequately capture model value.

Conceptual contribution

By reframing evaluation as a decision-theoretic, utility-mediated, and epistemically grounded process, the proposed framework provides a coherent conceptual basis for aligning materials AI with the interpretive demands of scientific practice. It moves beyond accuracy as a dominant evaluative norm and offers a structured lens for understanding how AI systems meaningfully contribute to materials discovery, design, and deployment under uncertainty. **Figure 1** illustrates the proposed decision-theoretic evaluation framework, emphasizing how model predictions, utility assignments, decision processes, and outcome assessments interact through iterative feedback to shape materials AI performance at the systems level.

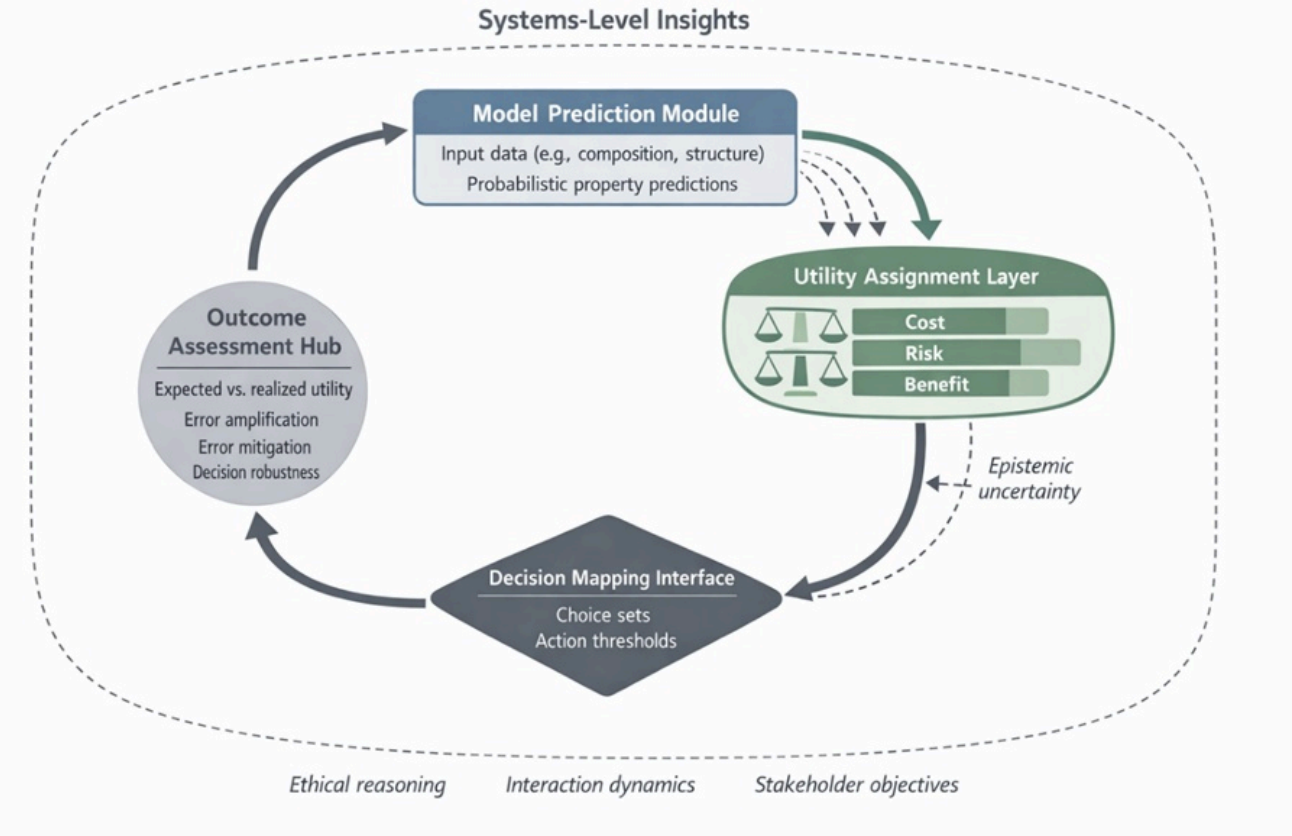

Figure 1. Conceptual decision-theoretic framework for evaluating materials AI models as components of decision ecosystems rather than isolated predictors. The framework emphasizes utility assignment, epistemic uncertainty, feedback, and systems-level interpretation.

Table 2 summarizes the core components of the proposed decision-theoretic framework, detailing their functional roles, epistemic implications, and contributions to systems-level evaluation.

Table 2. Components of the decision-theoretic evaluation framework for materials AI

Framework component	Primary function	Epistemic role	Analytical contribution
Model prediction module	Generates probabilistic estimates of material properties	Encodes data limitations and model assumptions	Serves as an input source for downstream decision interpretation
Utility assignment layer	Maps predictions to valued consequences	Translates uncertainty into context-sensitive preferences	Reveals trade-offs between risk, cost, benefit, and sustainability
Decision mapping interface	Integrates utilities into actionable choice sets	Mediates uncertainty through thresholds and selection rules	Aligns AI outputs with real-world decision constraints
Outcome assessment hub	Compares expected vs. realized utilities	Identifies epistemic gaps and misalignments	Enables feedback-driven refinement of models and utilities
Systems-level enclosure	Contextualizes evaluation across stakeholders and values	Surface's ethical and interaction dynamics	Ensures holistic interpretation beyond isolated metrics

Analytical implications

Utility-based interaction dynamics

The framework's utility-based perspective reveals the dynamics between AI model predictions and decision utilities, interpreting model outputs as facilitators of balanced outcomes in materials science workflows. Utility functions, by assigning values to prediction scenarios, reveal how model reliability interacts with stakeholder priorities, such as in the design of energy storage materials, where precision in capacity forecasts influences cost-benefit trade-offs [1]. This analytical implication highlights feedback structures in which decision outcomes loop back to refine model evaluations, promoting adaptive systems that enhance innovation in complex parameter spaces [2]. Ethical reasoning integrates naturally, as these dynamics prompt consideration of how utility assignments reflect societal goals, such as sustainability, and of ways to avoid misalignments that could hinder equitable materials advancement [3].

Risk trade-offs in materials decision systems

Risk trade-offs constitute a core analytical implication, with the framework interpreting how differential error costs shape decision logics under uncertainty [4]. In applications such as structural materials for aerospace, risk-averse utilities prioritize minimizing catastrophic failure risk, even at the expense of computational overhead, thereby revealing systems-level insights into how model uncertainties propagate through the design and testing phases [5]. This approach elucidates steering logics that balance aggressive exploration of novel materials with conservative validation, fostering resilience in decision ecosystems [6]. Epistemic reasoning further enriches this by analyzing how incomplete knowledge of material behaviors affects risk assessments, and it urges interpretive adjustments that align AI with real-world variabilities [7].

Epistemic reasoning in AI evaluation

Epistemic reasoning emerges as a pivotal analytical lens for interpreting model evaluations through the prism of knowledge limitations and data heterogeneity [8]. The framework exposes interaction dynamics where epistemic gaps in training data lead to skewed utility interpretations, as seen in predictive modeling of composite materials, where scale-dependent properties challenge model generalizability [9]. Systems-level insights include the potential for feedback structures to mitigate these gaps, through iterative utility recalibrations that enhance epistemic confidence [10]. Ethical dimensions are foregrounded, as this reasoning underscores the need for transparent interpretations to prevent bias perpetuation and guide more inclusive evaluation practices [11].

These analytical implications collectively advance a holistic view, where the framework's interpretive power transforms AI evaluation into a tool for uncovering hidden dynamics and trade-offs, ultimately supporting more robust decision-making structures in materials science [12].

Results and Discussion

The proposed framework, by embedding decision theory into materials AI evaluation, opens avenues for interpretive integration while surfacing conceptual challenges. A primary discussion revolves around the framework's ability to reveal utility dynamics, interpreting AI performance as embedded within decision-feedback structures that evolve with contextual demands [13]. For example, in sustainable materials innovation, this allows for analytical explorations of how environmental utilities interact with predictive uncertainties, though it demands vigilance against subjective bias in utility definitions [14]. Systems-level insights suggest that such integrations can amplify AI's role in multi-objective optimization. Yet, limitations arise from the epistemic complexities of materials datasets, where noise and sparsity may distort trade-off interpretations [15].

Another facet concerns risk and ethical reasoning, where the framework steers logics toward balanced assessments, highlighting interaction dynamics in high-impact applications like biomedical materials [16]. However, conceptual constraints include the potential for over-simplification of real-world uncertainties, necessitating further interpretive refinements to capture nuanced feedback from experimental validations [17]. Ethical considerations are paramount, as

the framework prompts discussion of how utility-based evaluations might inadvertently favor certain stakeholders, advocating epistemic inclusivity to ensure equitable outcomes [18].

Future conceptual developments could extend the framework to hybrid human-AI decision systems, analyzing how shared utilities foster collaborative dynamics [8]. This could yield insights into steering logics for emerging fields like quantum materials, where epistemic challenges are pronounced [19]. Overall, the discussion affirms the framework's value in bridging theoretical gaps, while calling for ongoing interpretive evolution to address its limitations and ensure alignment with the interpretive demands of materials science [20].

Conclusion

This conceptual paper has developed a decision-theoretic framework that reinterprets the evaluation of materials AI models beyond accuracy, focusing on utility dynamics, risk trade-offs, and epistemic reasoning to provide systems-level insights. By emphasizing interaction dynamics and feedback structures, the framework steers evaluation practices toward integrative interpretations that better align AI with the multifaceted decision contexts of materials science. Ethical and analytic implications underscore its potential to enhance trustworthiness and innovation, offering a novel lens for navigating uncertainties. Ultimately, this approach contributes to a more nuanced understanding of AI's role, fostering decision-making structures that advance sustainable and equitable materials advancements.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 26 May 2024 Revised: 08 Jul 2024 Accepted: 03 Aug 2024
Published online: 18 January 2025

References

Merchant A, Batzner S, Schoenholz SS, Aykol M, Cheon G, Cubuk ED. Scaling deep learning for materials discovery. Nature. 2023;624(7990):80-5.

Wang H, Fu T, Du Y, Gao W, Huang K, Liu Z, et al. Scientific discovery in the age of artificial intelligence. Nature. 2023;620(7972):47-60.

Ramprasad R, Batra R, Pilania G, Mannodi-Kanakkithodi A, Kim C. Machine learning in materials informatics: Recent applications and prospects. npj Comput Mater. 2017;3(1):54.

Fung V, Hu G, Ganesh P, Sumpter BG. Machine learned features from density of states for accurate adsorption energy prediction. Nat Commun. 2021;12(1):88.

Schleder GR, Focassio B, Fazzio A. Machine learning for materials discovery: Two-dimensional topological insulators. Appl Phys Rev. 2021;8(3):031409.

Zhou Q, Chen X, Wang J. Machine learning assisted material discovery: A small data approach. Acc Mater Res. 2024;5(5):571-84.

Kim KS. Machine learning for accelerating energy materials discovery: Bridging quantum accuracy with computational efficiency. Adv Energy Mater. 2024;14(40):2403356.

Mohammadiun S, Hu G, Gharahbagh AA, Li J, Hewage K, Sadiq R. Evaluation of machine learning techniques to select marine oil spill response methods under small-sized dataset conditions. J Hazard Mater. 2022;436:129282.

Harimi A, Majd Y, Gharahbagh AA, Hajihashemi V, Esmaileyan Z, Machado JJ, et al. Classification of heart sounds using chaogram transform and deep convolutional neural network transfer learning. Sensors (Basel). 2022;22(24):9569.

Kanase-Patil AB, Kaldate AP, Lokhande SD, Panchal H, Suresh M, Priya V. A review of artificial intelligence-based optimization techniques for the sizing of integrated renewable energy systems in smart cities. Environ Technol Rev. 2020;9(1):111-36.

Krishnan NA, Kodamana H, Bhattoo R. Machine learning for materials discovery: Numerical recipes and practical applications. Cham: Springer International Publishing; 2024.

Guo Z, Wu Y, Hartline JD, Hullman J. A decision theoretic framework for measuring ai reliance. In: Proceedings of the 2024 acm conference on fairness, accountability, and transparency. 2024. p. 221-36.

BaniHani I, Alawadi S, Elmrayyan N. Ai and the decision-making process: A literature review in healthcare, financial, and technology sectors. J Decis Syst. 2024;33(sup1):389-99.

Ben-Michael E, Greiner DJ, Huang M, Imai K, Jiang Z, Shin S. Does ai help humans make better decisions? A methodological framework for experimental evaluation. arXiv. 2024;arXiv:2403.12108.

Peterson JC, Bourgin DD, Agrawal M, Reichman D, Griffiths TL. Using large-scale experiments and machine learning to discover theories of human decision-making. Science. 2021;372(6547):1209-14.

Tolmeijer S, Christen M, Kandul S, Kneer M, Bernstein A. Capable but amoral? Comparing ai and human expert collaboration in ethical decision making. In: Proceedings of the 2022 chi conference on human factors in computing systems. 2022. p. 1-17.

Mohammadiun S, Hu G, Gharahbagh AA, Li J, Hewage K, Sadiq R. Intelligent computational techniques in marine oil spill management: A critical review. J Hazard Mater. 2021;419:126425.

von Lilienfeld OA, Müller KR, Tkatchenko A. Exploring chemical compound space with quantum-based machine learning. Nat Rev Chem. 2020;4(7):347-58.

Schuurman Y, Goulart de Araujo L, Vilcocq L, Fongarland P. Recent developments in the use of machine learning in catalysis kinetics. Catal Today. 2021;369:3-12.

Noack MM, Doerk GS, Li R, Streit JK, Vaia RA, Yager KG, et al. Autonomous materials discovery driven by gaussian process regression with inhomogeneous measurement noise and anisotropic kernels. Sci Rep. 2020;10(1):17663.

Dobrzański LA, Honysz R. Artificial intelligence and virtual environment application for materials design methodology. Arch Mater Sci Eng. 2010;45(2):69-94.

Xu P, Ji X, Li M, Lu W. Small data machine learning in materials science. npj Comput Mater. 2023;9(1):42.

Li S, You F. Genai for scientific discovery in electrochemical energy storage: State-of-the-art and perspectives from nano- and micro-scale. Small. 2024;20(50):2406153.

Li DZ, Chen L, Liu G, Yuan ZY, Li BF, Zhang X, et al. Porous metal–organic frameworks for methane storage and capture: Status and challenges. New Carbon Mater. 2021;36(3):468-96.

Cai J, Chu X, Xu K, Li H, Wei J. Machine learning-driven new material discovery. Nanoscale Adv. 2020;2(8):3115-30.

Dou B, Zhu Z, Merkurjev E, Ke L, Chen L, Jiang J, et al. Machine learning methods for small data challenges in molecular science. Chem Rev. 2023;123(13):8736-80.

Huang JS, Liew KM, Ademiloye A. Artificial intelligence in materials modeling and design. Arch Comput Methods Eng. 2021;28(5):3399-413.

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.

Guo K, Yang Z, Yu CH, Buehler MJ. Artificial intelligence and machine learning in design of mechanical materials. Mater Horiz. 2021;8(4):1153-72.

Lu B, Xia Y, Ren Y, Xie M, Zhou L, Vinai G, et al. When machine learning meets 2d materials: A review. Adv Sci. 2024;11(13):2305277.

Reiser P, Neubert M, Eberhard A, Torresi L, Zhou C, Shao C, et al. Graph neural networks for materials science and chemistry. Commun Mater. 2022;3(1):93.

Hao WJ, Tasir Z. Development of a theoretical framework of moocs with gamification elements to enhance students' higher-order thinking skills: A critical review of the literature. J Inf Technol Educ Res. 2024;23:1-25.

Benotsmane R, Dudás L, Kovács G. Survey on artificial intelligence algorithms used in industrial robotics. Multidiszciplináris Tudományok. 2020;10(4):194-205.