ORIGINAL RESEARCH **Open access**

A Conceptual Framework for Trustworthy AI Decisions Across the Materials Design Lifecycle

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

Abstract

Artificial intelligence (AI) is increasingly embedded across the materials design lifecycle. Yet, prevailing approaches to trustworthiness remain largely model-centric, emphasizing predictive accuracy while under-specifying how AI outputs translate into high-stakes material decisions. This limitation is particularly consequential in materials science, where decisions frequently commit resources to irreversible synthesis, deployment, and long-term societal or environmental impact. Here, we propose a novel decision-centric conceptual framework for trustworthy AI in materials design, defining trustworthiness as the justification of action recommendations under uncertainty—including decisions to select, reject, prioritize, stop, or redesign candidate materials—rather than as an intrinsic property of models alone. The framework structures the materials lifecycle as an iterative sequence of seven decision-bearing stages—from problem framing to revision—and introduces five validity gates—scope, domain, uncertainty, consequence, and sustainability—that serve as systematic filters between AI outputs and actionable commitments. Trust dimensions such as reliability, robustness, transparency, accountability, safety, and sustainability are conceptualized as emergent properties of gated lifecycle interactions rather than isolated criteria. By identifying where failures originate across the lifecycle and formalizing named failure modes with corresponding containment principles, the framework explicitly links uncertainty quantification, interpretability, and governance considerations to defensible decision-making in materials contexts. This work provides a unifying theoretical structure for understanding how trustworthy AI decisions can be operationalized in materials design, offering conceptual grounding for future methodological, institutional, and governance advances in applied artificial intelligence for materials science.

Keywords Failure modes, Validity gates, Trustworthy AI, Materials design lifecycle, Decision-centric AI, Uncertainty

*Correspondence:
Wei Chen
wei.chen@outlook.com
[1] Department of Computational Materials Science, School of Materials Engineering, Tsinghua University, Beijing, China

Introduction

In the rapidly evolving field of materials science, artificial intelligence (AI) has emerged as a transformative force, enabling unprecedented explorations of vast chemical spaces and complex property landscapes [1, 2]. However, the stakes in materials design are distinctly high, often involving critical applications in energy storage, aerospace, biomedical implants, and environmental remediation, where erroneous choices can lead to catastrophic failures, economic losses, or irreversible ecological harm [3-6]. Unlike domains such as image recognition or natural language processing, where errors might be tolerable or reversible, materials decisions frequently commit resources to physical synthesis, testing, and deployment, amplifying the consequences of uncertainty [7]. This paper defines a “trustworthy decision” as a justified action recommendation under uncertainty, encompassing options to select promising candidates, reject infeasible ones, prioritize experiments, stop unproductive paths, or redesign formulations. Such decisions must balance probabilistic

insights from AI with domain-specific risks, ensuring they are not merely accurate but also reliable, robust, transparent, accountable, safe, and relevant to sustainability [3, 8].

The rationale for a decision-centric focus stems from the recognition that trustworthiness in AI for materials cannot be confined to algorithmic performance alone. Traditional approaches often emphasize model accuracy or predictive power, overlooking how outputs translate into real-world actions [4, 9]. In materials design, uncertainties arise from incomplete data, multiscale phenomena, and extrinsic factors like manufacturing variability, making isolated model trust insufficient [10]. For instance, an AI suggestion to pursue a novel alloy might appear promising based on simulated properties, but without considering lifecycle implications—such as degradation over time or supply chain ethics—the resulting decision could prove untrustworthy [11]. Materials stakes differ profoundly due to their tangible, long-term impacts: a flawed battery material might endanger lives in electric vehicles, while unsustainable sourcing could exacerbate climate change [12]. This contrasts with lower-stakes AI applications, where iterative corrections are feasible without physical repercussions [13].

Moreover, the materials lifecycle—spanning problem framing, representation, model output, decision, deployment, monitoring, and revision—presents unique vulnerabilities. Problem framing involves defining objectives, such as optimizing for strength versus cost, where biases in initial assumptions can propagate [14]. Representation entails encoding material data, often grappling with sparse or noisy datasets from experiments [15]. Model outputs provide predictions, but their trustworthiness depends on how they handle epistemic and aleatoric uncertainties [16]. The decision stage integrates these into actions, requiring risk-tolerance thresholds [17]. Deployment actualizes choices in prototypes or products, monitoring tracks performance in real conditions, and revision loops back for refinement [18]. Across this cycle, trust erodes if decisions lack justification, particularly in high-uncertainty regimes common to novel materials discovery [19].

This decision-centric lens addresses gaps in existing paradigms, which often treat trustworthiness as a static model property rather than a dynamic, context-dependent outcome [20]. By centering decisions, the framework proposed herein integrates trust dimensions holistically: reliability ensures consistent performance under nominal conditions; robustness guards against perturbations, such as data shifts; transparency reveals reasoning paths; accountability assigns responsibility for outcomes; safety prevents harm; and sustainability evaluates environmental footprints [21]. In materials contexts, these dimensions intersect with domain imperatives, such as regulatory compliance in the pharmaceutical industry or durability in infrastructure [22]. The need for such a framework is underscored by recent syntheses highlighting AI's potential to accelerate materials innovation while warning of unchecked risks [23, 24]. To clarify how decision types and trust dimensions manifest across the materials lifecycle, **Table 1** systematically maps lifecycle stages to action categories and the dominant trust considerations for each.

Table 1. Mapping of lifecycle stages to AI-informed decision types and dominant trust dimensions in materials design. The table emphasizes trustworthiness as an emergent property of justified actions rather than isolated model performance

Lifecycle stage	Primary AI-informed decisions	Dominant trust dimensions	Typical risk if unjustified
Problem framing	Define objectives; constrain scope; stop ill-posed problems	Transparency, accountability	Misaligned goals; hidden value bias
Representation	Select descriptors; exclude variables; redesign encoding	Reliability, robustness	Proxy misalignment; information loss
Model output	Interpret predictions; accept/reject confidence	Reliability, uncertainty awareness	Overconfidence; epistemic blindness
Decision	Select, reject, prioritize, or halt	Accountability, safety	Premature commitment;

	candidates		unsafe action
Deployment	Scale synthesis; integrate into systems	Robustness, safety	Failure under real conditions
Monitoring	Detect drift; trigger alarms	Reliability, transparency	Undetected degradation
Revision	Redesign objectives or models	Accountability, sustainability	Error lock-in; repeated failure

Ultimately, this paper synthesizes theoretical insights to advance a novel conceptual structure, emphasizing that trustworthy AI decisions in materials design demand explicit consideration of uncertainty's role in action justification. This approach not only mitigates failures but also enhances the ethical integrity of AI-assisted science, paving the way for responsible innovation [25].

Theoretical Background and Literature Synthesis

Trustworthiness as a decision property (not a model property)

Trustworthiness in AI systems has traditionally been framed in terms of model attributes, such as accuracy and generalization. Still, in materials design, it is more aptly conceptualized as a property of the decisions derived from those models [1, 26]. This shift recognizes that AI outputs are intermediaries, not endpoints; their value lies in informing actions like material selection or redesign under uncertainty [2]. For example, a model's high predictive fidelity might falter when applied to decisions involving rare events or extrapolated domains, which are common in materials exploration [5]. The literature emphasizes that decision trustworthiness emerges from the interplay of epistemic justification—grounding recommendations in evidence—and pragmatic utility, ensuring that actions align with stakeholder goals [8, 9]. Unlike model-centric views, which isolate algorithmic traits, a decision-property perspective incorporates contextual factors, such as lifecycle stage, to evaluate whether a recommendation is defensible [13, 27]. This aligns with broader AI ethics discussions, where trustworthiness is tied to outcome accountability rather than isolated performance [20, 28].

Uncertainty, risk, and action thresholds in materials contexts

Uncertainty permeates materials AI, stemming from data scarcity, multiscale complexities, and extrinsic variables like processing conditions [10, 16]. Recent conceptual work distinguishes aleatoric (inherent randomness) from epistemic (knowledge gaps) uncertainty, arguing that trustworthy decisions require calibrated risk assessments to set action thresholds—points at which uncertainty determines whether a recommendation is actionable [21, 29]. In materials contexts, risks are amplified by physical consequences; for instance, uncertain predictions in alloy stability could lead to structural failures [7, 30]. Literature synthesizes how Bayesian frameworks conceptually guide threshold setting, balancing false positives (pursuing unviable materials) against false negatives (overlooking breakthroughs) [4, 31]. Action thresholds thus serve as conceptual guards, ensuring that decisions, such as prioritization, reflect not just expected values but also variance and tail risks [17, 32]. This is particularly salient in sustainability-driven materials, where long-term environmental risks demand conservative thresholds [12, 33].

Interpretability, accountability, and governance concepts in materials AI

Interpretability in materials AI involves elucidating how inputs, such as atomic structures, map to outputs, such as properties, thereby fostering trust by revealing causal pathways [3, 34]. Accountability extends this by assigning responsibility for decision outcomes, integrating human oversight with AI suggestions [11, 35]. Governance concepts, drawn from recent syntheses, advocate for structured protocols across the lifecycle, such as audit trails for transparency and ethical checklists for safety [18]. In materials AI, these are tailored to domain-specific needs: interpretability aids understanding of phase transitions, while accountability ensures traceability in supply chains [14]. The literature highlights hybrid governance models, in which AI decisions are framed within regulatory frameworks to enhance robustness and

sustainability [22]. This conceptual integration underscores that governance is not additive but foundational to trustworthy decisions [25].

Where failures originate across the lifecycle (taxonomy)

Failures in AI-assisted materials design originate at distinct lifecycle stages, forming a taxonomy that classifies them by source and propagation [6]. In problem framing, failures arise from misaligned objectives, such as overlooking sustainability, leading to biased explorations [15]. Representation failures involve inadequate encodings, such as the omission of defects, which can lead to downstream inaccuracies [19]. At model output, epistemic gaps manifest as overconfident predictions, eroding decision reliability [23]. Decision-stage failures include threshold miscalibration, resulting in premature actions [24]. Deployment exposes robustness issues, such as domain shifts in real-world testing [27]. Monitoring failures stem from undetected drifts, while revision loops can perpetuate errors if not governed [28]. This taxonomy, synthesized from recent work, emphasizes cascading effects in which early uncertainties amplify later risks, necessitating conceptual containment strategies [31]. **Table 2** summarizes the named failure modes, their lifecycle origins, associated validity gates, and conceptual containment principles.

Table 2. Taxonomy of decision-level failure modes in AI-assisted materials design and their containment via validity gates. The framework emphasizes prevention and revision over post-hoc correction

Failure mode	Lifecycle origin	Gate most involved	Typical consequence	Containment principle
Scope misalignment	Problem framing	Scope gate	Irrelevant or biased decisions	Modular scoping and objective redefinition
Domain drift	Representation/deployment	Domain gate	Invalid extrapolation	Continuous domain checks and OOD detection
Uncertainty escalation	Model output/decision	Uncertainty gate	Overconfident action	Probabilistic thresholds and abstention
Consequence oversight	Decision/deployment	Consequence and sustainability gates	Safety or environmental harm	Multi-stakeholder impact assessment

Proposed conceptual framework

The proposed framework conceptualizes trustworthy AI decisions as justified actions embedded within a cyclical materials design lifecycle, comprising seven interconnected stages: problem framing, representation, model output, decision, deployment, monitoring, and revision. This structure extends beyond linear pipelines by incorporating iterative feedback loops, ensuring decisions evolve with new insights. At each stage, decision points are identified as junctures where AI informs actions—select, reject, prioritize, stop, or redesign—under uncertainty. For instance, in problem framing, decisions involve scoping objectives; in representation, selecting data encodings; in model output, interpreting predictions; in decision, committing to paths; in deployment, scaling prototypes; in monitoring, assessing performance; and in revision, updating assumptions.

Central to the framework are five validity gates, serving as conceptual checkpoints to filter untrustworthy decisions: (1) Scope gate evaluates if the problem aligns with AI's conceptual strengths, rejecting overly vague or non-quantifiable framings; (2) Domain gate assesses data and model applicability, flagging extrapolations beyond trained regimes; (3) Uncertainty gate quantifies risk thresholds, halting actions where variance exceeds predefined limits; (4) Consequence gate weighs potential impacts, prioritizing safety in high-stakes scenarios; (5) Sustainability gate integrates environmental

and ethical considerations, vetoing decisions with undue ecological costs [2, 9, 16, 29, 34]. These gates operate sequentially yet interdependently, with failures at one prompting revisions upstream. **Figure 1** illustrates how validity gates operate as sequential decision filters, transforming raw AI outputs into justified or rejected actions in the face of uncertainty.

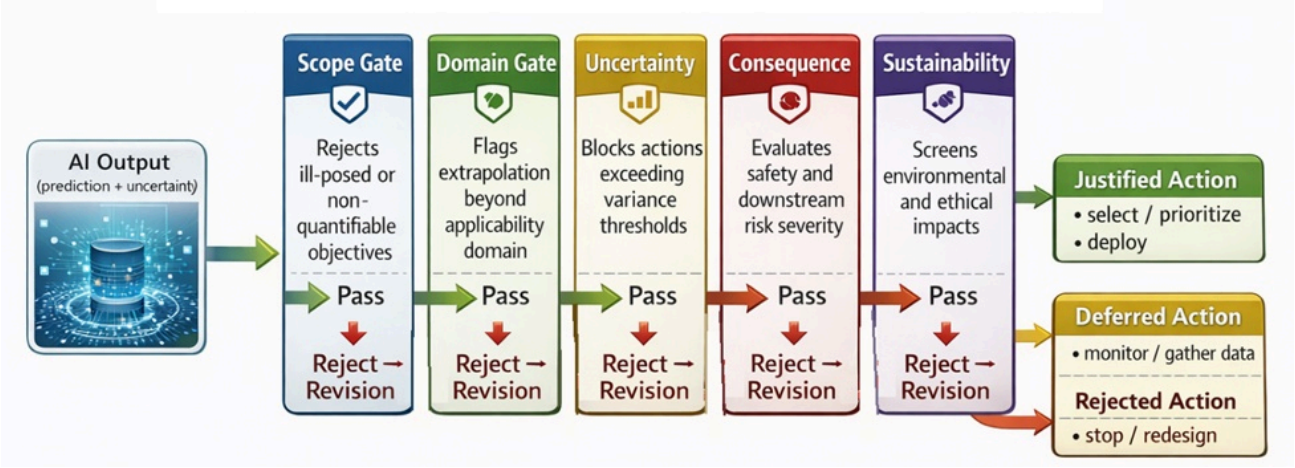

Figure 1. Conceptual depiction of validity gates as decision filters that convert AI outputs into justified actions. Gates operate sequentially, with failures triggering revision rather than execution, reinforcing decision-centric trustworthiness

The framework also delineates four named failure modes, each with associated containment principles: (1) Scope misalignment failure, where overambitious framing leads to irrelevant decisions, contained by modular scoping protocols; (2) Domain drift failure, arising from data shifts, mitigated through adaptive domain checks; (3) Uncertainty escalation failure, when ignored variances cascade, addressed via probabilistic thresholding; (4) Consequence oversight failure, overlooking downstream harms, contained by multi-stakeholder impact assessments [6, 10, 21]. These modes emphasize prevention over correction, aligning with trust dimensions by embedding reliability in gates, robustness in drifts, transparency in checks, accountability in assessments, safety in consequences, and sustainability throughout. As shown in **Figure 2**, the framework presents a gated, iterative lifecycle integrating trust dimensions across seven decision stages.

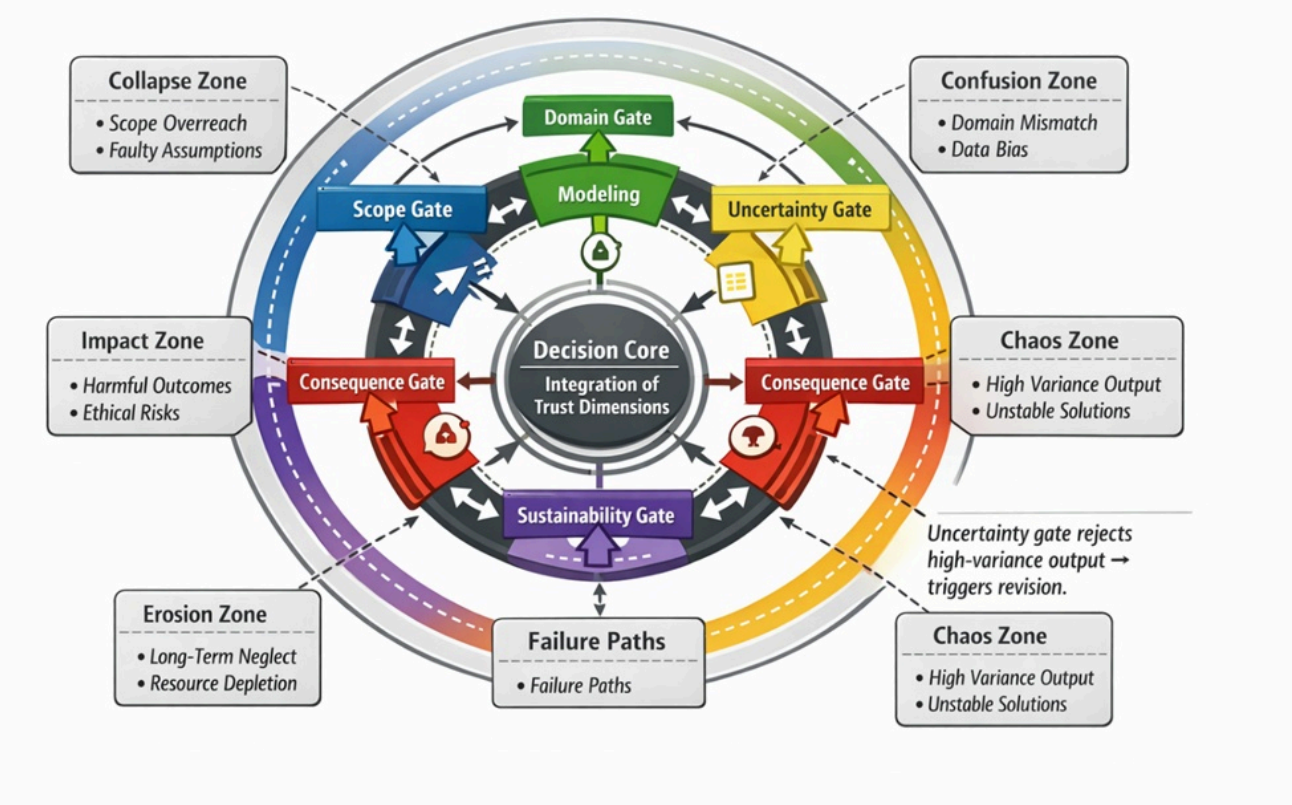

Figure 2. Circular framework illustrating seven iterative lifecycle stages linked by validity gates and a central decision core that integrates trust dimensions. Dashed lines and containment zones highlight failure modes and revision pathways

Propositions

This section advances three formal propositions derived from the proposed framework, linking its key elements—lifecycle stages, validity gates, failure modes, and containment principles—to anticipated outcomes in trustworthy AI decisions for materials design. These propositions serve as theoretical assertions, grounded in synthesized literature, to guide conceptual understanding without empirical validation [1, 3, 7].

Proposition 1: Integrating validity gates at decision points across the materials lifecycle enhances the justification of action recommendations under uncertainty, thereby increasing overall decision trustworthiness. Specifically, by filtering recommendations through scope, domain, uncertainty, consequence, and sustainability gates, the framework ensures that actions, such as selection or redesign, align with trust dimensions, including reliability and robustness. For instance, the uncertainty gate calibrates action thresholds to mitigate epistemic gaps, thereby preventing overconfident decisions in sparse-data regimes typical of novel materials [4, 10, 13]. This proposition posits that gated processes reduce the propagation of early-stage biases, fostering defensible outcomes in high-stakes contexts such as energy materials, where sustainability is paramount [17, 18, 21].

Proposition 2: Identification and containment of named failure modes within the lifecycle promote resilience in AI-assisted materials decisions, transforming potential trust erosions into opportunities for iterative revision. Scope misalignment failure, for example, can be contained through modular protocols that revisit problem framing, ensuring accountability and transparency [2, 11, 19]. Similarly, uncertainty escalation failure is addressed via probabilistic thresholding, which aligns with robustness by bounding risks in multiscale predictions [26, 27, 29]. This proposition asserts that proactive containment principles not only avert catastrophic outcomes, such as unsafe structural materials, but also amplify safety and sustainability by embedding feedback loops that refine decisions over time [20, 22, 30].

Proposition 3: A decision-centric focus, rather than a model-centric one, elevates the role of trust dimensions in materials AI, leading to more accountable and ethical innovation ecosystems. By conceptualizing trustworthiness as emergent from lifecycle interactions rather than as isolated attributes, the framework ensures that transparency and accountability govern actions such as prioritization and stopping [3, 6, 33]. For materials with long-term impacts, such as biodegradable polymers, this shift prioritizes consequence and sustainability gates to balance innovation with ethical imperatives [23, 31, 34]. The proposition maintains that this holistic approach cultivates governance structures that adapt to evolving uncertainties, ultimately yielding decisions that are not only justified but also socially responsible [8, 12, 35].

These propositions collectively underscore the framework's novelty in reorienting trustworthiness toward actionable outcomes, providing a theoretical basis for conceptual advancements in applied AI [5, 9, 14].

Results and Discussion

The proposed conceptual framework offers significant implications for the conceptualization of trustworthy AI in materials design, shifting the paradigm from algorithmic isolation to integrated decision processes. By emphasizing decision points across the lifecycle, it addresses the unique stakes of materials science, where uncertainties can lead to irreversible physical commitments [1, 15]. This decision-centric approach implies a reevaluation of how trust dimensions—reliability, robustness, transparency, accountability, safety, and sustainability—are operationalized, treating them as interdependent properties that emerge from gated interactions rather than static traits [7, 16]. For instance, in problem framing, the scope gate requires explicit boundary definitions to avoid misaligned objectives, which could otherwise compromise sustainability in resource-intensive designs [17, 18]. Such implications extend to governance, suggesting that AI tools in materials should incorporate built-in checkpoints to foster ethical oversight, aligning with broader calls for responsible innovation [2, 19, 33].

Moreover, the framework's validity gates provide a conceptual tool for mitigating risks inherent to materials contexts, such as data sparsity or extrinsic variabilities [4, 26]. The domain gate, for example, implies vigilance against extrapolation failures, which are crucial in diverse applications such as composites and nanomaterials, where domain shifts can undermine robustness [24, 29]. Similarly, the consequence gate encourages consideration of downstream impacts, thereby enhancing safety in critical sectors such as biomedicine [10, 28]. By incorporating failure modes and containment principles, the framework implies a proactive stance against cascading errors, such as uncertainty escalation leading to flawed deployments [27, 30]. This has broader implications for interdisciplinary collaboration, as it conceptualizes human-AI hybrid systems where accountability is shared, potentially reducing overreliance on opaque models [6, 32, 34].

However, the framework has conceptual limitations that warrant acknowledgment. Primarily, its reliance on qualitative gates may overlook nuanced interdependencies among trust dimensions, such as trade-offs between transparency and robustness in complex lifecycle revisions [3, 8]. For example, stringent sustainability gates might constrain exploratory decisions in early framing, potentially stifling innovation in uncertain regimes [22, 23]. Additionally, the taxonomy of failure modes, while novel, is not exhaustive, as emerging AI paradigms, such as generative models, could introduce unforeseen vulnerabilities not captured here [5, 9, 25]. The framework assumes a cyclical lifecycle, but in practice, non-linear dynamics or external disruptions—like regulatory changes—might challenge its applicability [11, 13, 20]. Furthermore, it does not fully address cultural or institutional variations in trust perceptions, limiting its universality across global materials research communities [12, 35].

Looking ahead, future conceptual directions could extend this framework by integrating additional dimensions, such as equity in AI decisions for inclusive materials innovation [14, 21]. One avenue is to develop nested sub-frameworks for specific lifecycle stages, such as enhanced uncertainty handling in monitoring [27, 31]. Another is exploring synergies with related fields, such as chemistry, to broaden the scope of trustworthiness concepts [1, 7, 16]. Conceptually, incorporating adaptive gates that evolve with AI advancements could enhance resilience [4, 18, 26]. Ultimately, these directions invite theoretical refinements to sustain the framework's relevance in an evolving AI landscape [15, 19, 33].

Conclusion

This work has advanced a decision-centric theoretical framework for trustworthy artificial intelligence in materials design, reframing trustworthiness as a property of justified actions under uncertainty rather than a static attribute of predictive models. By explicitly centering decisions—such as selection, rejection, prioritization, stopping, and redesign—within a seven-stage, iterative materials lifecycle, the framework addresses the uniquely high stakes of materials science, where AI-guided choices often lead to irreversible physical, environmental, and societal consequences.

The proposed structure integrates five validity gates—scope, domain, uncertainty, consequence, and sustainability—as conceptual mechanisms that regulate when and how AI outputs may legitimately inform action. Through these gates, core trust dimensions emerge dynamically from lifecycle interactions, embedding reliability, robustness, transparency, accountability, safety, and sustainability directly into decision pathways. The formalization of lifecycle-specific failure modes and associated containment principles further strengthens the framework by shifting attention from post hoc correction toward proactive prevention and revision, a critical requirement for responsible materials innovation.

The three propositions articulated herein establish a theoretical basis for understanding how gated decision processes can enhance justification, resilience, and ethical accountability in AI-assisted materials design. Collectively, they position trustworthiness not as an auxiliary constraint on AI systems, but as an organizing principle for how AI participates in scientific and engineering judgment. While the framework is intentionally conceptual and qualitative, it offers a rigorous scaffold for future work, including the development of quantitative gate criteria, adaptive governance mechanisms, and domain-specific instantiations across materials classes.

As global demands intensify for safe, sustainable, and rapidly deployable materials solutions, the need for AI systems that support not only accurate predictions but also defensible decisions becomes increasingly urgent. By aligning AI outputs with lifecycle-aware justification, uncertainty management, and ethical responsibility, this framework provides a durable theoretical foundation for cultivating trust in AI-enabled materials science and for guiding the next generation of applied AI methodologies toward responsible and societally aligned innovation.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 21 Jul 2023 Revised: 01 Oct 2023 Accepted: 29 Oct 2023
Published online: 18 January 2024

References

Feng J, Lansford JL, Katsoulakis MA, Vlachos DG. Explainable and trustworthy artificial intelligence for correctable modeling in chemical sciences. Sci Adv. 2020;6(47):eabc3204.

Ravi N, Chaturvedi P, Huerta EA, Liu Z, Chard R, et al. Fair principles for ai models with a practical application for accelerated high energy diffraction microscopy. Sci Data. 2022;9(1):571.

Simion M, Kelp C. Trustworthy artificial intelligence. Asian J Philos. 2023;2(1):1-12.

Olfatbakhsh T, Milani AS. A highly interpretable materials informatics approach for predicting microstructure-property relationship in fabric composites. Compos Sci Technol. 2022;219:109264.

Huang H, Magar R, Xu C, Farimani AB. Materials informatics transformer: A language model for interpretable materials properties prediction. arXiv. 2023;arXiv:2308.16259.

Li C, Zheng K. Methods, progresses, and opportunities of materials informatics. InfoMat. 2023;5(3):e12433.

Oviedo F, Ferres JL, Buonassisi T. Interpretable and explainable machine learning for materials science and chemistry. Acc Mater Res. 2022;3(6):597-607.

Dean J, Scheffler M, Purcell TAR, Barabash SV. Interpretable machine learning for materials design. J Mater Res. 2023;38(9):1735-48.

Korolev V, Protsenko P. Accurate, interpretable predictions of materials properties within transformer language models. Patterns. 2023;4(9):100807.

Badini S, Regondi S, Pugliese R. Unleashing the power of artificial intelligence in materials design. Materials. 2023;16(17):5927.

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.

Zhu F, Wu X, Zhou M, Sabri MMS, Huang J. Intelligent design of building materials: Development of an ai-based method for cement-slag concrete design. Materials. 2022;15(11):3833.

Li J, Lim K, Yang H, Ren Z, Raghavan S, Chen PY. Ai applications through the whole life cycle of material discovery. Matter. 2020;3(2):393-432.

Kalidindi SR. Feature engineering of material structure for ai-based materials knowledge systems. J Appl Phys. 2020;128(4):041103.

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(3):033001.

Epps RW, Volk AA, Reyes KG, Abolhasani M. Accelerated ai development for autonomous materials synthesis in flow. Chem Sci. 2021;12(18):6025-37.

Hardian R, Liang Z, Zhang X, Szekely G. Artificial intelligence: The silver bullet for sustainable materials development. Green Chem. 2020;22(22):7521-8.

Gomes CP, Fink D, Van Dover RB. Computational sustainability meets materials science. Nat Rev Mater. 2021;6(8):645-59.

Toniato A, Schilter O, Laino T. The role of ai in driving the sustainability of the chemical industry. Chimia. 2023;77(4):213-8.

Abolhasani M, Brown KA. Role of ai in experimental materials science. MRS Bull. 2023;48(5):413-5.

Sha W, Guo Y, Yuan Q, Tang S, Zhang X. Artificial intelligence to power the future of materials science and engineering. Adv Intell Syst. 2020;2(2):1900143.

Liu Y, Yang Z, Yu Z, Liu Z, Liu D, Lin H, et al. Generative artificial intelligence and its applications in materials science: Current situation and future perspectives. J Materiomics. 2023;9(4):687-704.

Goswami L, Deka MK, Roy M. Artificial intelligence in material engineering: A review on applications of artificial intelligence in material engineering. Adv Eng Mater. 2023;25(12):2300104.

Pitz E, Pochiraju K. Ai/ml for quantification and calibration of property uncertainty in composites. In: Machine learning applied to composite materials. Springer; 2022.

Tavazza F, Choudhary K, DeCost B. Approaches for uncertainty quantification of ai-predicted material properties: A comparison. arXiv. 2023;arXiv:2310.13136.

Honarmandi P, Arróyave R. Uncertainty quantification and propagation in computational materials science and simulation-assisted materials design. Integr Mater Manuf Innov. 2020;9(3):286-300.

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

Fernandez J, Chiachio M, Chiachio J, Munoz R. Uncertainty quantification in neural networks by approximate bayesian computation: Application to fatigue in composite materials. Eng Appl Artif Intell. 2022;108:104615.

Acar P, Tran A, Nikbay M, Mahadevan S. Uncertainty quantification in characterization, modelling, and design of materials. Front Mater. 2022;9:799081.

Bharadwaja B, Nabian MA, Sharma B. Physics-informed machine learning and uncertainty quantification for mechanics of heterogeneous materials. Integr Mater Manuf Innov. 2022;11(4):537-50.

Zhang J, Kailkhura B, Han TYJ. Leveraging uncertainty from deep learning for trustworthy material discovery workflows. ACS Omega. 2021;6(15):10345-56.

Seidel R, Schmidt K, Thielen N, Franke J. Trustworthiness of machine learning models in manufacturing applications using the example of electronics manufacturing processes. Procedia CIRP. 2022;107:1047-52.

Thuraisingham B. Trustworthy machine learning. IEEE Intell Syst. 2022;37(1):108-11.

Zhong X, Gallagher B, Liu S, Kailkhura B. Explainable machine learning in materials science. npj Comput Mater. 2022;8(1):1-12.

Heuer H, Breiter A. More than accuracy: Towards trustworthy machine learning interfaces for object recognition. In: Proceedings of the 28th acm conference on user modeling, adaptation and personalization. 2020.