ORIGINAL RESEARCH Open access

From Correlations to Design Rules: A Conceptual Model of Knowledge Extraction in Materials AI

Claire Dupont[1*], Julien Martin[1]

Abstract

The integration of artificial intelligence (AI) into materials science has substantially accelerated property prediction and materials screening. Yet, the predominance of data-driven correlations has exposed a persistent epistemic gap between predictive success and the derivation of interpretable, generalizable design rules. This conceptual manuscript develops a theoretical framework for knowledge extraction in materials AI that explicitly addresses this gap by reframing the transition from correlations to design rules as a staged epistemic process rather than a by-product of model performance. Drawing on literature in materials informatics, data bias, and philosophy of science, the framework organizes knowledge extraction into four interconnected stages—Correlation Mapping, Bias Interrogation, Value Integration, and Rule Synthesis—linked through continuous epistemic validation. The model foregrounds epistemic agency, requiring explicit scrutiny of assumptions, biases, and value commitments before causal inference. Six propositions articulate the conditions under which AI-derived correlations may legitimately support prescriptive design claims, emphasizing reflexive feedback and epistemic governance. By conceptualizing knowledge extraction as a norm-governed process of justification, this work provides a theoretical scaffold for transforming AI outputs into scientifically defensible design rules, contributing to a more reliable and responsible epistemology of materials discovery.

Keywords Materials informatics, Artificial intelligence, Data bias, Epistemic values, Knowledge extraction, Design rules

*Correspondence:
Claire Dupont
claire.dupont@gmail.com
[1] Department of Materials Data Analytics, Faculty of Engineering, University of Bordeaux, Bordeaux, France

Introduction

The advent of artificial intelligence (AI) in materials science represents a paradigm shift, transforming traditional experimental and computational approaches into data-driven enterprises capable of unprecedented efficiency. Over the past decade, AI techniques, particularly machine learning algorithms, have been instrumental in predicting material properties, optimizing synthesis routes, and accelerating discovery cycles that once spanned years into mere weeks or months [1, 2]. For instance, AI has facilitated the screening of millions of potential compounds for applications in batteries, catalysts, and semiconductors, leveraging large-scale databases to identify promising candidates with minimal human intervention [3, 4]. This surge in capability is underpinned by the exponential growth of materials data repositories and computational power, enabling models to discern patterns from complex, high-dimensional datasets that elude conventional analysis [5].

Yet, this progress is not without conceptual hurdles. A core limitation lies in the predominant focus on correlations derived from statistical associations in training data, which often fail to elucidate causal relationships or generalizable design rules

[6]. Design rules, defined here as principled guidelines linking material structure, composition, and processing to desired functionalities, are the bedrock of materials engineering. They enable not only prediction but also rational design, allowing scientists to engineer materials with targeted properties while minimizing trial-and-error experimentation [7]. In contrast, many AI models operate as “black boxes,” producing outputs without transparent reasoning, raising concerns about reliability, reproducibility, and the epistemic validity of derived insights [8]. This opacity exacerbates issues when models are applied to novel contexts, potentially perpetuating errors or overlooking fundamental physical constraints [9].

Compounding these challenges are pervasive data biases in materials informatics. Datasets in materials science are frequently skewed toward well-studied systems, such as common oxides or equilibrium phases, underrepresenting exotic or metastable materials [10, 11]. Such biases can stem from historical publication preferences, experimental feasibility, or institutional priorities, leading AI models to reinforce existing knowledge gaps rather than bridge them [12]. For example, biases in crystal synthesizability predictions have been shown to arise from imbalanced training data, where successful syntheses are overrepresented, distorting model outputs [6]. Moreover, integrating AI introduces epistemic and non-epistemic values into the scientific process. Epistemic values, such as predictive accuracy and simplicity, guide model selection, while non-epistemic values—like societal priorities for sustainable materials or ethical considerations in data sourcing—influence research directions and interpretations [13, 14]. The interplay of these values demands scrutiny, as unchecked influences can undermine the objectivity of scientific claims [15].

This manuscript addresses these issues through a purely conceptual lens, developing a novel theoretical framework for knowledge extraction in materials AI. Unlike empirical studies that rely on simulations or data analysis, this work synthesizes theoretical insights from materials informatics, the data bias literature, and the philosophy of science to propose a model that elevates correlations into design rules. The framework is original in its emphasis on epistemic agency, in which scientists actively interrogate AI outputs through bias-aware, value-sensitive processes, ensuring that extracted knowledge is not merely correlative but causally informed and ethically aligned [16]. It posits that knowledge extraction is an iterative epistemic endeavor, requiring transparency at each stage to mitigate risks of overreliance on AI [17].

The need for such a framework is acute in an era where AI-driven discoveries are increasingly touted as revolutionary, yet their epistemic foundations remain underexplored [18]. Materials science, as a field that intersects physics, chemistry, and engineering, is particularly susceptible to these tensions, given its reliance on multiscale phenomena in which correlations at one level may not translate into rules at another [19]. By conceptualizing knowledge extraction as a bridge between data and theory, this model aims to enhance the field’s theoretical maturity and promote sustainable practices that align with broader scientific values [20].

The manuscript proceeds as follows. The theoretical background synthesizes key literature, examining the evolution of materials informatics, data biases, and the role of values in AI-augmented science. This synthesis highlights gaps in current approaches, such as the neglect of causal inference in correlative models [9]. The proposed conceptual framework delineates a structured process for knowledge extraction, including a textual description of a conceptual figure that illustrates its components. Future sections (not included here) will articulate propositions, discuss implications, and conclude with recommendations for adoption.

In sum, this work advocates a reflective turn in materials AI, in which theoretical frameworks guide the transition from data-centric predictions to knowledge-centric design. By doing so, it contributes to a more robust, equitable, and epistemically sound discipline, ensuring AI serves as a tool for genuine scientific progress rather than a substitute for it [21].

Theoretical Background and Literature Synthesis

Evolution of materials informatics

Materials informatics has emerged as a transformative discipline at the intersection of data science and materials research, leveraging AI to navigate the vast combinatorial space of possible materials [1]. Historically, materials discovery relied on Edisonian trial-and-error methods or physics-based simulations, which were time-intensive and limited in scope [22]. The shift toward informatics began in earnest around 2010, with initiatives such as the Materials Genome Initiative, which emphasized high-throughput computational screening and data integration to accelerate development [3]. By 2020, AI had become central, with machine learning models predicting properties such as bandgaps and mechanical strength from structural descriptors [2, 23].

Recent advancements highlight AI's role in multiscale modeling, where algorithms integrate atomic-level simulations with macroscopic behaviors [4]. For example, graph neural networks have been employed to represent crystal structures, enabling predictions across diverse material classes [24]. However, this evolution reveals a tension: while AI excels at pattern recognition, it often prioritizes predictive accuracy over mechanistic understanding [19]. Literature syntheses underscore that, while powerful, informatics tools must evolve to incorporate domain knowledge to avoid superficial correlations [12]. This progression sets the stage for frameworks that extract deeper insights, as the field matures from data aggregation to knowledge synthesis [5].

Data challenges and bias in AI-Driven materials research

Data quality and bias represent foundational challenges in materials AI, where incomplete or skewed datasets can propagate errors throughout the discovery pipeline [10, 11]. Materials databases, such as those curated by national laboratories, often exhibit selection biases that favor stable, easily synthesizable compounds, leading to underrepresentation of high-entropy alloys and nanomaterials [6, 25]. Such biases manifest in model outputs, as seen in synthesizability predictions, where AI overestimates the feasibility of familiar systems [9].

The literature on bias mitigation emphasizes synthetic data generation and diversification strategies to address these issues [10, 26]. For instance, techniques such as generative adversarial networks have been proposed to augment datasets, thereby reducing epistemic gaps [27]. Yet, biases extend beyond data to algorithmic design, where choices in feature selection or loss functions can amplify disparities [12, 28]. Reviews highlight that without bias-aware validation, AI risks entrenching historical inequities in materials research, such as prioritizing industrially viable materials over those for niche applications [7, 29]. This synthesis reveals a need for conceptual models that embed bias detection as an integral step in knowledge extraction, ensuring outputs are not only accurate but epistemically reliable [8].

Epistemic and non-epistemic values in scientific AI

Values play a pivotal role in shaping AI applications in materials science, influencing everything from dataset curation to model interpretation [13, 15]. Epistemic values, such as coherence and explanatory power, guide the pursuit of truthful representations, while non-epistemic values—like sustainability or equity—direct research toward societal benefits [14, 16]. In AI contexts, these values intersect, as seen in alignment research where models are tuned to reflect human priorities, potentially introducing subjective biases [17, 18].

Philosophical analyses argue that AI opacity undermines epistemic agency, where scientists must "know when they do not know" to maintain integrity [14, 20]. Literature on epistemic risks warns that overreliance on AI could erode diverse inquiry modes, fostering monocultures that prioritize quantifiable outcomes over exploratory science [9, 15]. The values literature further posits that AI introduces new epistemic stances, requiring adaptive epistemologies that balance automation with human oversight [19, 21]. This synthesis underscores the necessity for frameworks that explicitly incorporate value reflection, ensuring AI enhances rather than supplants scientific values [13, 22].

Knowledge in science: From data to theory

The epistemic journey from data to theory in materials science involves transcending correlations to uncover causal structures [23, 24]. Traditional epistemology views knowledge as justified true belief, but AI complicates this by generating predictions without explicit justification [12]. The literature on scientific AI advocates paradigms that integrate machine outputs with theoretical validation, such as hybrid models combining neural networks with physical laws [25, 26].

Challenges arise in high-stakes contexts, where AI's lack of transparency erodes trust in the knowledge it produces [27, 28]. Syntheses emphasize that effective knowledge extraction requires mechanisms for causal reasoning, drawing on human cognition's forward-looking logic to complement AI's data-driven approach [29]. This background highlights the gap in current practices: while AI excels at data processing, theoretical frameworks are needed to elevate outputs to design rules, fostering a sustainable epistemic ecosystem [9, 20].

Proposed conceptual framework

The proposed conceptual framework formalizes knowledge extraction in materials AI as an explicitly epistemic transformation process, through which raw statistical correlations are progressively refined into robust, interpretable, and normatively defensible design rules. Rather than treating knowledge emergence as a by-product of predictive performance, the framework reconceptualizes it as a layered justificatory progression, wherein each stage incrementally strengthens the epistemic warrant of AI-generated insights. In doing so, the framework responds directly to the limitations of prevailing data-centric paradigms, which prioritize accuracy while leaving questions of bias, value influence, and causal validity largely implicit.

At its core, the framework integrates three dimensions that are typically treated in isolation within materials AI: bias mitigation, value alignment, and causal inference. Their integration is not procedural but structural: each dimension is embedded within a distinct epistemic stage, ensuring that knowledge claims are continuously interrogated rather than retrospectively justified. This orientation foregrounds epistemic agency, positioning scientists not as passive recipients of AI outputs but as active evaluators of their meaning, scope, and legitimacy. As such, the framework advances a conception of materials AI in which interpretability, ethical soundness, and theoretical grounding are constitutive features of knowledge rather than optional enhancements.

The framework is organized into four interconnected stages—Correlation Mapping, Bias Interrogation, Value Integration, and Rule Synthesis—arranged as an iterative epistemic cycle. While analytically distinct, these stages are dynamically coupled through feedback loops that enable refinement in response to emerging inconsistencies, uncertainties, or normative tensions. The progression from one stage to the next marks a shift from descriptive pattern recognition to prescriptive scientific guidance, reflecting increasing epistemic commitment.

Stage I: Correlation mapping

The first stage, Correlation Mapping, corresponds to the initial epistemic encounter between AI systems and materials data. Here, machine learning models analyze high-dimensional datasets to identify statistical associations between material descriptors—such as atomic composition, lattice topology, or processing parameters—and observed properties, including mechanical strength, conductivity, or stability. This stage draws on established materials informatics techniques to surface candidate relationships that may warrant further investigation.

Crucially, the framework rejects the tacit assumption that correlations are epistemically neutral. Instead, Correlation Mapping is explicitly constrained by assumption documentation, requiring transparency regarding data provenance, feature selection, representational choices, and modeling priors. By treating correlations as provisional epistemic objects rather than latent truths, this stage guards against overinterpretation and premature generalization [1, 2]. The output of Correlation Mapping is therefore not knowledge per se, but a structured space of proto-hypotheses whose epistemic status remains intentionally limited.

Stage II: Bias interrogation

Bias Interrogation constitutes the framework's primary epistemic filter, transforming raw correlations into more reliable candidates for knowledge extraction. At this stage, identified associations are systematically examined for distortions arising from data imbalance, historical research preferences, or algorithmic design choices. Techniques such as sensitivity analysis, uncertainty quantification, and distributional diagnostics are employed to reveal whether correlations disproportionately reflect overrepresented material classes or experimentally convenient regimes [10, 11].

Importantly, bias is treated not merely as a statistical artifact but as an epistemic risk: unexamined biases can masquerade as generalizable insights while encoding contingent research histories. By foregrounding bias interrogation, the framework strengthens the justificatory basis of subsequent reasoning, ensuring that retained correlations are epistemically resilient rather than artifactually robust [14]. This stage thus converts descriptive associations into vetted epistemic candidates, explicitly marked by their limitations and uncertainty profiles.

Stage III: Value integration

The third stage, Value Integration, introduces a deliberate moment of normative reflection into the knowledge extraction process. Here, correlations that have survived bias interrogation are evaluated against both epistemic values—such as explanatory coherence, simplicity, and theoretical plausibility—and non-epistemic values, including sustainability, safety, and societal relevance [13, 15]. Rather than treating values as external constraints, the framework acknowledges their constitutive role in shaping what counts as meaningful or actionable knowledge.

Decision matrices or structured deliberation tools are employed to make value trade-offs explicit, preventing implicit preferences from silently steering interpretation [16]. This stage is particularly critical in materials science, where AI-derived recommendations may influence long-term research trajectories or resource allocation. By embedding value integration before rule formation, the framework ensures that design rules are not only scientifically defensible but also normatively aligned, preserving the integrity of knowledge claims under diverse evaluative lenses.

Stage IV: Rule synthesis

Rule Synthesis represents the culmination of the epistemic progression, wherein value-vetted correlations are transformed into causal design rules. This transition marks a fundamental shift from associative reasoning to mechanistic inference, employing tools such as counterfactual analysis, physics-constrained modeling, or hybrid AI–theory approaches to establish causal plausibility [23, 28]. The resulting rules articulate conditional, context-sensitive guidance—e.g., "Increasing dopant concentration by *X*% enhances property *Y* under conditions *Z*"—that can inform rational materials design rather than mere prediction.

Epistemic closure at this stage is provisional rather than final. Synthesized rules remain subject to iterative validation and revision, with feedback loops enabling reassessment if empirical inconsistencies or value conflicts emerge. In this sense, Rule Synthesis does not terminate inquiry but stabilizes it temporarily, yielding actionable yet revisable knowledge. **Table 1** summarizes the epistemic role of each framework component, distinguishing descriptive associations from normatively and causally warranted design rules.

Table 1. Epistemic roles of framework stages in knowledge extraction

Framework stage	Primary epistemic function	Dominant risk addressed	Output status
Correlation mapping	Identification of statistical associations under explicit representational assumptions	Spurious pattern detection; ontological ambiguity	Provisional correlations (proto-hypotheses)
Bias interrogation	Epistemic filtering through uncertainty analysis and bias detection	Dataset skew; historical overrepresentation	Vetted correlations with documented limitations
Value integration	Normative evaluation of epistemic and non-epistemic priorities	Implicit value drift; instrumental overreach	Normatively admissible candidate relations

Rule synthesis	Causal stabilization via counterfactual and physics-constrained reasoning	Black-box inference; non-transferability	Context-dependent design rules
Epistemic feedback (cross-cutting)	Continuous justificatory reassessment across stages	Knowledge drift; ossification	Revisable, temporally robust knowledge
Epistemic governance (cross-cutting)	Traceability and accountability of assumptions and decisions	Loss of legitimacy; irreproducibility	Institutionally defensible knowledge claims

Epistemic logic and originality

The originality of the framework lies in its epistemic layering, in which each stage incrementally strengthens the justificatory strength of AI outputs. Knowledge extraction is thus conceptualized not as a linear pipeline but as a reflexive process of epistemic escalation, moving from descriptive correlation to prescriptive rule while maintaining transparency and accountability at each step. The framework assumes good-faith engagement by interdisciplinary teams and incorporates internal audit points to prevent epistemic drift or normative capture.

By positioning epistemic validation as a continuous, system-wide function rather than a terminal check, the framework distinguishes itself from reformulations of predictive workflows [3, 4]. As shown in **Figure 1**, it provides a theoretical scaffold for materials AI that foregrounds knowledge production as a norm-governed scientific practice, rather than an emergent property of optimization.

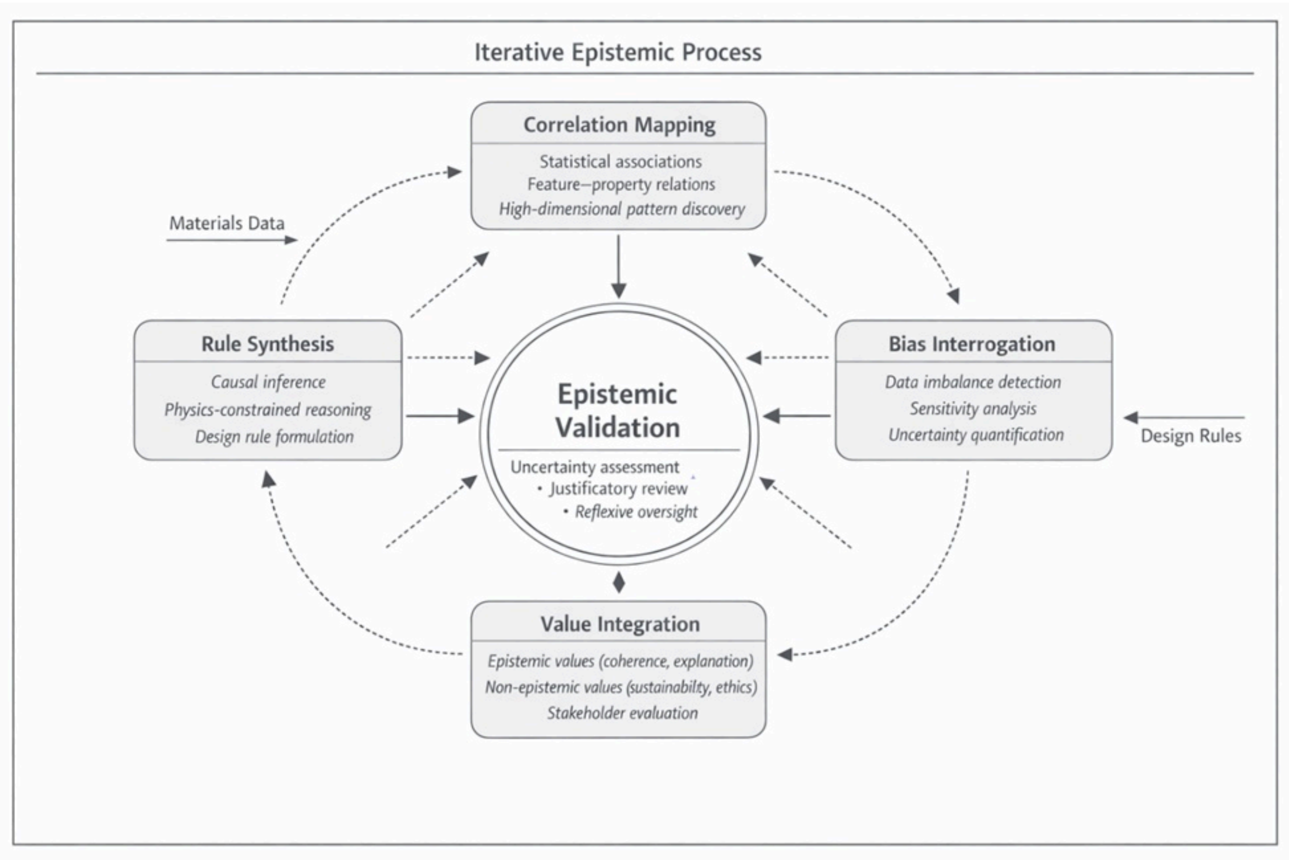

Figure 1. Conceptual framework for knowledge extraction in materials AI: from correlations to design rules

Propositions

Building on the proposed conceptual framework, this section articulates six propositions that formalize the theoretical implications of transitioning from correlations to design rules in materials AI. The propositions specify epistemic conditions governing inference, justification, and revision across the framework's stages, with particular emphasis on bias mitigation, value integration, and causal reasoning.

Proposition 1. *Correlation Mapping, when constrained by domain-specific ontological commitments, facilitates the identification of robust statistical associations but requires epistemic safeguards to prevent overgeneralization from biased datasets.*In materials informatics, initial correlations often emerge from high-dimensional data, yet without explicit ontological framing—such as distinguishing equilibrium from non-equilibrium states—these associations risk epistemic fragility [1, 5]. For instance, unfiltered correlations in crystal structure prediction can perpetuate biases toward stable oxides, undermining generalizability [6, 10]. Incorporating ontological constraints elevates correlations from mere patterns to proto-hypotheses, aligning them with epistemic values of coherence and predictive utility [13, 14] and reducing susceptibility to spurious inference [9, 19].

Proposition 2. *Bias Interrogation functions as an epistemic filter that enhances the justificatory strength of AI outputs by requiring systematic detection and correction of data skew before causal interpretation.* Biases, such as the underrepresentation of metastable phases, can distort model predictions and weaken knowledge claims [9, 11, 12]. Interrogation processes—including sensitivity analysis and dataset diversification—not only reduce distortion but also increase epistemic agency by rendering assumptions explicit [8, 15]. Evidence from alloy design studies suggests that bias-aware validation improves robustness where historical data imbalances have constrained exploration [25, 26], transforming fragile correlations into epistemically defensible inputs for rule synthesis [16, 20].

Proposition 3. *Value Integration ensures that extracted knowledge reflects a balanced interplay of epistemic and non-epistemic priorities, preventing the subordination of scientific justification to utilitarian optimization.* In AI-augmented materials research, values such as sustainability, interpretability, and explanatory adequacy shape interpretation, yet often remain implicit [13, 17]. Decision frameworks that explicitly weigh epistemic criteria against non-epistemic considerations —such as environmental impact—support normatively defensible inference, as demonstrated in value-sensitive approaches to energy materials [18, 21]. Embedding this stage renders knowledge extraction reflexive rather than purely instrumental [14, 22].

Proposition 4. *Rule Synthesis, grounded in causal inference mechanisms, yields design rules that transcend correlative limitations, provided that iterative validation supports epistemic closure.* Transforming correlations into design rules requires counterfactual reasoning and physics-constrained modeling to address the opacity of black-box predictors [23, 24]. Hybrid approaches integrating AI with physical laws enhance transferability and contextual validity [12, 28]. Iterative validation ensures that synthesized rules remain actionable and resilient, distinguishing epistemically grounded design guidance from performance-driven prediction alone [3, 4, 29].

Proposition 5. *Continuous epistemic feedback across framework stages is necessary to prevent knowledge drift and preserve justificatory integrity in iterative materials AI workflows.* As data distributions, modeling assumptions, and value priorities evolve, design rules that were initially warranted may lose epistemic validity [8, 19]. Without backward feedback from Rule Synthesis to Correlation Mapping and Bias Interrogation, such rules risk ossification, persisting despite degraded foundations [9, 23]. System-wide epistemic validation—rather than terminal assessment—maintains justificatory continuity by enabling reassessment in response to uncertainty, bias re-emergence, and contextual change [14, 15, 29].

Proposition 6. *Explicit epistemic governance enhances the transferability and legitimacy of AI-derived design rules by ensuring traceability of assumptions, values, and causal claims.* In high-stakes materials applications, the absence of transparent epistemic accountability undermines confidence in AI-mediated recommendations [16, 18]. Frameworks that encode bias auditing, value deliberation, and causal justification support responsible transfer across laboratories and application domains, even when predictive accuracy alone is insufficient [9, 12]. By situating knowledge extraction as an institutionally governed practice, such approaches align materials AI with norms of reproducibility, accountability, and ethical stewardship [13, 17, 22, 28].

Collectively, these propositions specify constraints on when and how AI-derived correlations in materials science may warrant interpretation as design rules. They do not prescribe specific algorithms or guarantee empirical success; rather, they delimit epistemic conditions under which transitions from association to prescription are defensible within scientific practice.

Results and Discussion

The proposed conceptual framework and its derived propositions offer a theoretical scaffold for advancing knowledge extraction in materials AI, addressing longstanding epistemic challenges while highlighting avenues for innovation. At its core, the model reconceptualizes AI as an epistemic tool rather than a mere predictor, enabling materials scientists to derive causally informed, value-aligned design rules. This shift has profound implications for the field, particularly in domains like sustainable materials development, where correlations alone fail to capture complex multiscale dynamics [2, 5]. By structuring extraction as iterative stages, the framework mitigates risks associated with opacity and bias, fostering transparency that aligns with epistemic values of justification and reliability [8, 13].

One key implication is the enhancement of interdisciplinary collaboration. Materials informatics often spans physics, chemistry, and data science, yet siloed approaches perpetuate knowledge gaps [19]. The framework's emphasis on value integration encourages stakeholder involvement and integrates diverse perspectives to refine rules for applications such as advanced composites or energy storage [4, 7]. This not only improves epistemic robustness but also addresses non-epistemic concerns, such as ethical data sourcing, ensuring that AI contributes to equitable scientific outcomes [15, 16]. Moreover, in high-stakes contexts, such as those involving critical infrastructure materials, the model's bias interrogation stage could reduce systemic errors, thereby promoting trust in AI-derived insights [9, 11].

However, limitations must be acknowledged. The framework assumes access to high-quality datasets and computational resources, which may not be universal, particularly in under-resourced settings [10, 12]. This could exacerbate global disparities in materials research, where biases in data availability favor well-studied systems [6, 25]. Additionally, while the model advocates causal inference in rule synthesis, implementing counterfactual reasoning in practice remains conceptually challenging, as AI models often lack inherent causality [23, 27]. Overreliance on the framework might also stifle exploratory science if value alignment overly constrains creativity [14, 20].

Future directions should explore extensions to emerging paradigms, such as generative AI for hypothetical materials [3, 24]. Conceptual refinements could incorporate dynamic feedback from real-world applications, testing propositions in specific subfields like nanomaterials [26, 28]. Philosophical inquiries into epistemic agency in human-AI hybrids would further enrich the model, examining how values evolve in collaborative ecosystems [9, 17]. Ultimately, this framework underscores the imperative for theoretical maturity in materials AI, positioning it as a catalyst for epistemically sound and socially responsible innovation [1, 21, 29].

Conclusion

This manuscript has advanced a conceptual framework for knowledge extraction in materials AI that addresses a foundational limitation of current data-centric paradigms: the tendency to conflate predictive correlation with scientific understanding. By articulating knowledge extraction as a staged epistemic process—encompassing correlation identification, bias interrogation, value integration, and causal rule synthesis—the framework clarifies the conditions under which AI-mediated outputs may warrant interpretation as design rules rather than provisional patterns.

The six propositions developed in this work specify constraints on inference, justification, and revision that operate across iterative materials AI workflows. Together, they emphasize that epistemic reliability depends not solely on model accuracy, but on the explicit management of bias, the articulation of value commitments, and the incorporation of causal reasoning supported by continuous epistemic feedback. This perspective reframes AI systems as contributors to scientific reasoning rather than substitutes for it, preserving the role of theory, explanation, and normative judgment in materials research.

While the framework does not prescribe specific algorithms or guarantee empirical outcomes, it provides a principled basis for evaluating when AI-derived correlations can support prescriptive claims about material design. In doing so, it contributes to the theoretical maturation of materials AI by situating knowledge extraction within established norms of scientific justification. As AI continues to shape materials discovery, such epistemically grounded approaches will be essential to ensuring that accelerated prediction translates into durable, trustworthy scientific knowledge.

Acknowledgements
None

Conflict of interest
None

Financial support
None

Ethics statement
None

Received: 15 Feb 2025 Revised: 04 Apr 2025 Accepted: 30 Jun 2025
Published online: 18 January 2026

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.

Chen C, Ong SP. A universal graph deep learning interatomic potential for the periodic table. Nat Comput Sci. 2022;2(11):718-28.

Katsura Y, Tanaka H, Takizawa H, Takahashi K. Systematic searches for new inorganic materials assisted by materials informatics. Sci Technol Adv Mater. 2024;25(1):2428154.

Wang Z, Sun Y, Liang J, Tang S, Dai J, Mao J, et al. Matgpt: A vane of materials informatics from past, present, to future. Adv Mater. 2024;36(9):2306733.

Ortega-Guerrero A, Espinosa-Ramos JI. Artificial intelligence reinventing materials engineering: A bibliometric review. Appl Sci. 2024;14(18):8143.

Foppa L, Sutton C, Ghiringhelli LM. Reproducibility in materials informatics: Lessons from ‘a general-purpose machine learning framework for predicting properties of inorganic materials’. Digit Discov. 2024;3(2):281-6.

Chen J, Ye X, Park TE, Yang Z, Barde A, Cairns K, et al. Materialsatlas.org: a materials informatics web app platform for materials discovery and survey of state-of-the-art. npj Comput Mater. 2022;8(1):72.

Gong S. The significance of materials informatics on material science. Appl Comput Eng. 2024;58:208-14.

Schwartz R, Vassilev A, Greene K, Perine L, Burt A, Hall P. Towards a standard for identifying and managing bias in artificial intelligence. NIST Spec Publ. 2022;1270:1-86.

Leavy S, O’Sullivan B, Siapera E. Data, power and bias in artificial intelligence. AI Soc Good Workshop. 2020:1-6.

Dunjic M. Values in science and ai alignment research. Soc Epistemol. 2024;38(4):458-75.

Koskinen I. We have no satisfactory social epistemology of ai-based science. Soc Epistemol. 2024;38(4):458-75. https://doi.org/10.1080/02691728.2023.2286253.

Gomez-Bombarelli R, et al. Materials informatics: A review of ai and machine learning tools, platforms, data repositories, and applications to architectured porous materials. Mater Today Commun. 2024;40:110198.

Chaudhuri A, et al. Artificial intelligence in materials by design: Critical review and perspectives on materials informatics to generative and agentic intelligence. Mater Des. 2024;in press.

Merchant A, et al. Data integrity in materials science in the era of ai: Balancing accelerated discovery with responsible science and innovation. J Mater Chem A. 2024;in press.

Chen A, McCloskey P, Sorger VJ. Bias in ai-based models for medical applications: Challenges and mitigation strategies. npj Digit Med. 2024;7(1):60.

Asooja K, et al. Systematic literature review on bias mitigation in generative ai. AI Ethics. 2024;in press.

Ratwani RM, Sutton K, Galarraga JE. Addressing ai algorithmic bias in health care. JAMA. 2024;332(13):1051-2.

Hamidieh K, et al. Researchers reduce bias in ai models while preserving or improving accuracy. arXiv. 2024;2406.16846.

Wu X, et al. Addressing bias in generative ai: Challenges and research opportunities in information management. J Bus Res. 2024;172:114425.

Leavy S. Gender bias in artificial intelligence: The need for diversity and gender theory in machine learning. In: Proceedings of the 1st international workshop on gender equality in software engineering; 2018:14-6.

Leavy S. Uncovering gender bias in media coverage of politicians with machine learning. arXiv preprint arXiv:2005.07734. 2020.

Ashik Shahul Hameed M, et al. Bias mitigation via synthetic data generation: A review. Electronics. 2024;13(19):3909.

Koskinen I. We still have no satisfactory social epistemology of ai-based science: A response to peters. Soc Epistemol Rev Reply Collect. 2024;13(5):11-7.

Sikimic V. The use of AI and epistemic values in science. Eindhoven University of Technology; 2024.

Stehr N. Social scientific knowledge about knowledge and information. Epistemol Philos Sci. 2023;60(3):131-70.

Baeva L. Epistemic status of artificial intelligence in medical practice: Ethical challenges. J Digit Diagn. 2024;5(3):319-25.

Flores L, Kim S, Young SD. Addressing bias in artificial intelligence for public health surveillance. J Med Ethics. 2024;50(3):190-4.

Agarwal R, Bjarnadottir M, Rhue L, Dugas M, Crowley K, Clark J, et al. Addressing algorithmic bias and the perpetuation of health inequities: An ai bias aware framework. Health Policy Technol. 2023;12(1):100702.