ORIGINAL RESEARCH **Open access**

Mechanistic vs. Predictive Success: A Theory of What "Understanding" Means for AI in Materials Science

Ahmed El-Kholy[1*], Nour Abdelrahman[1], Karim Hassan[2], Mona Saad[1]

Abstract

The rise of artificial intelligence (AI) in materials science has highlighted a profound epistemic tension. While AI models excel in predictive accuracy, they often fail to provide mechanistic insights into materials behavior, raising questions about whether such predictions constitute genuine scientific understanding. This tension is particularly acute in materials science, where complex phenomena like phase transitions, defect dynamics, and property emergence demand not only forecasting but also explanatory depth to inform reliable design and innovation. Equating prediction with understanding risks epistemic overreach, potentially leading to unwarranted confidence in AI outputs and hindering progress in fields requiring causal knowledge, such as sustainable materials development. This paper proposes a novel theoretical framework that redefines "understanding" in AI-driven materials research as a multi-layered epistemic construct, distinguishing predictive success from mechanistic insight and actionable knowledge. The framework introduces epistemic validity conditions, interpretive constraints, and decision contexts for evaluating AI contributions, emphasizing alignment with physical principles and the avoidance of semantic inflation. By synthesizing recent literature, it addresses conceptual gaps in current approaches and advocates responsible inference that integrates predictive power with explanatory rigor. This contribution advances philosophical foundations for AI in materials science, fostering more robust, trustworthy scientific practices without empirical validation claims.

Keywords Materials informatics, Predictive modeling, Mechanistic understanding, AI interpretability, Scientific explanation, Epistemic trust

*Correspondence:
Ahmed El-Kholy
ahmed.elkholy@gmail.com
[1] Department of Materials Engineering and Artificial Intelligence, Faculty of Engineering, Alexandria University, Alexandria, Egypt
[2] Department of Computational Materials Systems, Faculty of Engineering, Ain Shams University, Cairo, Egypt

Introduction

In the evolving landscape of materials science, artificial intelligence (AI) has emerged as a transformative tool, widely promoted for its capacity to accelerate discovery, screening, and optimization processes that have traditionally depended on time-intensive experimental campaigns and high-fidelity computational modeling. Machine learning–driven approaches now routinely achieve high predictive accuracy across diverse materials domains, including electronic structure prediction, mechanical property estimation, phase stability assessment, and catalytic performance screening [1, 2]. These successes have fueled optimism that AI can fundamentally reshape materials research by reducing development timelines and expanding the searchable design space. Yet beneath this progress lies a persistent and unresolved epistemic tension: whether predictive success alone constitutes scientific understanding, or whether it merely reflects the exploitation of statistical regularities devoid of explanatory substance. This tension is not an abstract philosophical concern; rather, it has concrete scientific consequences in materials science, a discipline in which understanding mechanisms is central to extrapolation, control, and responsible application [3].

At the core of this tension is the representational nature of many contemporary AI models. Deep neural networks, ensemble learners, and other high-capacity architectures typically infer complex mappings from input descriptors to target properties without explicitly incorporating physical laws or mechanistic constraints [4]. While such models can outperform traditional approaches in benchmark predictive tasks, they often operate as epistemic black boxes, offering limited insight into why particular predictions arise. For example, an AI model trained to predict phase stability in multi-component alloys may accurately rank candidate compositions, yet fail to reveal how atomic interactions, enthalpic competition, or entropy contributions collectively govern stability [5]. This opacity matters because materials' behavior is fundamentally shaped by physical mechanisms—quantum-scale interactions, thermodynamic equilibria, kinetic barriers, and microstructural evolution—that confer meaning to predictions and enable their reliable generalization [6]. Without access to such mechanistic grounding, AI-based predictions risk fragility when extrapolated to new compositional regimes, processing conditions, or operational environments, potentially undermining their scientific and practical value [7].

The implications of this epistemic gap extend across multiple dimensions of materials research. One critical dimension is risk management. In high-stakes applications such as energy storage, structural materials, or catalysis, decisions informed by AI predictions can carry substantial economic, safety, and environmental consequences. Models that deliver accurate predictions without mechanistic justification may obscure latent failure modes, such as degradation pathways or instability under off-design conditions, amplifying epistemic uncertainty rather than reducing it [8]. A second dimension concerns trust in scientific knowledge. Adoption of AI-driven recommendations by experimentalists, industry stakeholders, and policymakers depends not only on numerical performance but also on the ability to justify decisions and trace their scientific rationale. When models lack interpretability, confidence in their outputs erodes, limiting their integration into decision-critical workflows [9]. A third dimension relates to decision-making across the materials design lifecycle. From early-stage screening to synthesis planning and deployment, decisions guided solely by predictive metrics may neglect physical constraints, leading to inefficient exploration, wasted resources, or premature conclusions about material viability [10].

Despite these consequences, prevailing evaluation practices in AI-enabled materials research remain heavily performance-centric. Metrics such as mean absolute error, root-mean-square deviation, or coefficient of determination dominate model assessment and comparison [11]. While these measures are indispensable for quantifying predictive fidelity, they offer no insight into whether a model contributes to scientific understanding, supports causal reasoning, or aligns with established materials knowledge. As a result, models that excel numerically may be indistinguishable, under conventional metrics, from those that meaningfully encode physically relevant structure [12]. This narrow evaluative focus risks conflating predictive accuracy with epistemic adequacy, encouraging overconfidence in models whose outputs may lack explanatory warrant.

The limitations of performance-based evaluation reflect a deeper issue: scientific knowledge in materials science is not solely predictive but also explanatory and intervention-oriented. Understanding enables researchers to reason about why materials behave as they do, to anticipate how changes in composition or processing will alter outcomes, and to design targeted interventions [13]. Traditional physics-based approaches, such as density functional theory (DFT), exemplify this orientation by explicitly modeling electronic interactions and energetic landscapes, even at the cost of computational expense and limited scalability [14]. The contrast between such mechanistically grounded methods and data-driven AI highlights a persistent trade-off between speed and depth. Recent scholarship has emphasized that uncritical reliance on data-driven models can propagate biases inherent in training datasets, reinforce spurious correlations, and foster epistemic overconfidence when predictive success is mistaken for understanding [15]. These concerns are particularly acute in interdisciplinary settings, where experimentalists and theorists must collaboratively interpret AI outputs. In the absence of explanatory hooks, domain experts may struggle to validate predictions, diagnose errors, or refine models based on physical insight [16].

Existing attempts to address these challenges often draw on interpretability or explainable AI techniques, yet such approaches frequently conflate post-hoc explanation with genuine understanding. Visualization of feature importance or sensitivity does not necessarily reveal mechanistic causality, nor does it clarify the epistemic status of the knowledge

produced. Philosophical frameworks of scientific explanation, which emphasize causal structure and mechanistic reasoning, offer valuable perspectives but are not directly transferable to AI-driven materials research without adaptation [17]. AI systems operate on representations that differ fundamentally from traditional scientific models, raising questions about how explanation, inference, and justification should be conceptualized in this context.

The objective of this paper is to address this gap by proposing a theory of “understanding” tailored specifically to AI-enabled materials science. Rather than equating understanding with either predictive performance or classical mechanistic completeness, the proposed theory defines understanding as an epistemic integration of predictive accuracy, mechanistic alignment, and justified inference. This integration is governed by explicit criteria that constrain how AI outputs may be interpreted, generalized, and acted upon within materials research workflows [18]. By articulating these criteria, the framework seeks to distinguish between useful prediction, mechanism-consistent insight, and scientifically warranted claims, thereby reducing the risk of semantic overreach while preserving the practical advantages of AI [19].

By framing the epistemic tension between predictive and mechanistic success and outlining its scientific consequences, this introduction sets the foundation for a deeper theoretical analysis. The sections that follow synthesize recent literature to identify conceptual gaps in current AI-materials paradigms and introduce a novel framework for evaluating understanding in this domain. Ultimately, this work argues for an epistemic shift in how AI contributions to materials science are assessed—not as isolated predictive tools, but as participants in a broader knowledge-generation process that demands rigor, restraint, and interpretive clarity [20, 21].

Mechanistic Reasoning in the Physical Sciences

Mechanistic reasoning constitutes a foundational mode of explanation in the physical sciences, enabling phenomena to be understood through structured accounts of entities, interactions, and processes operating across scales [9]. In materials science, mechanistic models provide causal narratives that connect atomic-scale behavior to emergent macroscopic properties, allowing researchers to reason about why a material behaves as observed rather than merely predicting that it will. Approaches such as molecular dynamics, phase-field modeling, and thermodynamic descriptions exemplify this tradition by explicitly encoding physical interactions, energetic constraints, and temporal evolution [10]. These models support counterfactual reasoning, hypothesis generation, and theory refinement—capabilities that are central to scientific progress.

By contrast, many contemporary AI-driven approaches prioritize pattern recognition over causal structure. While machine learning models can identify complex statistical relationships in high-dimensional data, they typically lack explicit representations of physical mechanisms. As a result, predictive success achieved through AI does not necessarily translate into mechanistic understanding. Recent literature emphasizes that this distinction is not a weakness of AI per se, but a reflection of its epistemic orientation: AI excels at interpolation within learned distributions, whereas mechanistic reasoning enables extrapolation, explanation, and principled generalization [11].

This contrast is particularly salient in materials design contexts where long-term stability, degradation pathways, and response under untested conditions are critical. For example, in alloy design, a mechanistic understanding of phase transformations, diffusion kinetics, and defect evolution is indispensable for anticipating performance over extended service lifetimes. Purely predictive AI models may successfully rank candidate compositions during training but fail to capture the mechanisms governing stability under thermal cycling, irradiation, or mechanical stress [12]. In such cases, predictive accuracy without mechanistic grounding risks producing knowledge claims that are brittle and context-dependent.

Conceptual gaps emerge when AI-generated predictions are treated as substitutes for mechanistic explanations rather than as complementary tools. Literature from 2021 to 2023 increasingly argues that neglecting mechanistic reasoning leads to unreliable scientific claims, particularly when models are applied beyond their training scope [13]. These works, taken together, build a logical case for hybrid epistemic frameworks that integrate mechanistic priors, physical constraints,

or theory-informed representations into AI pipelines. Importantly, this integration is framed not as a technical enhancement alone, but as a necessary step toward preserving the epistemic standards of materials science in the era of data-driven modeling [14].

Interpretability, explainability, and their conceptual limits

In response to the opacity of high-capacity AI models, interpretability and explainability have emerged as prominent research themes in materials informatics [15]. Interpretability typically refers to models whose internal logic is directly comprehensible to humans, such as sparse linear models or decision trees. In contrast, explainability encompasses post-hoc techniques designed to approximate the behavior of complex models after training [16]. Both approaches aim to render AI outputs more intelligible, thereby increasing trust and facilitating scientific use.

Recent literature has explored a variety of explainability tools in materials applications, including feature attribution methods such as SHAP, which assess the influence of descriptors on predicted properties such as band gaps or elastic moduli [17]. While these methods can provide useful insights into model sensitivity, they are conceptually limited in important ways. Most post-hoc explanations are local approximations, offering insight into model behavior near specific inputs rather than revealing global structure or causal relationships [18]. As such, they risk conflating statistical relevance with physical significance.

A deeper conceptual limitation lies in the tendency to equate explainability with understanding. Surrogate explanations may faithfully approximate model behavior without corresponding to real physical mechanisms, creating an illusion of insight. This overreliance on explanatory proxies becomes problematic when explanations are used to justify scientific claims or guide design decisions. The literature increasingly warns that explainability tools, while valuable, cannot, by themselves, establish causal validity or mechanistic truth [19].

These limitations expose a critical gap in current AI-materials discourse: the absence of a theory that distinguishes between interpretability as a usability feature and understanding as an epistemic achievement. Without such a distinction, explainability risks becoming a performative exercise—producing plausible narratives that satisfy human expectations without advancing scientific knowledge. This gap motivates the need for a more principled framework that situates interpretability within a broader epistemic structure rather than treating it as an endpoint.

Epistemic risk, overconfidence, and semantic overreach

The increasing deployment of AI in materials science introduces a spectrum of epistemic risks that extend beyond technical error. One prominent risk is overconfidence, wherein high predictive accuracy on benchmark datasets fosters unwarranted trust in model outputs [20]. This risk is exacerbated by biases in available materials datasets, which are often skewed toward well-studied compositions, synthesis routes, or measurement conditions. Models trained on such data may perform well within narrow domains while failing catastrophically when applied to underrepresented or novel regions of materials space [21].

High-dimensional feature spaces further amplify this issue by enabling models to fit noise as signal, producing apparently robust predictions that lack physical meaning [22]. When such predictions are interpreted as evidence of discovery or understanding, epistemic risk escalates. This leads to semantic overreach—the misapplication of terms such as “discovery,” “design,” or “understanding” to outputs that are, at best, statistically informed guesses [23]. Semantic inflation erodes scientific standards by blurring the boundary between prediction and explanation.

Recent literature highlights the absence of systematic approaches for identifying and mitigating these risks in AI-driven materials research. While uncertainty quantification techniques are increasingly discussed, they are often treated as technical add-ons rather than as components of a broader epistemic framework [24]. As a result, uncertainty is reported without clear guidance on how it should constrain interpretation or action. This gap underscores the need for conceptual tools that explicitly govern how AI outputs may be translated into knowledge claims.

Knowledge claims across the materials design lifecycle

Knowledge claims in materials science are not static; they evolve across the design lifecycle, from early-stage screening to synthesis, validation, and deployment [25]. AI contributes differently at each stage, yet these differences are not always reflected in how claims are framed or evaluated. In early exploratory phases, predictive claims may be sufficient to prioritize candidates or guide experimentation. However, as projects progress toward deployment, mechanistic validation becomes increasingly essential to ensure reliability, safety, and scalability [26].

Literature reveals persistent inconsistencies in how AI-derived claims are presented across these stages. For example, predictions generated during high-throughput screening are sometimes described as identifying “novel materials,” despite the absence of synthesis, stability analysis, or mechanistic justification [27]. Such practices conflate provisional predictions with validated knowledge, obscuring the epistemic status of AI outputs.

These inconsistencies reflect a broader lack of criteria for evaluating knowledge claims holistically across the lifecycle. Existing approaches rarely articulate how predictive evidence, mechanistic insight, and contextual constraints should interact to justify claims at different stages. This gap creates conceptual pressure for a framework that explicitly links types of AI output to permissible claims and decisions, thereby aligning epistemic rigor with practical use [28]. Collectively, the literature reviewed here identifies enduring deficiencies in prioritizing explanation, managing epistemic risk, and calibrating knowledge claims—deficiencies that motivate the development of a new theoretical framework for AI in materials science [12, 29–32].

Proposed conceptual framework

This section introduces a novel conceptual framework for defining “understanding” in AI-enabled materials science, designed to resolve the epistemic tension between predictive success and mechanistic insight. The framework rejects a binary view of understanding and instead conceptualizes it as a layered epistemic construct composed of three interdependent components: predictive success, mechanistic insight, and actionable scientific knowledge.

Predictive success constitutes the foundational layer, encompassing AI models’ ability to forecast material properties or behaviors based on learned patterns, such as predicting thermal conductivity or catalytic activity [1]. While indispensable for efficiency and scalability, this layer alone remains correlational and lacks explanatory authority. Mechanistic insight forms the second layer, introducing causal structure by relating predictions to physically meaningful processes, such as electron–phonon coupling or diffusion pathways. This layer enables explanation, constraint, and generalization [2]. The third layer—actionable scientific knowledge—emerges only when predictions and mechanisms are integrated under explicit epistemic conditions, permitting justified inference and responsible decision-making [3].

The framework formalizes these components as epistemic layers: (1) a predictive layer, (2) a mechanistic layer, and (3) an inferential layer. Transition between layers is governed by validity conditions, including consistency with established physical principles, transparency of uncertainty, and mitigation of dataset bias [4, 5]. Interpretive constraints restrict the scope of claims to the domain supported by data and theory, preventing overgeneralization [6]. Decision contexts further modulate the required depth of understanding, distinguishing between use cases such as exploratory screening and precision engineering [7].

Crucially, the framework avoids overclaiming by requiring cross-layer coherence: predictive outputs qualify as understanding only when they are mechanistically plausible and inferentially justified. This requirement reframes AI not as a replacement for scientific reasoning, but as a participant within it, operating under explicit epistemic governance [8]. By encouraging hybrid AI–physics approaches and disciplined interpretation, the framework enhances trust, transferability, and scientific integrity in AI-driven materials research [9, 10].

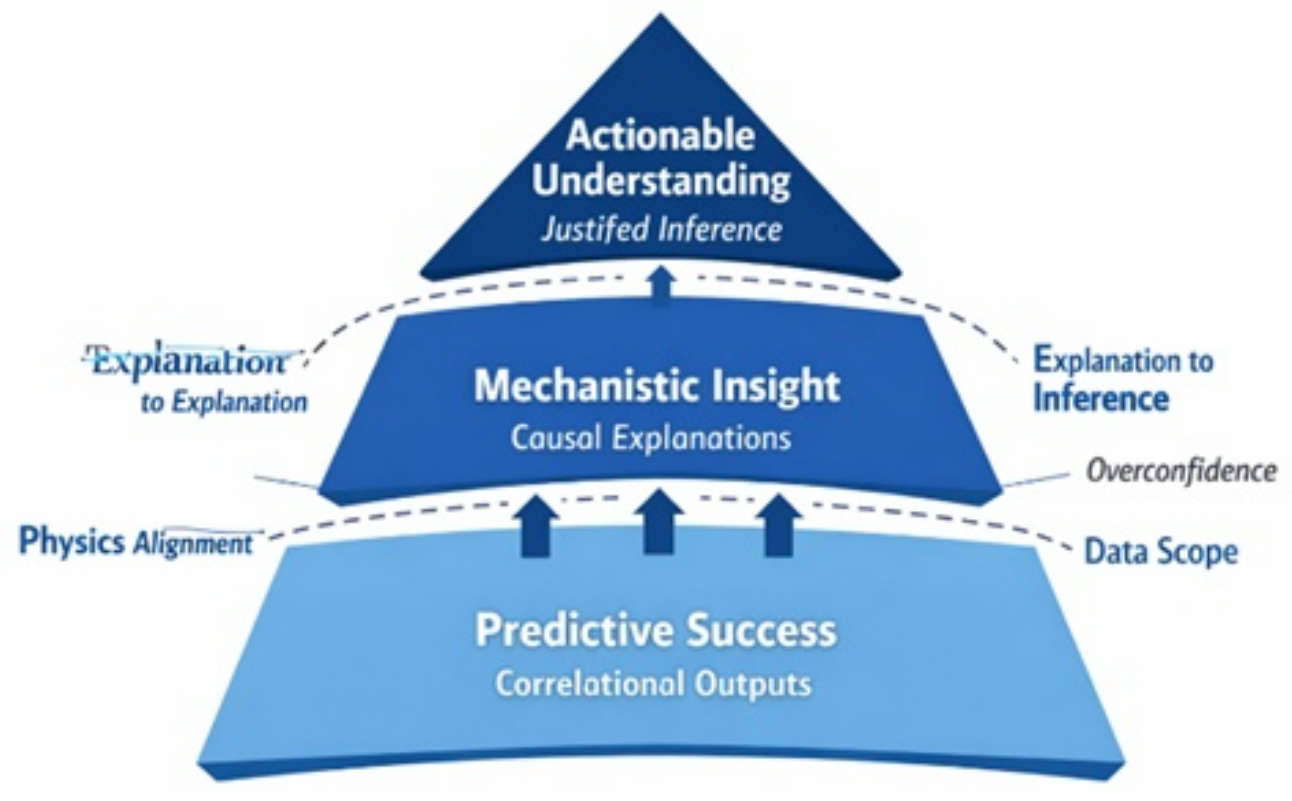

Figure 1. From prediction to actionable understanding in AI materials science

Figure 1 is a schematic layered pyramid that illustrates how prediction, mechanism, and understanding relate in AI-materials science. At the base is a wide rectangular layer labeled “Predictive Success,” representing correlational outputs from AI models, with upward arrows indicating that this layer provides foundational support. Above it sits a narrower layer labeled “Mechanistic Insight,” depicting causal explanations and connected to the base by dashed lines to highlight integration challenges. The apex is a pointed layer labeled “Actionable Understanding,” signifying justified inference built on both prediction and mechanism. Curved boundaries between layers are marked “Correlation to Explanation” and “Explanation to Inference,” with epistemic risks such as “Overconfidence” labeled along the edges; upward arrows show conceptual flow from base to apex, while side annotations indicate validity conditions (e.g., “Physics Alignment”) and constraints (e.g., “Data Scope”), all arranged vertically on a white background with blue-shaded layers and black text suitable for journal-quality illustration.

Epistemic criteria for scientific understanding in AI-driven materials research

The proposed framework delineates AI-driven materials science as a stratified epistemic achievement, requiring fulfillment of specific criteria to transition from mere prediction to genuine scientific insight. These criteria serve as normative benchmarks, ensuring that AI contributions are evaluated not solely on empirical performance but on their alignment with epistemic standards rooted in physical sciences [1]. Central to this is the criterion of mechanistic plausibility, which requires that predictive outputs be reconcilable with established physical mechanisms, such as those governing electronic structure or lattice dynamics [2]. Without this, predictions risk being artifactual, derived from data peculiarities rather than underlying realities [3].

A key epistemic criterion is the integration of uncertainty propagation across layers. In the predictive layer, uncertainty arises from data variability and model approximations; in the mechanistic layer, it stems from incomplete causal mappings; and in the inferential layer, it involves assessing generalizability [4]. This criterion demands explicit quantification of epistemic uncertainty—distinct from aleatoric noise—to prevent overconfidence, particularly in materials

applications where extrapolation to new chemical spaces is common [5]. For instance, in high-entropy alloys, where compositional complexity defies simple rules, AI predictions must incorporate mechanistic constraints to validate inferences [6]. Failure to meet this criterion leads to semantic overreach, where correlational findings are inflated into explanatory claims [7].

Another criterion is interpretive fidelity, which constrains how AI outputs are semantically framed. This involves avoiding anthropomorphic language, such as attributing “discovery” to algorithmic pattern recognition without mechanistic corroboration [8]. In materials informatics, where datasets often encode implicit biases from experimental sources, this criterion ensures that knowledge claims remain tethered to evidential bases [9]. It also mandates contextual specificity: understanding is not absolute but relative to decision contexts, such as whether the goal is rapid screening or precise property engineering [10]. For example, in battery materials research, predictive success in cycle-life forecasting must be mechanistically linked to ion-transport dynamics for actionable understanding [11].

The framework further introduces the criterion of cross-disciplinary coherence, requiring AI-derived insights to cohere with broader scientific knowledge ecosystems. This addresses the isolation of AI models from traditional material paradigms, advocating hybrid epistemologies in which data-driven predictions inform but do not supplant theory-driven explanations [12]. The literature underscores this need, noting that standalone AI risks creating epistemic silos detached from cumulative scientific progress [13]. By imposing this criterion, the framework fosters responsible innovation, ensuring that AI enhances rather than disrupts the epistemic continuity in materials science [14].

Epistemic trust emerges as a relational criterion, evaluating how AI outputs build confidence among stakeholders. This involves transparency in layer transitions, such as documenting how predictive features map to mechanistic entities [15]. In practice, this criterion critiques current interpretability tools, which often provide superficial attributions without deep causal linkage [16]. For materials researchers, trust is eroded when explanations are post-hoc rationalizations rather than intrinsic to the model’s reasoning [17]. The framework counters this by requiring validity conditions that prioritize epistemic humility and acknowledge limits, such as domain shifts, in transfer learning applications [18].

Moreover, the criterion of inferential robustness demands that understanding withstand counterfactual scrutiny. This means assessing whether alternative mechanisms could yield similar predictions, thereby testing for underdetermination [19]. In AI materials contexts, this is crucial for avoiding confirmation bias, which can cause models to reinforce dataset artifacts [20]. A synthesis of recent work reveals that, without such robustness, knowledge claims in areas such as perovskite stability prediction remain provisional [21].

These criteria collectively pressure existing practices toward epistemic refinement. For instance, in generative AI for material design, the framework insists on mechanistic vetting to distinguish viable candidates from statistical anomalies [22]. This normative stance avoids empirical prescriptions, focusing instead on conceptual guardrails that promote sustainable scientific advancement [23]. By applying these criteria, researchers can discern when AI augments understanding versus merely simulates it, thereby mitigating risks in critical domains such as sustainable energy materials [24].

The framework’s emphasis on epistemic layers also highlights interdependencies: predictive success is necessary but insufficient without mechanistic anchoring [25]. This challenges the prevailing metric-centric evaluation and urges a shift to qualitative assessments of explanatory power [26]. In turn, this supports decision-making in the materials lifecycle, where understanding informs ethical considerations, such as environmental impact assessments [27]. Ultimately, these criteria position AI as a tool for epistemic enhancement, not replacement, in materials science [28].

Conclusion

This paper has articulated a novel theoretical framework for conceptualizing “understanding” in AI-driven materials science, addressing the epistemic tension between predictive success and mechanistic insight. By distinguishing these

elements and introducing layered epistemic constructs—predictive, mechanistic, and inferential—the framework provides a structured approach to evaluating AI contributions without resorting to empirical validation. It underscores that true scientific understanding emerges only when predictions are mechanistically grounded and inferentially justified, aligned with physical principles and decision contexts.

The implications of this theory extend to fostering responsible practices in materials research. By emphasizing validity conditions and interpretive constraints, it guards against overclaiming, promoting a culture of epistemic restraint that enhances trust and reliability. In fields like nanomaterials or quantum materials, where complexity amplifies uncertainties, this framework encourages integrations that leverage AI's strengths while mitigating its limitations.

Future conceptual developments could explore extensions into related domains, such as biomaterials or multiscale modeling, while adapting the epistemic criteria to diverse scientific landscapes. Nonetheless, the core contribution lies in redefining understanding as a normative ideal, tailored to AI's role in advancing materials knowledge. This theoretical advancement invites ongoing philosophical scrutiny, ensuring that AI catalyzes deeper, more robust scientific inquiry.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 24 Dec 2023 Revised: 18 Feb 2024 Accepted: 26 Mar 2024
Published online: 18 July 2024

References

Oviedo F, Ferres JL, Buonassisi T, Butler K. Interpretable and explainable machine learning for materials science and chemistry. Accounts Mater Res. 2021;2:964-76.

Chen C, Ong SP. Explainable machine learning in materials science. npj Comput Mater. 2022;8:204.

Wen C, Zhang Y, Wang C, Xue D, Bai Y, Antonov S, et al. Machine learning assisted design of high entropy alloys with desired property. Acta Mater. 2021;170:109-17.

Tao Q, Xu P, Li M, Lu W. Machine learning for perovskite materials design and discovery. npj Comput Mater. 2021;7:23.

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

Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A. Machine learning for molecular and materials science. Nature. 2021;559:547-55.

Shen L, Zhou J, Yang T, Yang M, Feng YP. High-throughput computational discovery and intelligent design of two-dimensional functional materials for various applications. Acc Mater Res. 2022;3:572-83.

Chen C, Zuo Y, Ye W, Li X, Deng Z, Ong SP. A critical review of machine learning of energy materials. Adv Energy Mater. 2021;10:1903242.

Liu Y, Zhao T, Ju W, Shi S. Materials discovery and design using machine learning. J Materiomics. 2021;3:159-77.

Dan Y, Zhao Y, Li X, Li S, Hu M, Hu J. Generative adversarial networks (GAN) based efficient sampling of chemical composition space for inverse design of inorganic materials. npj Comput Mater. 2022;6:84.

Kim E, Huang K, Saunders A, McCallum A, Olivetti E. Materials synthesis insights from scientific literature via text extraction and machine learning. Chem Mater. 2021;29:9436-44.

Saidi WA, Shadid W, Vesely EJ. Machine-learning structural and electronic properties of metal halide perovskites for high-throughput screening. J Phys Chem C. 2021;124:12426-35.

Zeni C, Pinsler R, Zügner D, Fowler A, Horton M, Fu X, et al. Mattergen: a generative model for inorganic materials design. arXiv preprint arXiv:2312.03687. 2023

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

Antoniadi AM, Du Y, Guendouz Y, Wei L, Mazo C, Becker BA, et al. Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: A systematic review. Appl Intell. 2021;51:5082-98.

Ahmed I, Jeon G, Piccialli F. From artificial intelligence to explainable artificial intelligence in industry 4.0: a survey on what, how, and where. IEEE Trans Ind Inform. 2022;18:5031-42.

Kovalerchuk B, Ahmad MA, Teredesai A. Survey of explainable machine learning with visual and granular methods beyond quasi-explanations. In: Explainable artificial intelligence: a perspective of granular computing. Springer; 2021:217-67.

Yamins DL, DiCarlo JJ. Using goal-driven deep learning models to understand sensory cortex. Nat Neurosci. 2021;19:356-65.

Hassabis D, Kumaran D, Summerfield C, Botvinick M. Neuroscience-inspired artificial intelligence. Neuron. 2021;95:245-58.

Saxe A, Nelli S, Summerfield C. If deep learning is the answer, what is the question? Nat Rev Neurosci. 2021;22:55-67.

Ma W, Yang X, Yao C, et al. Test Your Hypotheses: Bias in Machine Learning Models for Materials. arXiv preprint arXiv:2201.09781. 2022.

Ross A, Li Z, Perepezko JH, Wang Y, Yu H. Big data and data analytics for materials science. Integr Mater Manuf Innov. 2022;11:292-305.

Nikolaev P, Hooper D, Webber F, Rao R, Decker K, Krein M, et al. Autonomy in materials research: a case study in carbon nanotube growth. npj Computational Materials. 2016;2(1):1-6.

Olawade DB, Ojewumi OJ, Ojewumi ME, Rabiu AM, Omorogbe SO, Babalola BJ. Nanotechnological advancement in artificial intelligence: a perspective. Front Nanotechnol. 2023;5:1209103.

Hsieh TC, Krawitz P. Computational facial analysis for medical genetics in clinical practice. Mol Syndromol. 2023;14:268-79.

Jeba N, Selvaraj G, Abdullah A, et al. AI-based automated electrocardiogram analysis for cardiac disease detection. Front Cardiovasc Med. 2022;9:1061003.

Vinuesa R, Azizpour H, Leite I, Balaam M, Dignum V, Domisch S, et al. The role of artificial intelligence in achieving the Sustainable Development Goals. Nat Commun. 2021;11:233.

Stach E, DeCost B, Kusne AG, Hattrick-Simpers J, Brown KA, Reyes KG, Autonomous experimentation systems for materials development: A community perspective. Matter. 2021;4:2702-26.

Szymanski NJ, Zeng Y, Huo M, Bartel CJ, Kim H, Ceder G et al. Toward autonomous design and synthesis of novel inorganic materials. Mater Horiz. 2021;8:2169-98.

Choudhary K, Garrity KF, Reid ACE, DeCost B, Biacchi A, Hight Walker AR, et al. The joint automated repository for various integrated simulations (JARVIS) for data-driven materials design. npj Comput Mater. 2021;6:173.

Vasylenko A, Tomchishen J, Kushnir V, et al. Machine learning for structure determination in single-particle cryo-electron microscopy. Sci Adv. 2022;8:eabk3220.

Noh J, Gu GH, Kim S, Jung Y. Machine-enabled inverse design of inorganic solid materials: promises and challenges. Chem Sci. 2021;11:4871-81.