ORIGINAL RESEARCH Open access

Algorithmic Simplicity as Scientific Virtue: A Conceptual Tension in Materials AI Design

Siti Rahman[1*], Ahmad Zaki[1], Nurul Huda[2], Amir Faisal[1], Lim Wei[2]

Abstract

In the rapidly evolving field of Artificial Intelligence for Materials Science, algorithmic simplicity is frequently championed as an epistemic virtue, with practitioners prioritizing linear models, shallow architectures, and parsimonious descriptors under the assumption that simpler solutions inherently promote scientific insight and reliability. This paper critically examines the assumption that simplicity is always a scientific virtue in materials AI design, arguing instead that an overemphasis on simplicity introduces high epistemic costs by obscuring the multifaceted, nonlinear, and multi-scale nature of materials phenomena. The analysis unfolds through four interconnected critique points: first, the inherent trade-off between simplicity and predictive accuracy in capturing complex interactions; second, the risk of simplicity functioning as an obscurant that produces misleading yet confident representations; third, the problematic conflation of simplicity with interpretability, where the two concepts are treated as synonymous despite their distinct epistemic roles; and fourth, the fundamental mismatch between simplicity-prioritizing approaches and the intrinsic complexity demanded by real materials systems. These critiques reveal substantial consequences of simplicity bias, including missed opportunities for discovery, underestimation of uncertainty, premature model acceptance, and inefficient research pathways. Ultimately, the paper proposes alternative approaches that embrace appropriate complexity—matching model sophistication to problem demands, employing regularized complexity, leveraging ensemble methods, designing structured complex architectures, and adopting complexity-aware evaluation frameworks—thereby advocating for a more nuanced valuation of model complexity in service of genuine scientific understanding in materials discovery.

Keywords Materials AI, Algorithmic simplicity, Occam’s razor, Model complexity, Interpretability trade-off, Epistemic virtue

*Correspondence:
Siti Rahman
siti.rahman@outlook.com
[1] Department of Intelligent Materials Engineering, University of Malaya, Kuala Lumpur, Malaysia
[2] Department of Computational Materials Analytics, Universiti Teknologi Malaysia, Johor Bahru, Malaysia

Introduction

Materials AI has increasingly embraced algorithmic simplicity as a guiding principle in model design, often favoring linear regression techniques, small decision trees, or low-dimensional embeddings for property prediction tasks. This preference stems from a deeply ingrained belief that simpler models align with core scientific ideals of clarity, generalizability, and trustworthiness. Yet the central question persists: is simpler always better when confronting the intricate realities of materials systems? This paper offers a conceptual critique of the assumption that algorithmic simplicity serves as an unalloyed scientific virtue in materials AI design [1-4].

While simplicity can indeed mitigate overfitting and facilitate human comprehension, materials science frequently involves phenomena governed by high-dimensional interactions, emergent behaviors, and multi-scale hierarchies that resist reduction to parsimonious forms. The seed reference by Hansen [3] articulates precisely this conceptual tension, positioning algorithmic simplicity against the demands of accurate materials representation. Historical appeals to Occam's razor, as discussed by Jefferys and Berger, suggest that simpler explanations should be preferred when they suffice. Yet, Bayesian perspectives from MacKay highlight how complexity penalties must be weighed against data fit [1, 2]. In materials contexts, however, such principles risk oversimplification when applied uncritically.

The problem manifests across workflows where researchers default to simple descriptors over graph-based or deep representations, or opt for shallow networks instead of architectures capable of learning hierarchical features. Butler *et al.* survey machine learning applications in molecular and materials science, noting the appeal of straightforward approaches for rapid prototyping, while Schmidt *et al.* document the proliferation of machine learning in solid-state materials, often with an implicit bias toward parsimony [4, 5]. Chen *et al.* [6] introduce graph networks as a more expressive framework, yet many subsequent works revert to simpler baselines under the guise of virtue.

This paper argues that the valorization of simplicity in materials AI carries epistemic risks that outweigh its benefits in many scenarios. Materials properties—ranging from electronic band structures to mechanical responses in disordered alloys—often depend on subtle, nonlinear couplings that simple models systematically undervalue. Zunger's discussion of inverse design underscores the need for sophisticated search strategies in functional materials, implicitly challenging overly reductive modeling [7]. By tracing the historical roots of simplicity as virtue and surveying its manifestations in contemporary materials AI, the subsequent sections build toward a structured critique. The goal is not to reject simplicity outright but to distinguish its legitimate applications from problematic overextensions, ultimately calling for a balanced epistemology that values appropriate complexity where materials demand it [3, 8, 9].

Figure 1 illustrates the conceptual architecture through which simplicity bias in materials AI translates a historically valued epistemic heuristic into recurrent modeling practices, four forms of epistemic distortion, downstream scientific consequences, and complexity-aware corrective alternatives.

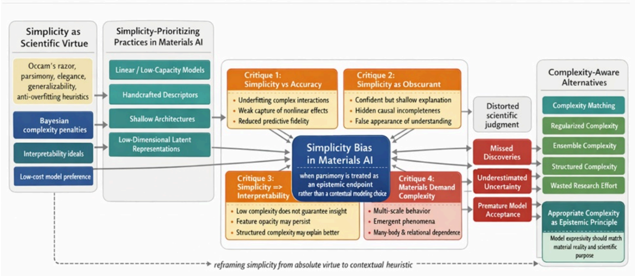

Figure 1. Conceptual Architecture of Simplicity Bias in Materials AI.

Figure 1 shows how simplicity, originally valued as a scientific heuristic, becomes operationalized as recurrent low-complexity modeling practices in materials AI, generating four epistemic distortions: reduced predictive adequacy, obscurant simplification, conflation of simplicity with interpretability, and mismatch with the ontological complexity of materials systems. These distortions produce downstream scientific consequences, including missed discoveries, underestimated uncertainty, premature model acceptance, and wasted research effort. **Figure 1** concludes with five complexity-aware alternatives that reposition simplicity as a contextual heuristic within a broader principle of appropriate complexity.

Simplicity as Scientific Virtue

The notion of simplicity as a scientific virtue traces back centuries, crystallized in Occam's razor, which posits that entities or explanations should not be multiplied beyond necessity. Jefferys and Berger explore this principle through a Bayesian lens, demonstrating how simpler models often receive higher posterior probability when they adequately explain observations [1]. MacKay's work on Bayesian interpolation further formalizes how complexity is penalized in model selection, promoting parsimony as a safeguard against overfitting [2]. In philosophy of science, parsimony functions as an epistemic heuristic, guiding inference toward explanations that invoke fewer assumptions while maintaining explanatory power.

Legitimate roles for simplicity abound in scientific practice. Simple models prevent overfitting by constraining parameter space, thereby enhancing generalization to unseen data—a point echoed in various machine learning contexts. They also promote interpretability, allowing researchers to trace causal pathways or physical mechanisms more readily. In materials informatics, parsimonious descriptors can distill essential chemical or structural features, facilitating rapid screening and hypothesis generation. Vasudevan *et al.* [9] emphasize the value of parsimony alongside Bayesianity and causality in moving beyond off-the-shelf deep learning for materials problems.

Yet simplicity's virtue is contextual rather than absolute. Bayesian model comparison, as MacKay outlines, balances fit and complexity through evidence metrics, but this assumes the true generating process lies within the considered model class [2]. When reality exceeds simple assumptions—as is common in condensed matter systems—parsimony can mislead. The seed paper by Hansen [3] identifies this tension explicitly in materials AI, where algorithmic simplicity collides with the need for expressive representations.

Parsimony in scientific inference also serves aesthetic and pragmatic ends. Elegant equations or low-parameter models evoke a sense of beauty and economy, aligning with broader scientific values. Butler *et al.* [4] highlight how machine learning for materials benefits from interpretable baselines before scaling to complexity. Schmidt *et al.* [5] review advances where simpler machine learning variants provide initial insights into solid-state phenomena. However, these benefits hold primarily when the underlying phenomena admit compact descriptions. In cases of emergent complexity, such as phase transitions or defect interactions, enforcing simplicity distorts understanding rather than illuminating it.

Acknowledging these legitimate roles does not preclude critique. Simplicity functions best as a regulative ideal or starting point, not an endpoint. Over time, scientific progress often involves moving from simple approximations to more nuanced models as data and theory accumulate. The history of physics, from Newtonian mechanics to quantum field theory, illustrates this progression. In materials AI, a similar trajectory is observable but frequently arrested by an undue attachment to parsimony. Chen *et al.*'s graph networks exemplify a shift toward greater expressivity for molecular and crystal systems, challenging the default simplicity bias [6]. Zunger advocates inverse design strategies that inherently require navigating complex search spaces [7].

Thus, while simplicity retains value in preventing gratuitous complexity and fostering generalization, its elevation to an unqualified virtue risks stifling inquiry in domains where complexity is ontologically necessary. The following sections examine how this plays out specifically in materials AI applications [1-3, 8, 10].

Table 1 distinguishes legitimate simplicity from epistemically harmful simplicity by showing that parsimony contributes to scientific understanding only when it remains proportionate to the causal and representational demands of the materials problem.

Table 1. Distinguishing legitimate simplicity from epistemically harmful simplicity in materials AI

Dimension	Legitimate simplicity	Epistemically harmful simplicity	Material illustra
Epistemic role	Serves as a provisional heuristic or baseline	Treated as an intrinsic scientific virtue regardless of context	Using a l baseline an init compar
Relation to phenomenon	Matches relatively low-complexity or well-behaved systems	Imposed despite nonlinear, emergent, or multi-scale material behavior	Applyi compa descripto disorde alloys amorph syster
Model capacity	Constrained but sufficient for the causal or predictive burden of the task	Underpowered relative to the interaction depth in the system	Shallo predictor bandga phas behavior higher-o depende
Interpretability status	Supports insight when variables retain a clear physical meaning	Assumed to be interpretable merely because architecture is simple	Sparse n built c opaqu enginee descrip
Uncertainty behavior	Offers bounded generalization when assumptions	Masks unmodeled variance and produces overconfident outputs	Simp stability proper mode extrapol beyond

	are transparent		trainir regim
Discovery potential	Useful for screening familiar patterns and generating first-pass hypotheses	Narrows the search space prematurely and suppresses novel regimes	Prefere for sim composi based ru high-ent alloy explora
Validation demand	Requires empirical checking proportional to claims made	Gains credibility through elegance or convenience rather than stress-testing	Accept compa mode because are eas inspe
Relation to complexity	Filters gratuitous complexity while allowing escalation when needed	Avoids complexity a priori, even when ontology demands it	Refusa adopt gı or struct models relatio materi proble
Scientific function	Clarifies without erasing essential dynamics	Simplifies by truncating causally relevant structure	Averag away lo heteroge in meta glasses defec
Normative implication	Simplicity is a contextual modeling choice	Simplicity is mistaken for a universal epistemic endpoint	Defau preferenc parsimor architect across materi tasks

Simplicity in Materials AI

Within materials AI, algorithmic simplicity manifests in multiple recurring forms, each underpinned by the assumption that parsimonious models yield superior scientific outcomes. Linear models for property prediction remain popular due to their transparency and low computational cost. Researchers often employ simple descriptors—such as elemental compositions or basic geometric features—rather than complex graph or embedding representations. Shallow networks or decision trees are preferred over deep architectures, and low-dimensional latent spaces are favored for visualization and analysis [4, 5, 9].

Butler *et al.* [4] provide a comprehensive overview of machine learning for molecular and materials science, noting the widespread use of straightforward regression techniques as entry points for property prediction. Schmidt *et al.* [5] document recent advances, where many studies begin with baseline simple models before considering more elaborate alternatives. Chen *et al.* [6] propose graph networks as a universal framework, yet their work is often cited alongside simpler baselines that assume lower complexity suffices for many crystals and molecules.

The implicit assumption throughout much of the literature is that simpler equates to better—more robust, more interpretable, and more aligned with physical intuition. Vasudevan *et al.* [9] call for parsimony in representations while critiquing unbridled deep learning, reinforcing the notion that excessive complexity should be curtailed. Lu *et al.* apply interpretable machine learning strategies to soft-magnetic properties in metallic glasses, favoring approaches that maintain relative simplicity [10]. Zhong *et al.* discuss explainable machine learning in materials science, where simpler surrogate models are frequently invoked to approximate complex black-box predictions [8].

Surveying the field reveals a pattern: simple linear or tree-based models for bandgap prediction, formation energy estimation, or stability assessment dominate initial explorations. Oviedo *et al.* [11] address interpretable and explainable methods for materials and chemistry, often contrasting them with more complex neural approaches. Desai *et al.* [12] introduce parsimonious neural networks that learn physical laws, blending simplicity with targeted expressivity. Gallegos *et al.* [13] explore explainable chemical AI, where parsimony aids in deriving insights from molecular properties.

In solid-state contexts, Zhou *et al.* [14] employ machine learning for phase design in high-entropy alloys, beginning with relatively compact models. Singh *et al.* [15] augment interpretable models with language techniques, yet retain core simplicity in base structures. Moitzi *et al.* [16] and Kim *et al.* [17] present frameworks for multi-principal element alloys and metamaterials that, while advancing capability, still navigate tensions between simple arithmetic operations in latent spaces and underlying complexity.

La Cava *et al.* [18] and Korolev *et al.* [19] contribute to symbolic regression and transformer-based predictions, where simplicity in symbolic forms or careful regularization is prized. Bell *et al.* [20] empirically examine accuracy-explainability trade-offs, underscoring the pull toward simpler models in practical deployments. Dziugaite *et al.* [21] analyze statistical impacts of enforcing interpretability through simplicity constraints.

This prevalence of simplicity in materials AI is not accidental but rooted in the epistemic valorization traced earlier [1-3]. Papers routinely justify model choices by invoking reduced complexity as a virtue, citing improved generalization or easier validation. Yet this survey also hints at underlying strains. When simple models underperform on validation sets involving disordered or multi-component systems, the field often responds by incremental tweaks rather than reconsidering the simplicity imperative. The next sections critique this pattern through targeted lenses [4, 5].

Critique Point 1: Simplicity versus Accuracy

A primary critique of simplicity-prioritizing approaches in materials AI centers on the unavoidable trade-off with predictive accuracy. Simple models, by design, constrain representational capacity, often failing to capture the nonlinear and high-order interactions that govern many material properties. For instance, a linear regression model applied to electronic bandgap prediction may adequately fit simple semiconductors but systematically deviates when confronted with complex band structures influenced by orbital hybridizations, spin-orbit coupling, and lattice distortions [5, 6].

In high-entropy alloys, simple compositional descriptors frequently miss entropic stabilization effects and local chemical ordering that emerge only through intricate many-body interactions. Zhou *et al.* illustrate machine learning guided exploration of phase design rules, where compact models risk overlooking subtle compositional thresholds [14]. Similarly, for thermal stability in Fe-based metallic glasses, Lu *et al.* demonstrate that interpretable yet relatively simple strategies must be augmented to handle amorphous complexity adequately [10].

A second materials-specific example arises in interface phenomena, such as grain boundaries or heterostructures. Simple low-dimensional embeddings cannot encode the multi-scale energetics and electronic reconstructions at play. Zunger's inverse design framework implicitly requires navigating far richer spaces than parsimonious models afford [7]. Chen *et al.*'s graph networks succeed precisely because they move beyond simplistic pairwise descriptors to capture crystal connectivity more faithfully [6].

Third, in metamaterial design, Kim *et al.* show that even simple arithmetic in latent spaces benefits from underlying complex representations; purely reductive models fail to generate diverse functional responses [17]. These cases argue that accuracy should, in targeted scenarios, override the default preference for simplicity. While regularization techniques can temper complexity, as noted in broader discussions of model selection [2], materials AI often encounters regimes where the data-generating process demands greater expressive power.

The critique does not advocate unchecked complexity but identifies the epistemic cost of systematically subordinating accuracy to simplicity. When simple models produce lower errors on training data through underfitting rather than genuine insight, they undermine the very scientific progress they purport to support. Butler *et al.* and Schmidt *et al.* document workflows where initial simple models serve as benchmarks, yet persistent reliance on them delays adoption of more accurate, albeit complex, alternatives [4, 5]. Vasudevan *et al.* caution against off-the-shelf approaches lacking sufficient parsimony balanced with needed capacity [9].

In summary, simplicity versus accuracy constitutes a genuine tension rather than a false dichotomy resolvable by defaulting to the former. Materials AI would benefit from explicit acknowledgment that, for certain properties and systems, accepting higher complexity yields superior epistemic returns [3, 8, 11, 20, 22, 23].

Critique Point 2: Simplicity as Obscurant

Overly simple models in materials AI can function not as clarifiers but as obscurants, generating predictions that appear confident while masking underlying physical realities. A simple descriptor-based model for phase stability might fit available data yet ignore critical vibrational entropy or electronic entropy contributions, leading to overgeneralized rules that fail under varied conditions. This produces a false sense of understanding, where the model "explains" observations through reductive mechanisms that do not correspond to actual causal structures [9, 14].

Consider disordered systems such as amorphous materials or glasses. A parsimonious linear combination of average coordination numbers may correlate with macroscopic properties but obscures local environment heterogeneity responsible for property distributions. Lu *et al.*'s work on soft-magnetic properties highlights how even interpretable strategies must confront this risk when simplicity truncates relevant variance [10]. Similarly, in high-entropy alloys, simple rule-based models for phase formation can obscure the role of competing electronic and configurational effects, as explored in machine learning appraisals [14].

A further example involves defect dynamics and diffusion in solids. Simple activation energy models based on averaged barriers neglect correlated jumps or trap interactions that dominate transport in real microstructures. Such models yield smooth but misleading predictions, obscuring pathways for materials optimization. Oviedo *et al.* discuss challenges in achieving genuine interpretability when simplicity hides these nuances in chemistry and materials contexts [11].

Simplicity here acts as an obscurant by promoting surface-level correlations as deep insights. Gallegos *et al.* [13] note that while explainable AI can derive properties, overly parsimonious forms risk confident but incomplete narratives. Desai *et al.* [12] advocate parsimonious networks for physical laws, yet warn against extremes that distort learned relationships.

The epistemic harm lies in how such models discourage deeper investigation. Researchers may accept "simple and sufficient" explanations prematurely, halting inquiry into emergent or context-dependent behaviors. Zhong *et al.* [8] address explainable machine learning where surrogate simplicity can mask the true complexity of underlying processes. This critique distinguishes productive simplification—which distills without loss of essential dynamics—from obscurant simplicity that erases critical information.

In materials AI, where validation against experiment or higher-fidelity simulation is resource-intensive, the allure of simple models amplifies this danger. The field must remain vigilant against simplicity that conceals rather than reveals [1-3, 16, 17].

Critique Point 3: The Simplicity-Interpretability Conflation

A pervasive issue in materials AI lies in the frequent conflation of algorithmic simplicity with interpretability, as if the two were interchangeable or even synonymous. Many practitioners assume that reducing model complexity—through linear regressions, shallow trees, or sparse descriptors—automatically yields greater insight into underlying mechanisms. Zhong *et al.* highlight the common tradeoff where increased model complexity challenges explainability, yet this does not imply that all simple models are inherently interpretable or that complex ones must remain opaque [8]. The seed conceptual paper by Hansen [3] directly engages this tension, questioning whether prioritizing simplicity truly advances understanding in materials design.

In practice, simple models can obscure interpretability in subtle ways. A linear model with dozens of hand-crafted or high-dimensional descriptors may appear parsimonious in functional form but becomes opaque when tracing how each coefficient contributes to a prediction, especially if features are collinear or derived through opaque preprocessing. Butler *et al.* [4] survey machine learning applications and note that while baseline simple models aid initial exploration, their interpretability diminishes when features lose direct physical mapping. Conversely, structured complex models—such as graph networks with attention mechanisms or physics-informed architectures—can offer clearer mechanistic insights by explicitly encoding symmetries or hierarchical relations. Chen *et al.* [6] demonstrate this with graph networks that provide universal yet expressive frameworks for molecules and crystals, allowing interpretation through learned connectivity rather than enforced sparsity.

Three materials-specific examples illustrate the conflation's pitfalls. First, in predicting properties of Fe-based metallic glasses, Lu *et al.* employ interpretable strategies that balance simplicity with domain relevance; however, forcing excessive parsimony risks losing insight into local atomic environments that drive magnetic and thermal behaviors, where moderately complex representations better reveal structure-property links [10]. A purely linear descriptor set might yield a compact equation, but it fails to disentangle competing effects like compositional fluctuations and short-range order. Second, for high-entropy alloys, simple rule-based or low-parameter models, as critiqued in phase design explorations by Zhou *et al.*, conflate ease of formulation with true interpretability; emergent configurational entropy and local ordering require models that can articulate multi-element interactions without reductive averaging that hides causal drivers [14]. Third, in metamaterial or inverse design contexts, Zunger emphasizes sophisticated search strategies; here, overly simple latent spaces or embeddings may produce numerically stable outputs but obscure how geometric or electronic degrees of freedom couple, whereas structured complex models can trace pathways to target functionalities more transparently [7].

This conflation carries epistemic costs by discouraging the development of models that achieve interpretability through design rather than reduction. Oviedo *et al.* discuss how interpretable and explainable techniques for materials science and chemistry must navigate trade-offs among completeness, understandability, and scientific validity, showing that complex models can yield domain-grounded explanations when properly structured [11]. Vasudevan *et al.* argue for parsimony, Bayesianity, and causality beyond off-the-shelf deep learning. Yet, their call implicitly decouples raw simplicity from genuine insight, advocating representations that are compact yet causally informative [9].

Distinguishing the concepts is essential: simplicity refers to low parameter count or shallow architecture, while interpretability concerns the degree to which a model's internal logic maps to domain concepts accessible to experts. Legitimate simplicity—such as regularization for generalization—differs from problematic oversimplification that sacrifices mapping fidelity. The field would benefit from explicit decoupling, evaluating models on both dimensions independently rather than assuming one entails the other [1, 2, 20-22]. By maintaining this distinction, materials AI can pursue expressive yet insightful architectures without defaulting to reductive heuristics.

Critique Point 4: Materials Demand Complexity

Materials systems are ontologically complex, exhibiting multi-scale hierarchies, emergent phenomena, and nonlinear couplings that simple models fundamentally struggle to accommodate. High-entropy alloys, for example, derive unique properties from configurational disorder, local chemical heterogeneities, and competing electronic interactions across multiple principal elements—features that resist capture by low-dimensional or linear approximations. Moitzi *et al.* and related works on multi-principal element alloys underscore how ab initio frameworks reveal trade-off relationships only when complexity is explicitly modeled [16]. Enforcing simplicity here risks collapsing rich phase spaces into misleading averages.

A second example concerns interfaces and defects in crystalline or amorphous materials. Grain boundaries, heterointerfaces, and defect clusters involve long-range elastic fields, charge redistributions, and vibrational modes that operate across disparate length and time scales. Simple descriptor-based approaches often average these effects, missing critical emergent behaviors such as localized electronic states or diffusion pathways. Kim *et al.* show that even latent-space arithmetic for metamaterials benefits from underlying complex representations to generate diverse responses [17]. Third, in soft-magnetic or stability predictions for metallic glasses, Lu *et al.* demonstrate that thermal and magnetic responses depend on subtle short-to-medium range ordering; parsimonious models may fit global trends but fail to predict outliers or guide optimization where local complexity dominates [10].

Materials science thus demands appropriate complexity rather than its avoidance. Schmidt *et al.* review advances in solid-state machine learning, where expressive frameworks increasingly supplant simplistic baselines for accurate representation of crystal and molecular systems [5]. Zunger's inverse design paradigm requires navigating vast, high-dimensional spaces to identify target functionalities, implicitly rejecting the notion that parsimony suffices for discovery [7]. Chen *et al.*'s graph networks succeed by embracing relational complexity inherent to atomic connectivity [6].

The critique identifies that simplicity bias stems from a mismatch between model assumptions and material reality. While parsimony aids in well-behaved, low-variance regimes, many materials problems involve irreducible complexity arising from quantum-mechanical many-body effects, thermodynamic ensembles, or processing histories. Butler *et al.* acknowledge that machine learning for materials benefits from scalable representations, yet warn against underpowered models that cannot generalize across chemical spaces [4]. Vasudevan *et al.* call for balanced parsimony integrated with causality, recognizing that off-the-shelf simplicity often falls short [9].

Distinguishing legitimate simplification (e.g., effective coarse-graining grounded in physics) from problematic reduction (arbitrary truncation of relevant degrees of freedom) is crucial. Materials AI should adopt a complexity-matching principle: model expressivity scaled to the intrinsic demands of the phenomenon rather than an a priori preference for parsimony. This shift would better align algorithmic tools with the epistemic goals of uncovering mechanisms in inherently complex systems [3, 8, 11, 23-29].

Consequences of Simplicity Bias

The overvaluation of algorithmic simplicity in materials AI produces several interlocking consequences that hinder scientific progress.

Missed discoveries of complex phenomena. By favoring parsimonious models, researchers may overlook subtle emergent behaviors or rare compositional regimes where high-order interactions dominate. For instance, in high-entropy alloys or disordered systems, simple descriptors can miss novel phases or property extrema that graph-based or hierarchical models would surface [14, 16]. This narrows the discovery frontier, channeling effort toward incremental refinements of known simple patterns rather than genuine innovation.

Underestimation of uncertainty. Simple models often produce overconfident predictions by underrepresenting variance from unmodeled factors. In property prediction tasks involving metastable or multi-scale materials, this leads to misleading reliability assessments, as the model appears robust within its constrained space but fails when extrapolated. Zhong *et al.* and related discussions on explainable methods note how complexity-aware approaches better quantify epistemic uncertainty [8].

Premature model acceptance. Simplicity serves as a proxy for validation, where compact models gain acceptance based on elegance rather than rigorous testing against diverse conditions or higher-fidelity references. Oviedo *et al.* highlight that interpretability techniques must still ensure scientific validity beyond surface transparency [11]. This risks embedding flawed assumptions into downstream design pipelines.

Wasted research effort. Forcing complex materials problems into simplistic frameworks consumes resources on hyperparameter tuning, feature engineering, or post-hoc explanations that ultimately cannot compensate for representational inadequacy. Vasudevan *et al.* critique off-the-shelf approaches lacking sufficient capacity balanced with parsimony and causality [9]. Effort diverts from developing tailored, complexity-appropriate methods toward retrofitting ill-suited baselines.

These consequences compound, creating feedback loops that reinforce simplicity bias across the community [1-5]. While simplicity offers short-term pragmatic gains, its systematic prioritization erodes long-term epistemic robustness in materials discovery.

Alternative Approaches

To counter simplicity bias, materials AI should embrace frameworks that value appropriate complexity.

Complexity as needed—match model sophistication explicitly to problem demands through diagnostic assessments of intrinsic material complexity (e.g., via multi-scale analysis or information-theoretic measures) before model selection [6, 7]. Regularized complexity—employ penalties that discourage gratuitous parameters while permitting expressive capacity where data and physics justify it, extending Bayesian approaches like those of MacKay to materials-specific priors [2, 9]. Ensemble complexity—combine multiple simple base learners or modular components whose interactions capture emergent relations without monolithic black-box architectures, leveraging strengths while mitigating individual weaknesses [4, 11].

Structured complexity—design architectures with built-in domain symmetries, hierarchies, or causal graphs (e.g., physics-informed graph networks) that remain interpretable despite higher capacity. Chen *et al.* and Zhong *et al.* point toward such pathways [6, 8]. Complexity-aware evaluation—assess models not only on accuracy and parsimony but on metrics capturing mechanistic fidelity, uncertainty calibration, and discovery potential across regimes of varying material complexity [3, 5, 20].

These alternatives promote a nuanced epistemology that distinguishes productive from obstructive simplicity while harnessing complexity where materials ontology requires it. Implementation demands interdisciplinary collaboration between AI practitioners and domain experts to define “appropriate” thresholds contextually.

Table 2 converts the manuscript’s critique into a complexity-aware evaluation framework that reorients model selection from default parsimony toward scientifically adequate expressivity, calibration, and discovery potential.

Table 2. Complexity-aware evaluation framework for model choice in materials AI

Evaluation dimension	Simplicity-biased question	Complexity-aware question	What be eva
Phenomenon complexity	Can the task be forced into a compact representation?	What level of expressivity is required by the material system?	Deg nonlin emer hierarc relat deper
Predictive adequacy	Is the simple model acceptable on average error alone?	Does the model sustain performance across hard regimes and edge cases?	Er distril ou behav extrap sta
Mechanistic fidelity	Does the model look scientifically neat?	Do learned relations align with domain mechanisms or plausible structure-property logic?	Phy cohe relat faithfu a mech plaus
Interpretability quality	Is the model small or shallow?	Does the explanation map onto expert-understandable materials concepts?	Fea transp conc trace a expla val
Uncertainty calibration	Does the model provide a single confident answer?	Does it represent uncertainty arising from model and data limitations?	Calib confi reliabi epis unce
Discovery potential	Does the model reproduce known trends efficiently?	Can it reveal non-obvious regimes, interactions, or candidate spaces?	Nov sens reg expa a canc dive
Validation burden	Is the model easy to justify because it is simple?	Has the model been stress-tested at a level matching its scientific claims?	Cross-valid higher comp a experi corrob
Complexity control	How much complexity can be removed?	How can complexity be structured, regularized, or modularized productively?	Pri regular enser a struc archit
Scientific efficiency	Does simplicity save immediate computation time?	Does the chosen model reduce long-run epistemic waste?	Trac betv shor conve a down rese effic
Final selection principle	Prefer the simplest model by default	Prefer the least complex model that remains scientifically adequate	Integ judg acro dimer ab

Conclusion

This critical critique has examined the assumption that algorithmic simplicity constitutes an unqualified scientific virtue in materials AI design. Tracing its historical roots in Occam’s razor and Bayesian parsimony, surveying its prevalence in property prediction and descriptor choices, and dissecting four core problems—accuracy trade-offs, obscurant effects, interpretability conflation, and mismatch with material demands—reveals substantial epistemic costs. Consequences range from missed discoveries to distorted uncertainty quantification and inefficient workflows.

Rather than rejecting simplicity outright, the analysis distinguishes its legitimate regulative roles from overextensions that hinder understanding of complex, multi-scale, and emergent materials phenomena. Alternative approaches that embrace matched, regularized, ensemble, structured, and evaluated complexity offer a path forward.

Materials AI must move beyond default parsimony toward a mature valuation of appropriate complexity in service of deeper insight and more reliable discovery.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 05 Feb 2024 Revised: 16 Jun 2024 Accepted: 10 Jul 2024
Published online: 18 January 2025

References

Gross F. Occam’s razor in molecular and systems biology. Philos Sci. 2019;86(5):1134-45.

Hanin B, Zlokapa A. Bayesian interpolation with deep linear networks. Proc Natl Acad Sci U S A. 2023;120(23):e2301345120.

Hansen KB. The virtue of simplicity: On machine learning models in algorithmic trading. Big Data Soc. 2020;7(1):2053951720926558.

Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A. Machine learning for molecular and materials science. Nature. 2018;559(7715):547-55.

Schmidt J, Marques MR, Botti S, Marques MA. Recent advances and applications of machine learning in solid-state materials science. npj Comput Mater. 2019;5(1):83.

Chen C, Ye W, Zuo Y, Zheng C, Ong SP. Graph networks as a universal machine learning framework for molecules and crystals. Chem Mater. 2019;31(9):3564-72.

Zunger A. Inverse design in search of materials with target functionalities. Nat Rev Chem. 2018;2(4):0121.

Zhong X, Gallagher B, Liu S, Kailkhura B, Hiszpanski A, Han TY. Explainable machine learning in materials science. npj Comput Mater. 2022;8(1):204.

Vasudevan RK, Ziatdinov M, Vlcek L, Kalinin SV. Off-the-shelf deep learning is not enough, and requires parsimony, Bayesianity, and causality. npj Comput Mater. 2021;7(1):16.

Lu Z, Chen X, Liu X, Lin D, Wu Y, Zhang Y, et al. Interpretable machine-learning strategy for soft-magnetic property and thermal stability in Fe-based metallic glasses. npj Comput Mater. 2020;6(1):187.

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

Desai S, Strachan A. Parsimonious neural networks learn interpretable physical laws. Sci Rep. 2021;11(1):12761.

Gallegos M, Vassilev-Galindo V, Poltavsky I, Martín Pendás Á, Tkatchenko A. Explainable chemical artificial intelligence from accurate machine learning of real-space chemical descriptors. Nat Commun. 2024;15(1):4345.

Zhou Z, Zhou Y, He Q, Ding Z, Li F, Yang Y. Machine learning guided appraisal and exploration of phase design for high entropy alloys. npj Comput Mater. 2019;5(1):128.

Singh C, Askari A, Caruana R, Gao J. Augmenting interpretable models with large language models during training. Nat Commun. 2023;14(1):7913.

Moitzi F, Romaner L, Ruban AV, Hodapp M, Peil OE. Ab initio framework for deciphering trade-off relationships in multi-component alloys. npj Comput Mater. 2024;10(1):152.

Kim N, Lee D, Kim C, Lee D, Hong Y. Simple arithmetic operation in latent space can generate a novel three-dimensional graph metamaterials. npj Comput Mater. 2024;10(1):236.

La Cava WG, Lee PC, Ajmal I, Ding X, Solanki P, Cohen JB, et al. A flexible symbolic regression method for constructing interpretable clinical prediction models. NPJ Digit Med. 2023;6(1):107.

Korolev V, Protsenko P. Accurate, interpretable predictions of materials properties within transformer language models. Patterns. 2023;4(10):100803.

Bell A, Solano-Kamaiko I, Nov O, Stoyanovich J. It's just not that simple: An empirical study of the accuracy-explainability trade-off in machine learning for public policy. In: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency. New York, NY: ACM; 2022. pp. 248-66.

Dziugaite GK, Ben-David S, Roy DM. Enforcing interpretability and its statistical impacts: Trade-offs between accuracy and interpretability. arXiv preprint arXiv:2010.13764. 2020 Oct 26.

Linardatos P, Papastefanopoulos V, Kotsiantis S. Explainable AI: A review of machine learning interpretability methods. Entropy. 2020;23(1):18.

Sullivan E. Understanding from machine learning models. Br J Philos Sci. 2022;73(1):109-33.

Damewood J, Karaguesian J, Lunger JR, Tan AR, Xie M, Peng J, et al. Representations of materials for machine learning. Annu Rev Mater Res. 2023;53(1):399-426.

Liu Y, Jovanovic M, Mallayya K, Maddox WJ, Wilson AG, Klemenz S, et al. Materials expert-artificial intelligence for materials discovery. Commun Mater. 2025;6(1):212.

Belle V, Papantonis I. Principles and practice of explainable machine learning. Front Big Data. 2021;4:688969.

Bengio Y, Goodfellow I, Courville A. Deep learning. Cambridge, MA, USA: MIT Press; 2017.

Gubernatis JE, Lookman TJ. Machine learning in materials design and discovery: Examples from the present and suggestions for the future. Phys Rev Mater. 2018;2(12):120301.

Silveyra JM, Ferrara E, Huber DL, Monson TC. Soft magnetic materials for a sustainable and electrified world. Science. 2018;362(6413):eaao0195.