ORIGINAL RESEARCH **Open access**

A Theory of Justifiable Abstraction in Multi-Scale Materials AI

Oliver Grant[1], David Clark[1*], Sophia Nguyen[2]

Abstract

Multi-scale materials AI depends on abstraction as a necessary but inherently risky operation: researchers must simplify systems spanning electronic, atomic, microstructural, and macroscopic scales to achieve computational tractability, yet every simplification discards degrees of freedom, interactions, or information whose relevance cannot be known a priori. Abstraction, therefore, stands at the heart of every coarse-grained molecular-dynamics run, every surrogate model, and every continuum approximation, yet the epistemic costs of these choices remain largely unexamined. Without explicit justification, abstracted models risk producing predictions that appear accurate within narrow validation regimes while failing catastrophically when deployed on new tasks, new materials, or new operating conditions. This paper argues that abstraction cannot be taken for granted and instead requires a principled theory of justifiable abstraction. The proposed theory rests on three core principles—task-relative justification, information-preservation criterion, and multi-scale validation—supported by five operational criteria that together allow researchers to decide, for any given modeling context, whether an abstraction is defensible or whether higher-fidelity reference calculations must be retained. The framework further distinguishes four canonical types of abstraction (spatial, temporal, compositional, and physical) that appear repeatedly across the literature on multi-scale machine learning for materials. By making justification explicit and evaluable, the theory shifts multi-scale materials AI from an ad-hoc practice to a disciplined epistemic activity, ensuring that computational gains do not come at the expense of scientific reliability or technological trustworthiness. The implications extend beyond individual papers to the design of benchmarks, the standards of peer review, and the very architecture of future hierarchical modeling platforms.

Keywords Justifiable abstraction, Multi-scale materials AI, Coarse-graining, Scale-bridging, Information preservation, Hierarchical modeling

*Correspondence:
David Clark
david.clark@gmail.com
[1] Department of Materials Informatics and AI, University of Glasgow, Glasgow, United Kingdom
[2] Department of Smart Materials Systems, National University of Singapore, Singapore, Singapore

Introduction

Multi-scale materials AI has emerged as one of the most powerful paradigms for accelerating the discovery and design of advanced materials. By integrating machine-learning surrogates with physics-based models at multiple length and time scales, researchers can now explore vast compositional spaces, predict emergent properties, and optimize processing routes that would be inaccessible to purely first-principles or purely empirical approaches [1-7]. Yet every successful multi-scale workflow rests on a foundational, often invisible operation: abstraction. Whether through coarse-graining of atomistic degrees of freedom, homogenization of microstructural heterogeneities, or construction of surrogate models that replace expensive density-functional-theory calculations, abstraction is the mechanism that renders otherwise intractable problems computable. The literature already demonstrates the practical power of such abstractions; molecular-dynamics simulations must operate at multiple scales to capture realistic material behavior [1], while multi-scale machine learning can link quantum-mechanical descriptors to macroscopic performance metrics [2]. Similar themes

appear in the graph-network frameworks that abstract local atomic environments [7] and in the broader surveys of machine learning for materials science [4, 5].

Nevertheless, the very success of these methods conceals a deeper epistemic vulnerability. Abstraction is never lossless [8-15]. Every time a modeler neglects certain electronic correlations, ignores vibrational modes, or replaces discrete atomic interactions with continuum fields, information is irreversibly discarded. The question is not whether abstraction occurs—because it must—but whether the particular abstraction chosen is justifiable for the specific scientific or engineering task at hand. When is abstraction justified? When does it produce misleading results? These questions have received surprisingly little systematic attention in the materials AI community [3]. Most papers report impressive predictive accuracies on curated test sets without disclosing the scale at which critical information was sacrificed or the conditions under which that sacrifice becomes unacceptable [16-19]. The consequence is a growing gap between computational convenience and scientific rigor.

This paper addresses that gap by proposing a theory of justifiable abstraction tailored to multi-scale materials AI. The theory begins from the recognition that abstraction is not an absolute property of a model but a relational one: its validity depends on the task, the target property, the required accuracy, and the downstream decision context [3]. Building on earlier conceptual work in the field, the framework articulates explicit criteria that researchers can apply before, during, and after model construction. It further develops a typology of abstraction types that recur across the literature and analyzes the consequences that arise when abstractions are adopted without justification. The goal is not to discourage abstraction—indeed, abstraction remains indispensable—but to elevate it from an unexamined habit to a deliberate, auditable epistemic practice. By doing so, the theory aims to strengthen the reliability of multi-scale predictions, reduce the incidence of irreproducible claims, and provide clearer guidance for when higher-fidelity reference calculations are non-negotiable [1, 2]. In short, it seeks to make the invisible infrastructure of abstraction visible and accountable.

Abstraction in Multi-Scale Materials Science

Abstraction, in the context of multi-scale materials AI, can be formally defined as follows:

Abstraction is the deliberate process of simplifying a material system by neglecting, aggregating, or approximating selected degrees of freedom, interactions, or scales of description to render computation tractable while preserving the information required for a specified set of tasks.

Common abstractions in materials AI, therefore, fall into recognizable categories. Spatial abstraction reduces the resolution of atomic or microstructural detail; temporal abstraction coarsens time steps or eliminates fast vibrational modes; compositional abstraction replaces complex multi-element systems with pseudo-binary or proxy representations; and physical abstraction omits specific interaction types such as spin-orbit coupling or explicit phonon scattering. Each of these simplifications appears repeatedly in the surveyed literature. For instance, multi-scale machine-learning frameworks routinely abstract electronic-structure details into lower-dimensional descriptors [2], while graph networks abstract local atomic environments into message-passing representations that ignore long-range interactions beyond a cutoff radius [7]. Inverse-design philosophies similarly abstract target functionalities into searchable property spaces, accepting the loss of microscopic detail in exchange for accelerated discovery [6].

What unites these examples is the recognition that abstraction is not optional. The enormous disparity between electronic-structure time and length scales and those relevant to macroscopic engineering performance makes direct all-scale simulation impossible even with exascale computing [1]. Abstraction is, therefore, the pragmatic bridge that allows materials AI to operate across scales. Yet the literature also reveals that abstraction is rarely accompanied by an explicit justification protocol [3]. Most studies validate the final model against experimental or higher-fidelity data for a narrow set of properties without documenting which information was deliberately discarded or why that discard was deemed acceptable [17]. This omission is the central motivation for the theory developed in the subsequent sections.

Why Abstraction Needs Justification

Abstraction is often treated as a routine methodological step in multi-scale materials modeling, yet such treatment obscures the epistemic risks it introduces. At its core, abstraction entails the systematic removal of degrees of freedom, and with this removal comes an unavoidable loss of information that may later prove consequential for the property under investigation [10]. What appears negligible at one scale can re-emerge through non-linear interactions at another, undermining the stability of the abstraction when applied beyond its immediate context. Evidence from coarse-graining studies in polymers illustrates this tension, where subtle atomistic features—initially deemed irrelevant—exert measurable influence on macroscopic observables such as glass-transition temperatures and mechanical response once the model is extended beyond its calibration regime [10].

This vulnerability is further intensified by the presence of scale coupling, which complicates any attempt to localize relevance within a single level of description. Material properties frequently depend on interactions that traverse scales in ways that are neither linear nor intuitively predictable [17]. An abstraction that appears well-justified at the atomic level may therefore fail when confronted with microstructural phenomena, where neglected fine-scale variations propagate upward in amplified or transformed form. Machine-learning studies of nanoconfinement in porous systems make this dependence explicit, demonstrating how microscopic fluctuations can induce macroscopic transport behaviors that are not recoverable from simplified representations alone [17].

The challenge becomes more acute when considering emergent phenomena, whose defining characteristic is that they cannot be reduced to isolated components without losing their explanatory structure. Many macroscopic behaviors in materials arise from collective interactions and correlations that are inherently distributed across scales, making them particularly susceptible to erasure under aggressive abstraction [11]. Coarse-grained machine-learning models provide a revealing example: their ability to capture stable crystal structures depends critically on whether the abstraction retains sufficient information to encode the ordering processes that give rise to phase transitions. When this condition is not met, the model may perform adequately on limited tasks while failing to detect the very phenomena that define the system's behavior [11].

A further complication arises from the asymmetry between validation and generalization. An abstracted model may exhibit strong agreement with reference data for a specific task, creating the impression of adequacy, yet this apparent success often masks a deeper fragility. When applied to new conditions or alternative queries, the same model may diverge significantly, revealing that its validity was contingent rather than robust [18]. Multi-fidelity modeling approaches highlight this gap, showing how surrogate models trained on one level of fidelity can deliver accurate short-term predictions while deteriorating as soon as the domain of application shifts [18]. This discrepancy underscores the difficulty of inferring epistemic reliability from localized validation alone.

Taken together, these considerations converge on a central conclusion: abstraction is not a neutral simplification but an epistemically consequential operation that must be justified rather than assumed [4]. Without explicit criteria for evaluating what has been removed and what has been preserved, it becomes impossible to determine whether the resulting model retains the causal structure necessary for reliable prediction. The remainder of this analysis, therefore, turns toward the development of a theoretical framework capable of rendering such justification systematic, transparent, and reproducible [3].

A Theory of Justifiable Abstraction

The theory of justifiable abstraction developed here reframes abstraction as an explicit epistemic commitment, governed by principles that allow it to be evaluated rather than taken for granted. Central to this perspective is the recognition that abstraction acquires meaning only in relation to the task it is intended to support. There is no context-independent standard against which an abstraction can be judged; its adequacy is determined by whether it preserves the structures necessary to answer a particular class of questions [3]. A representation that is sufficient for predicting elastic moduli, for instance, may be entirely inadequate for capturing fracture behavior or optical responses, because the relevant mechanisms differ fundamentally across these domains. This task-relativity aligns with established multi-scale machine-learning strategies, where representational choices are explicitly conditioned on the target property of interest [5].

Building on this foundation, the theory introduces a criterion of information preservation that shifts the focus from aggregate accuracy to the retention of causally relevant

structure. An abstraction is defensible only if it maintains the information that is demonstrably necessary and sufficient for the task at hand [9]. This requirement cannot be reduced to conventional performance metrics, as high predictive accuracy may still be achieved through correlations that bypass underlying mechanisms. Instead, justification demands that the retained features encode the pathways through which microscopic variables give rise to macroscopic observables. The energy-renormalization framework for epoxy systems illustrates this principle in practice, where the validity of the coarse-grained model depends on its ability to preserve the effective free-energy landscape governing mechanical response [9].

The final component of the theory addresses the limitations of single-scale validation by requiring that abstraction be assessed across all relevant levels of description. Validation confined to the output scale provides only partial assurance, as it may conceal inconsistencies that emerge when the model is interrogated at higher or intermediate fidelities. A robust justification, therefore, entails systematic comparison with higher-fidelity reference calculations across scales, ensuring that the abstraction remains coherent within the broader hierarchical structure of the system [1, 2]. This requirement aligns with the principles of hierarchical modeling articulated in the literature, where consistency across scales is treated as a prerequisite for reliability [16, 20-22].

Taken together, these principles establish a coherent framework in which abstraction is considered justifiable only when it is explicitly aligned with a defined task, demonstrably preserves the information required to support that task, and withstands validation across the scales through which the system is constituted. Rather than prescribing a single optimal representation, the theory provides a structured basis for evaluating alternative abstractions, enabling them to be defended or rejected through transparent and reproducible criteria [3].

Table 1 maps the three core principles of justifiable abstraction onto the five operational criteria, clarifying their functional interdependencies."

Table 1. Mapping core principles to operational criteria in justifiable abstraction

Core principle	Associated criteria	Functional role in justification	Failure risk if violated
Task-relative justification	Task alignment; transferability	Aligns abstraction with the prediction objective	Misaligned predictions; context-specific failure
Information preservation	Information sufficiency; scale separation	Ensures causal features are retained	Loss of emergent behavior; biased outputs
Multi-scale validation	Error bounds; transferability	Verifies robustness across scales and conditions	Hidden errors; failure under distribution shift
Cross-principle interaction	All five criteria jointly	Integrates epistemic validity across the modeling pipeline	False confidence; irreproducible findings

Criteria for Justifiable Abstraction

The theoretical framework becomes practically meaningful only when translated into criteria that can guide evaluation in concrete modeling settings. In this respect, justifiable abstraction depends on the extent to which scale separation is genuinely achieved rather than merely assumed. Many abstractions rely on the premise that interactions across scales can be decoupled, yet in materials systems, such independence is often only approximate. Residual couplings may persist in subtle forms, and when these are neglected, the abstraction introduces systematic distortions that are difficult to detect within limited validation regimes. Evidence from nanoconfinement studies illustrates how continuum-level simplifications can obscure atomistic fluctuations that continue to influence macroscopic transport behavior, revealing that apparent scale separation may conceal rather than eliminate cross-scale dependencies [17].

Closely related to this concern is the requirement that the abstracted representation retains sufficient information to

support the target task. This condition cannot be reduced to overall predictive accuracy, as high performance may arise from correlations that do not faithfully encode the underlying structure of the system. Instead, the relevant question is whether the abstraction preserves the information necessary to reconstruct the property of interest within an acceptable tolerance [10]. From an information-theoretic perspective, this can be examined through the mutual information shared between the full and reduced representations with respect to the task variable, providing a more precise account of what has been retained and what has been lost.

The issue of information retention naturally leads to the question of quantifying the consequences of what has been discarded. Any abstraction introduces error, but its epistemic acceptability depends on whether this error can be estimated and bounded. Without such bounds, validation remains contingent and case-specific, offering little assurance of reliability beyond the immediate dataset. Establishing analytical approximations or employing statistical resampling techniques allows the uncertainty introduced by abstraction to be characterized explicitly, transforming error from an implicit byproduct into a measurable quantity that can inform model selection and interpretation [14].

Beyond these internal considerations, justification also depends on the alignment between the abstraction and the specific task it is intended to support. Representations that perform well for one class of properties may fail when applied to others, particularly when the underlying mechanisms differ in structure or scale. Coarse-grained models optimized for equilibrium thermodynamic quantities, for example, may prove inadequate for capturing non-equilibrium transport phenomena, as demonstrated across multiple studies [8, 12]. This dependency underscores that abstraction is not universally valid but context-sensitive, requiring evaluation relative to the predictive or decision context in which it is deployed [19].

A final dimension of justification emerges when the model is exposed to conditions beyond those encountered during its construction. Transferability under distribution shift—whether in composition, temperature, or loading regime—serves as a critical test of whether the abstraction has preserved the structural features necessary for generalization. Models that appear robust within a narrow domain may fail when confronted with new regimes, revealing that their validity was contingent on the original distribution. Incorporating transferability assessment into the justification process, therefore, ensures that abstraction is not only locally adequate but remains defensible under the broader range of conditions characteristic of real-world materials systems [21].

Figure 1 presents a hierarchical decision architecture in which task-specific queries condition abstraction through core principles and operational criteria, culminating in an explicit justification outcome.

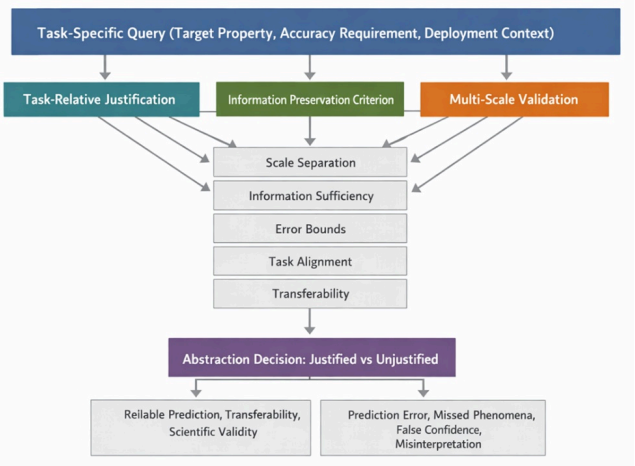

Figure 1. A hierarchical decision architecture in which task-specific queries condition abstraction through core principles and operational criteria.

Collectively, the criteria convert the abstract theory into a practical checklist that can be embedded in modeling workflows, reported in publications, and scrutinized during peer review [3, 5].

Types of Abstraction In Materials AI

A typology of abstraction in multi-scale materials AI clarifies the distinct mechanisms through which simplification occurs and thereby makes the justification process more precise. The typology distinguishes four primary types—spatial, temporal, compositional, and physical—each of which appears repeatedly in the literature yet carries its own epistemic signature. By classifying abstractions in this manner, the theory proposed here equips researchers with a diagnostic vocabulary that links specific simplification choices to the criteria of justifiable abstraction developed earlier.

Spatial abstraction

Spatial abstraction consists of reducing the resolution of atomic or microstructural detail by replacing discrete entities with averaged or homogenized representations. Coarse-graining of atomistic degrees of freedom into effective beads, homogenization of microstructural heterogeneities into continuum fields, and the use of cutoff radii in graph-network representations all exemplify this type. Machine-learning-informed energy renormalization for epoxy resins [9] and systematic coarse-graining of model polymers [8] illustrate how spatial abstraction enables simulations over length scales that would otherwise remain inaccessible. Spatial abstraction is justified when scale separation is demonstrably clean [17] and when the discarded spatial fluctuations do not contribute to the target macroscopic property, as in the prediction of bulk elastic moduli under uniform loading. It becomes risky, however, when emergent phenomena such as dislocation patterning or grain-boundary sliding depend on the very spatial correlations that have been averaged away [11], leading to systematic underestimation of yield strength or fracture toughness.

Temporal abstraction

Temporal abstraction coarsens the time domain by enlarging integration steps, eliminating fast vibrational modes, or imposing steady-state assumptions that ignore transient dynamics. Time-step coarsening in molecular-dynamics trajectories and the replacement of explicit phonon scattering with effective damping terms fall into this category. The machine-learning-enabled coarse-grained models for epoxies over wide temperature ranges [12] rely on temporal abstraction to reach experimentally relevant timescales. This type is justified when the target task concerns equilibrium thermodynamics or long-time relaxation, where fast modes contribute only to averaged noise rather than to the observable of interest [13]. Temporal abstraction turns risky whenever non-equilibrium processes—such as shock-wave propagation or rate-dependent plasticity—depend on the precise sequencing of fast events that have been removed [14], producing predictions that diverge sharply once the model is confronted with dynamic loading conditions.

Compositional abstraction

Compositional abstraction replaces multi-element systems with pseudo-binary or proxy representations, ignoring dilute dopants, trace impurities, or specific elemental interactions. Elemental proxies in high-entropy alloys and the reduction of complex solid solutions to effective medium approximations belong here. Inverse-design strategies that abstract target functionalities into searchable property spaces routinely employ compositional abstraction [6], as do graph-network frameworks that map local atomic environments without enumerating every possible coordination [7]. Compositional abstraction is justified when the discarded compositional degrees of freedom exert only perturbative effects on the primary property landscape [5]. It becomes risky, however, in systems where emergent electronic or magnetic states hinge on precise stoichiometry, as in the prediction of band gaps or Curie temperatures, where even small compositional deviations can flip phase stability [2].

Physical abstraction

Physical abstraction selectively neglects entire classes of interactions—such as spin-orbit coupling, explicit electron-phonon scattering, or relativistic corrections—while retaining the remaining physics. Surrogate models that omit many-body terms or continuum approximations that ignore quantum-mechanical details exemplify this type. The multi-scale machine-learning pipelines surveyed in the literature frequently abstract away selected physical mechanisms to achieve tractability [15, 19]. Physical abstraction is justified when the neglected interactions lie far outside the energy window relevant to the task, such as when predicting room-temperature mechanical properties without resolving cryogenic magnetic ordering. It is risky whenever the omitted physics couples back into the retained degrees of freedom through higher-order effects, producing unphysical artifacts in optical, thermal, or transport predictions [4].

Across all four types, the typology reveals that abstraction is not monolithic. Each type maps onto different subsets of the five justification criteria, enabling researchers to perform targeted audits. For instance, spatial and compositional abstractions are most sensitive to scale separation and information sufficiency, while temporal and physical abstractions more directly challenge error bounds and transferability. By making these distinctions explicit, the typology transforms an otherwise opaque modeling choice into a structured epistemic decision space.

Table 2 systematizes the four abstraction types by linking each to its dominant justification criteria and characteristic failure modes.

Table 2. Typology of abstraction types and their epistemic risk profiles

Abstraction type	Mechanism of simplification	Primary criteria sensitivity	Typical justificatio condition
Spatial	Reduction of spatial resolution	Scale separation; information sufficiency	Clean separation of micro/macr scales
Temporal	Time-step coarsening; removal of fast dynamics	Error bounds; transferability	Equilibrium or long-tim behavior dominance
Compositional	Simplified element representation	Information sufficiency; task alignment	Weak influence o omitted species
Physical	Omission of interaction classes	Error bounds; transferability	Negligible contributio of omitted physics

Consequences of Unjustified Abstraction

When abstraction is implemented without explicit justification, its consequences extend beyond localized modeling errors and begin to erode the broader reliability of multi-scale materials AI. The first indication of this erosion typically appears as a divergence between abstracted predictions and higher-fidelity references, particularly for properties that depend on information eliminated during simplification. Within narrowly defined validation regimes, such discrepancies may remain concealed, giving the impression of adequacy. However, once the model is applied to new compositions or operating conditions, these hidden deficiencies surface, often amplifying in non-linear ways as the neglected degrees of freedom become consequential [10, 19]. What initially appears as a minor approximation error thus evolves into a systematic limitation on predictive scope.

This degradation in predictive reliability is accompanied by a more profound loss at the level of scientific representation. Abstraction does not merely approximate phenomena; it can remove the very scale at which those phenomena originate. When fine-grained correlations are excluded, emergent behaviors—such as phase transitions, defect self-organization, or collective excitations—may no longer be expressible within the model. The result is not simply reduced accuracy but an altered ontology, in which key aspects of material behavior cease to exist within the representational framework. Studies on coarse-grained systems demonstrate that such omissions are not incidental but structural, reflecting the inability of oversimplified models to capture correlations that are essential to emergent ordering [11, 17].

A further complication arises from the way validation practices interact with these omissions. When assessment is confined to the final output scale, models can appear robust despite having lost critical causal structure. Agreement with experimental or high-fidelity data under a specific set of conditions may be interpreted as evidence of general validity, even though the underlying representation lacks the capacity to support related queries. Multi-fidelity studies reveal how surrogate models can achieve strong performance within their training domain while failing abruptly when asked to address adjacent properties or slightly altered conditions [18]. This creates a form of false confidence, in which apparent success obscures deeper epistemic fragility.

Over time, these effects accumulate into a more subtle but consequential distortion of scientific understanding. When models systematically omit relevant scales or interactions, the explanations derived from them risk becoming artifacts of the abstraction itself rather than faithful accounts of material reality. Researchers may infer causal mechanisms that are consistent within the reduced representation yet incorrect in the full system, thereby embedding error into subsequent cycles of theory development and materials design. Such misattributions are particularly problematic because they propagate forward, shaping experimental priorities and guiding future modeling efforts in directions that reflect the limitations of the abstraction rather than the structure of the underlying phenomena [4, 15].

Table 3 links the major consequences of unjustified abstraction to specific violated criteria, providing a diagnostic structure for prevention.

Table 3. Consequences of unjustified abstraction and corresponding preventive criteria

Consequence	Root cause (unmet criterion)	Manifestation in models	Prever criter
Prediction error	Information sufficiency failure	Deviations under extrapolation	Informa sufficie error bo
Missed phenomena	Scale separation failure	Absence of emergent behavior	Sca separa multi-s valida
False confidence	Incomplete validation	Overfitting to narrow validation regimes	Multi-s validat transfer
Scientific misunderstanding	Task misalignment	Incorrect causal inference	Tas alignm informa preserv

These consequences are not hypothetical; they follow directly from the four reasons why abstraction requires justification. Without task-relative evaluation, information-preservation checks, and multi-scale validation, the epistemic costs remain invisible until downstream applications expose them. The theory of justifiable abstraction, therefore, treats these consequences as preventable outcomes of methodological laxity rather than inevitable features of multi-scale modeling.

Relation to Existing Concepts

The theory of justifiable abstraction does not exist in isolation; it articulates a missing epistemic layer that connects and sharpens several established concepts in computational materials science. Model reduction, for example, supplies systematic mathematical procedures for simplifying governing equations, yet it rarely asks whether the reduced model remains valid for every downstream task [23-29]. The proposed theory supplies precisely that missing question by demanding task-relative justification and information-preservation checks before any reduction is accepted.

Surrogate modeling approximates expensive simulations with fast statistical emulators and has become ubiquitous in materials informatics [15, 16]. While surrogate techniques excel at computational acceleration, they inherit the same information-loss vulnerabilities outlined earlier. The theory reframes surrogate construction as an instance of justifiable abstraction, requiring explicit error bounds and transferability tests rather than relying on cross-validation scores alone.

Multi-fidelity methods combine models of varying accuracy to balance cost and precision [19, 22]. These approaches already acknowledge scale hierarchies, yet they seldom formalize when a low-fidelity abstraction is permissible versus when higher-fidelity references must be retained. The five criteria developed here provide the decision rules that multi-fidelity workflows have previously lacked, turning ad-hoc fidelity switching into a principled, auditable process.

Finally, renormalization-group techniques from statistical physics formalize scale transformations and information flow across scales [1]. The theory of justifiable abstraction extends this physical intuition into the machine-learning era by insisting that every renormalization step—whether spatial, temporal, compositional, or physical—must be justified relative to the scientific query at hand. In this way, the framework unifies classical physics-based scale-bridging with modern data-driven methods under a single epistemic standard.

By relating justifiable abstraction to these concepts, the theory does not replace them but supplies the normative scaffold that renders their application more robust and reproducible.

Implications for Materials AI Practice

Adoption of the theory of justifiable abstraction requires concrete changes across three stakeholder groups.

For authors, three practices become mandatory: first, every manuscript must explicitly state the abstraction type or types employed and map them onto the typology presented here; second, justification must be provided relative to the specific task using the five criteria; third, validation results must be reported at multiple scales rather than at the final output alone. These requirements elevate abstraction from

an implicit modeling habit to a transparent epistemic commitment [3, 5].

For reviewers, the checklist derived from the criteria offers a practical evaluation tool: reviewers should ask whether scale separation was verified, whether information sufficiency was demonstrated, and whether error bounds and transferability were addressed. When any criterion is left unexamined, the manuscript should be returned for revision rather than accepted on the basis of predictive accuracy metrics alone.

For the broader community, two initiatives are essential: the development of standardized benchmarks that test abstraction quality across the four types and the establishment of reporting standards that require authors to include a "justification statement" in supplementary materials. Such benchmarks and standards will accelerate the maturation of multi-scale materials AI by making epistemic rigor as visible as predictive performance [2, 18].

Collectively, these practice-level changes ensure that computational gains are accompanied by commensurate gains in scientific trustworthiness. The theory thereby transforms multi-scale materials AI from a collection of powerful but opaque techniques into a disciplined, auditable discipline.

Conclusion

The theory of justifiable abstraction proposed in this paper reframes abstraction as a necessary but epistemically risky operation that must be defended rather than assumed. By grounding justification in three core principles—task-relative justification, information preservation, and multi-scale validation—and by operationalizing those principles through five explicit criteria, the framework supplies materials to AI researchers with a coherent standard for deciding when abstraction is appropriate and when higher-fidelity reference calculations remain indispensable. The accompanying typology of spatial, temporal, compositional, and physical abstractions further equips the community with a diagnostic vocabulary that links modeling choices to their epistemic consequences.

Unjustified abstraction produces prediction errors, missed phenomena, false confidence, and scientific misunderstanding; the theory offers a preventive remedy. Its implications extend from individual workflows to community-wide standards of peer review and benchmarking. As multi-scale materials AI continues to expand its reach, justifiable abstraction must become a non-negotiable requirement rather than an optional afterthought. Only by making the invisible infrastructure of abstraction visible and accountable can the field ensure that computational power translates into reliable scientific insight and technologically trustworthy materials design.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 21 Sep 2022 Revised: 08 Dec 2022 Accepted: 19 Jan 2023
Published online: 18 July 2023

References

Hollingsworth SA, Dror RO. Molecular dynamics simulation for all. Neuron. 2018;99(6):1129-43.

Lei H, Xie P, Zhang L. Machine learning-assisted multi-scale modeling. J Math Phys. 2023;64(7).

Bonaccorso G. Machine learning algorithms: Popular algorithms for data science and machine learning. Birmingham: Packt Publishing; 2018.

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.

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

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.

Shireen Z, Weeratunge H, Menzel A, Phillips AW, Larson RG, Smith-Miles K, et al. A machine learning enabled hybrid optimization framework for efficient coarse-graining of a model polymer. npj Comput Mater. 2022;8(1):224.

Giuntoli A, Hansoge NK, van Beek A, Meng Z, Chen W, Keten S. Systematic coarse-graining of epoxy resins with machine learning-informed energy renormalization. npj Comput Mater. 2021;7(1):168.

Ricci E, Vergadou N. Integrating machine learning in the coarse-grained molecular simulation of polymers. J Phys Chem B. 2023;127(11):2302-22.

Goodall RE, Parackal AS, Faber FA, Armiento R, Lee AA. Rapid discovery of stable materials by coordinate-free coarse graining. Sci Adv. 2022;8(30):eabn4117.

Duan K, He Y, Li Y, Liu J, Zhang J, Hu Y, et al. Machine-learning assisted coarse-grained model for epoxies over wide ranges of temperatures and cross-linking degrees. Mater Des. 2019;183:108130.

Chan H, Cherukara MJ, Narayanan B, Loeffler TD, Benmore C, Gray SK, et al. Machine learning coarse grained models for water. Nat Commun. 2019;10(1):379.

Swinburne TD. Coarse-graining and forecasting atomic material simulations with descriptors. Phys Rev Lett. 2023;131(23):236101.

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

Nyshadham C, Rupp M, Bekker B, Shapeev AV, Mueller T, Rosenbrock CW, et al. Machine-learned multi-system surrogate models for materials prediction. npj Comput Mater. 2019;5(1):51.

Lubbers N, Agarwal A, Chen Y, Son S, Mehana M, Kang Q, et al. Modeling and scale-bridging using machine learning: Nanoconfinement effects in porous media. Sci Rep. 2020;10(1):13312.

Yang W, Peng L, Zhu Y, Hong L. When machine learning meets multiscale modeling in chemical reactions. J Chem Phys. 2020;153(9).

Batra R, Sankaranarayanan S. Machine learning for multi-fidelity scale bridging and dynamical simulations of materials. J Phys Mater. 2020;3(3):031002.

Mianroodi JR, Rezaei S, Siboni NH, Xu BX, Raabe D. Lossless multi-scale constitutive elastic relations with artificial intelligence. npj Comput Mater. 2022;8(1):67.

Joglekar AS, Thomas AG. Machine learning of hidden variables in multiscale fluid simulation. Mach Learn Sci Technol. 2023;4(3):035049.

Ingólfsson HI, Bhatia H, Aydin F, Oppelstrup T, López CA, Stanton LG, et al. Machine learning-driven multiscale modeling: Bridging the scales with a next-generation simulation infrastructure. J Chem Theory Comput. 2023;19(9):2658-75.

Kovachki N, Liu B, Sun X, Zhou H, Bhattacharya K, Ortiz M, et al. Multiscale modeling of materials: Computing, data science, uncertainty and goal-oriented optimization. Mech Mater. 2022;165:104156.

Alber M, Buganza Tepole A, Cannon WR, De S, Dura-Bernal S, Garikipati K, et al. Integrating machine learning and multiscale modeling—perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences. NPJ Digit Med. 2019;2(1):115.

Peng GCY, Alber M, Tepole AB, Cannon WR, De S, Dura-Bernal S, et al. Multiscale modeling meets machine learning: What can we learn? Arch Comput Methods Eng.

2021;28(3):1017-37. https://doi.org/10.1007/s11831-020-09405-5.

Yi W, Xiao P, Liu X, Zhao Z, Sun X, Wang J, et al. Recent advances in developing active targeting and multi-functional drug delivery systems via bioorthogonal chemistry. Signal Transduct Target Ther. 2022;7(1):386.

de Haan P, Cohen TS, Welling M. Natural graph networks. In: NeurIPS. Red Hook, NY: Curran Associates; 2020. p. 3636-46.

Peng J, Schwalbe-Koda D, Akkiraju K, Xie T, Giordano L, Yu Y, et al. Human-and machine-centred designs of molecules and materials for sustainability and decarbonization. Nat Rev Mater. 2022;7(12):991-1009.

West AR. Solid state chemistry and its applications. Hoboken, NJ: John Wiley & Sons; 2022.