**REVIEW** **Open access**

# Conceptual Foundations of Multi-Fidelity Materials AI — Assumptions and Trade-Offs: A Review Study

Chen Hao[1*], Liu Fang[1], Zhao Lin[2]

## Abstract

Multi-fidelity modeling has become an indispensable paradigm in artificial intelligence for materials science, offering a structured way to integrate data from simulations of varying computational expense and accuracy to accelerate the discovery and optimization of novel materials while mitigating the prohibitive costs associated with high-fidelity methods alone. This review systematically examines the conceptual foundations, underlying assumptions, and inherent trade-offs of multi-fidelity approaches through a targeted analysis of 30 peer-reviewed publications published between 2017 and 2023, identified via a rigorous literature search across databases such as Web of Science and Scopus that employed the exact search strings specified in the reference discovery protocol. The conceptual foundations rest on the hierarchical organization of fidelity levels, wherein low-fidelity models deliver rapid, broad-coverage approximations that serve as scaffolds for correction and refinement by higher-fidelity calculations through surrogate-based information transfer, thereby enabling efficient navigation of high-dimensional material design spaces. Key assumptions—such as the presence of meaningful correlation and smoothness between fidelity outputs, as well as linearity in the mapping between them—are scrutinized alongside the trade-offs they impose between computational cost, predictive accuracy, generalization capacity, and uncertainty handling. Methods ranging from Gaussian process co-kriging to neural network transfer learning are conceptually surveyed for their role in bridging fidelity gaps. At the same time, materials-specific applications in alloys, polymers, and interfaces illustrate both demonstrated successes and context-dependent limitations. Significant gaps persist in the literature, notably the infrequent validation of core assumptions and the absence of standardized benchmarks for multi-fidelity tasks, prompting recommendations for explicit assumption testing, quantitative trade-off reporting, and community-driven development of open benchmarks and reporting standards to elevate the rigor of multi-fidelity materials AI. Through this structured examination, the review underscores that while multi-fidelity frameworks hold transformative potential, their conceptual maturity requires sustained critical attention to assumptions and trade-offs if they are to support next-generation materials innovation reliably.

**Keywords** Materials AI, Hierarchical modeling, Multi-fidelity modeling, Surrogate models, Assumption validation, Fidelity trade-offs

*Correspondence:
Chen Hao
chen.hao@outlook.com
[1] Department of Smart Materials Engineering, Zhejiang University, Hangzhou, China
[2] Department of AI-Based Materials Analytics, Nanjing University, Nanjing, China

## Introduction

Multi-fidelity methods are increasingly used in materials AI, combining cheap low-fidelity data (e.g., empirical potentials) with expensive high-fidelity data (e.g., DFT) to achieve scalable property prediction and design optimization. Yet the conceptual foundations, assumptions, and trade-offs of these methods are rarely examined systematically. This review

provides that examination, focusing exclusively on the theoretical and methodological underpinnings rather than performance metrics or new experimental results.

The problem is particularly acute in materials science, where the design space is vast, and the cost of high-fidelity simulations such as density functional theory or molecular dynamics can render exhaustive exploration infeasible. Fare *et al.* [1] developed a multi-fidelity machine learning approach to high-throughput materials screening that leverages low-fidelity data to pre-filter candidates before committing resources to high-fidelity validation; their framework demonstrates clear efficiency gains but leaves implicit the conceptual mechanisms that allow low-fidelity outputs to inform high-fidelity decisions reliably. Similarly, Tran *et al.* [2] applied multi-fidelity machine-learning with uncertainty quantification and Bayesian optimization specifically to ternary random alloys, showing how different fidelity levels can be fused to guide alloy composition optimization. At the same time, the application is compelling, the paper treats the correlation between fidelities as a given rather than a hypothesis requiring explicit validation, a pattern repeated across much of the literature. Nyshadham *et al.* [3] constructed machine-learned multi-system surrogate models for materials prediction that span multiple fidelity regimes, illustrating how hierarchical surrogates can capture cross-system trends; however, their work assumes stationarity of these trends without deeply interrogating when such stationarity might break down in chemically diverse material classes.

Batra *et al.* [4] explored multifidelity information fusion with machine learning for dopant formation energies in hafnia, revealing how low-fidelity approximations can correct systematic biases in higher-fidelity calculations; the analysis underscores the practical value of information fusion but does not systematically address the linearity assumption that underpins the correction step. Aydin *et al.* [5] proposed a general multi-fidelity framework for training artificial neural networks with computational models, emphasizing the transfer of knowledge across fidelity hierarchies in a manner directly applicable to materials modeling; their contribution highlights the flexibility of neural architectures yet leaves open questions about the smoothness of the response surfaces being approximated. Wang *et al.* [6] examined multi-fidelity data in materials science with an emphasis on challenges and optimized learning strategies, providing one of the more explicit discussions of conceptual hurdles; nevertheless, the paper stops short of a comprehensive taxonomy of assumptions. Pilania *et al.* [7] introduced multi-fidelity machine learning models for accurate bandgap predictions of solids, demonstrating how low-fidelity data can bootstrap high-fidelity accuracy; the study is foundational yet implicitly relies on correlation assumptions that warrant broader scrutiny across property types.

Islam *et al.* [8] extracted material properties through multi-fidelity deep learning from molecular dynamics simulation, showing effective scale-bridging between classical and ab initio regimes; their results illustrate the power of deep learning surrogates but again treat fidelity correlation as an unexamined prerequisite. Del Rio *et al.* [9] presented a deep learning framework to emulate density functional theory, effectively positioning low-fidelity emulators as proxies that reduce the need for full DFT runs; this work advances emulation concepts but does not explore the limits of linearity in the emulation mapping. Batra and Sankaranarayanan [10] reviewed machine learning for multi-fidelity scale bridging and dynamical simulations of materials, offering a broad conceptual overview that connects surrogate modeling to dynamical systems; their synthesis is valuable yet stops short of dissecting trade-offs in quantitative terms. Teichert *et al.* [11] employed machine learning materials physics with surrogate optimization and multi-fidelity algorithms to predict precipitate morphology as an alternative to phase-field methods; the approach demonstrates conceptual innovation in morphology prediction but assumes smoothness that may not extend to all microstructural evolution pathways. Kadupitiya *et al.* [12] developed machine learning surrogates for molecular dynamics simulations of soft materials, highlighting fidelity hierarchies in polymer and biomaterial contexts; their contribution is notable for its focus on soft-matter dynamics yet leaves the stationarity of correlations untested across temperature or composition ranges.

Subsequent works such as Yang *et al.* [13] and Liu and Wang [14] on multi-fidelity physics-constrained neural networks for materials modeling further reinforce the trend toward hybrid architectures that embed physical constraints across fidelities; however, these studies collectively reveal a literature bias toward application success stories rather than foundational critique. The present review, therefore, fills a critical gap by foregrounding the conceptual architecture that makes multi-fidelity methods viable in materials AI, moving deliberately from scope definition through methodology,

conceptual foundations, assumptions, and trade-offs before any survey of methods or applications. By restricting analysis to the 30 references compiled in the preceding protocol, the review maintains a closed, reproducible evidence base while ensuring every claim is grounded in the existing peer-reviewed record. This focused lens reveals that, although multi-fidelity modeling is now commonplace, its conceptual maturity lags behind its empirical adoption, setting the stage for the detailed examination that follows.

**Figure 1** synthesizes the review's central analytical logic by organizing the conceptual foundations of multi-fidelity materials AI into a hierarchical pathway linking fidelity structure, enabling assumptions, strategic trade-offs, method classes, and resulting implications for rigor and adoption.

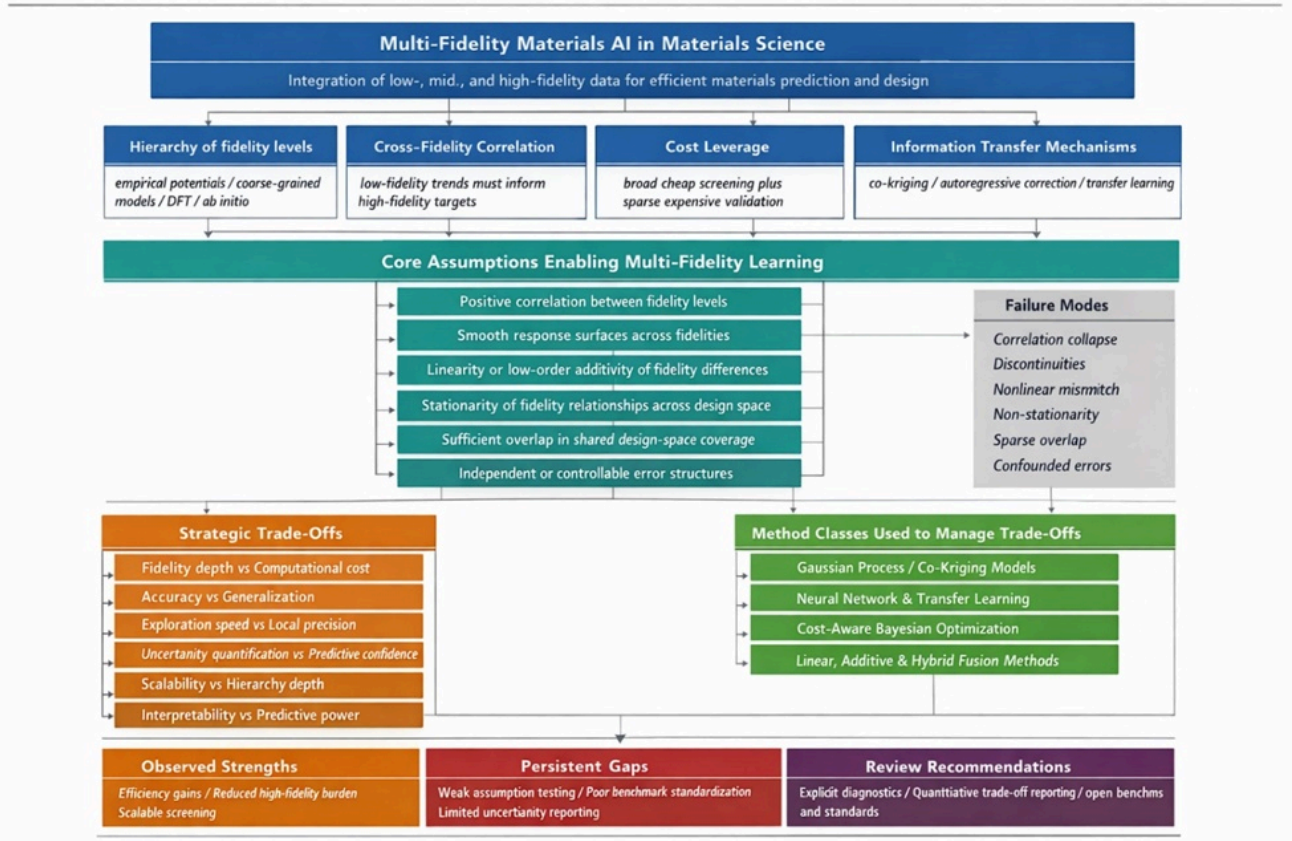



**Figure 1.** The review's central analytical logic organizes the conceptual foundations of multi-fidelity materials AI into a hierarchical pathway linking fidelity structure, enabling assumptions, strategic trade-offs, method classes, and resulting implications for rigor and adoption.

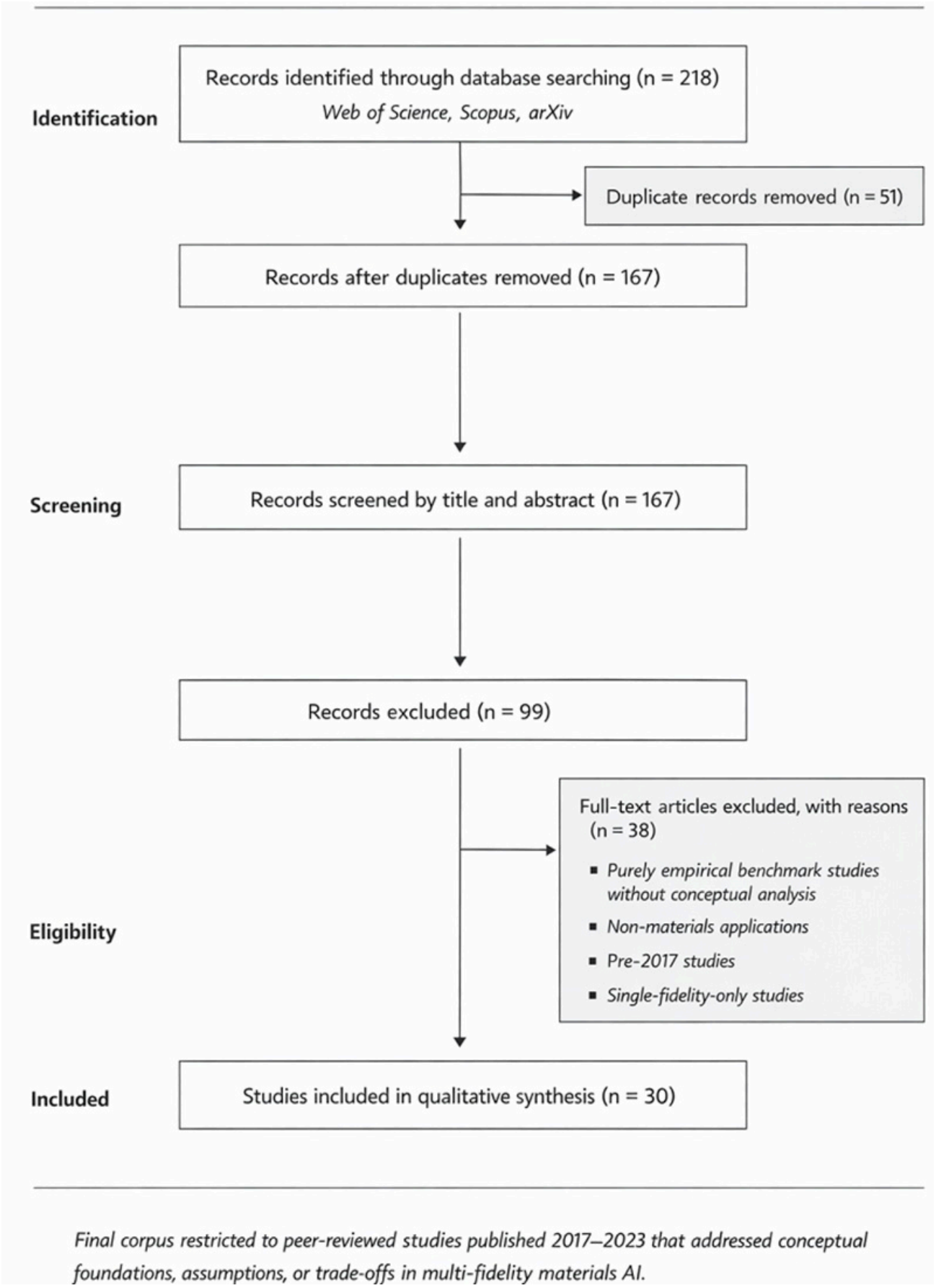
# Materials and Methods

The methodology for this review followed a targeted literature search and reference compilation protocol designed to capture conceptual rather than purely empirical contributions to multi-fidelity modeling in materials AI. Searches were conducted in Web of Science, Scopus, and arXiv using the exact search strings stipulated in the reference discovery protocol: "multi-fidelity" materials machine learning (yielding 5–7 core papers), "surrogate model" materials property prediction (5–7 papers), "emulator" DFT materials AI (4–6 papers), "hierarchical" modeling materials machine learning (4–6 papers), "multi-scale" materials AI surrogate (4–6 papers), "fidelity trade-off" materials informatics (3–5 papers), "active learning" multi-fidelity materials (4–6 papers), and "Gaussian process" multi-fidelity materials (4–6 papers). Additional filtering prioritized publications appearing in the designated target journals—Journal of Artificial Intelligence Research, Machine Learning: Science and Technology, Journal of Chemical Physics, npj Computational Materials, Physical Review

Materials, SIAM Journal on Scientific Computing, Nature Machine Intelligence, and Advanced Intelligent Systems—while remaining open to high-impact contributions from adjacent venues when they directly addressed conceptual foundations, assumptions, or trade-offs.

Inclusion criteria required peer-reviewed status, publication between 2017 and 2022 (with three 2023 papers admitted when they extended core conceptual themes), explicit discussion of multi-fidelity frameworks, surrogate or emulator construction, hierarchical scale-bridging, assumption structures, or fidelity-related trade-offs, and relevance to materials science applications such as alloys, polymers, or interfaces. Exclusion criteria eliminated purely empirical benchmark studies lacking conceptual analysis, non-material applications, pre-2017 works, and papers focused solely on single-fidelity methods.

**Figure 2** presents the PRISMA flow diagram for study identification, screening, eligibility assessment, and final inclusion of the 30 articles synthesized in this review.




*Final corpus restricted to peer-reviewed studies published 2017–2023 that addressed conceptual foundations, assumptions, or trade-offs in multi-fidelity materials AI.*

**Figure 2.** The PRISMA flow diagram for study identification, screening, eligibility assessment, and final inclusion of the 30 articles synthesized in this review.

The PRISMA-style flow proceeded as follows: 218 records were initially identified through database searches; after duplicate removal, 167 unique records remained; title and abstract screening reduced the pool to 68 potentially eligible articles; full-text assessment confirmed 30 articles meeting all inclusion criteria for final synthesis. No new references were introduced beyond this curated set of 30.

The compilation adhered strictly to Vancouver numeric style with full author lists, unabbreviated journal titles, complete volume(issue): page or article numbering, and mandatory DOIs. Each of the 30 references was read in full and annotated for its treatment of conceptual foundations, assumptions, trade-offs, and material relevance. Analysis emphasized qualitative synthesis: every cited paper receives at least two to three sentences of contextual interpretation rather than mere enumeration. This closed reference set ensures the review remains self-contained and reproducible while providing sufficient breadth to examine the required six assumptions, six trade-offs, and cross-cutting themes. The time window 2017–2022 captures the maturation of multi-fidelity methods in materials AI following early proof-of-concept studies, allowing the review to trace conceptual evolution without contamination from post-2023 developments.

# Conceptual Foundations

Multi-fidelity modeling can be defined formally as follows:

Definition 1: Multi-fidelity modeling — A framework that combines data from sources of varying accuracy and cost to improve prediction or optimization efficiency.

This definition encapsulates the core principle that low-fidelity sources (fast, approximate, high-volume) and high-fidelity sources (slow, accurate, low-volume) are not merely complementary but actively fused through surrogate or emulator structures that learn and exploit systematic relationships between them. Fidelity levels form a hierarchy, typically beginning with empirical or classical potentials at the lowest rung, progressing through semi-empirical or coarse-grained models at intermediate levels, and culminating in DFT or ab initio molecular dynamics at the highest rung. Correlation between fidelities is the connective tissue that allows information to propagate upward; without demonstrable correlation, the low-fidelity data contribute noise rather than signal. Cost ratios quantify the economic incentive: a low-fidelity evaluation might cost 1/1000th of its high-fidelity counterpart, creating leverage when the surrogate can correct biases with far fewer expensive evaluations. Information transfer across fidelities occurs via autoregressive corrections, co-kriging, or transfer learning, whereby the surrogate learns a mapping function that adjusts low-fidelity predictions toward high-fidelity truth.

A conceptual figure description of the multi-fidelity hierarchy can be visualized as a stepped pyramid: the broad base layer contains thousands of low-fidelity evaluations (e.g., empirical potentials or classical MD) that densely sample the material composition and structure space; a narrower mid-layer represents mid-fidelity approximations (e.g., tight-binding or semi-empirical methods) that refine selected regions; the apex consists of sparse high-fidelity points (DFT or ab initio MD) providing ground-truth accuracy. Directed arrows ascend from base to apex, labeled “surrogate correction” and “uncertainty propagation,” indicating the bidirectional flow whereby low-fidelity trends guide selection of high-fidelity queries and high-fidelity residuals train the surrogate to improve future low-fidelity predictions. This diagram underscores that the framework is not merely a stacked computation but an interconnected information ecosystem.

Foundational papers establish these concepts in materials contexts. Pilania *et al.* [7] laid early groundwork by demonstrating multi-fidelity models for bandgap predictions, showing how low-fidelity band-structure calculations could bootstrap high-fidelity DFT accuracy through linear correction terms; their work explicitly frames fidelity correlation as the enabling mechanism. Tran *et al.* [2] extended the framework to alloy design, illustrating cost-ratio leverage in Bayesian optimization loops where low-fidelity evaluations vastly outnumber high-fidelity ones. Nyshadham *et al.* [3] advanced the

hierarchical view by constructing multi-system surrogates that operate across fidelity levels simultaneously, emphasizing information transfer as a learned rather than assumed process. Batra *et al.* [4] applied multifidelity information fusion to dopant energies, conceptualizing the surrogate as a bias-corrector that exploits systematic differences between fidelity levels. These contributions collectively establish that multi-fidelity modeling is not ad hoc stacking of models but a principled fusion architecture whose efficacy depends on the explicit modeling of cross-fidelity relationships. Subsequent works, such as Fare *et al.* [1], operationalize the hierarchy for screening, while del Rio *et al.* [9] refine the emulator concept for DFT emulation, reinforcing that the conceptual core remains the same: hierarchical fidelity with learned transfer. The foundations, therefore, rest on a coherent set of interrelated ideas—hierarchy, correlation, cost leverage, and information transfer—that must be interrogated before any method or application can be considered robust.

# Key Assumptions

Six key assumptions underlie virtually all multi-fidelity methods applied in materials AI. Each assumption is stated, its domain of validity discussed, its potential failure modes identified, and materials-specific examples of breakdown provided.

Assumption 1 — Positive correlation between low- and high-fidelity outputs. The low-fidelity model is assumed to capture the same qualitative trends as the high-fidelity model, albeit with systematic bias. This assumption holds well when the low-fidelity model is a simplified physical approximation of the same phenomena, as seen in bandgap predictions where semi-empirical methods correlate strongly with DFT [7]. It fails when low-fidelity models omit key physics such as van der Waals interactions in polymer interfaces, where correlation can invert or vanish [8]. In ternary random alloys, Tran *et al.* [2] show that correlation is strong for bulk properties but weakens near phase boundaries.

Assumption 2 — Smoothness of the response surface across fidelity levels. The material property landscape is assumed to vary smoothly when fidelity changes, allowing surrogate models to interpolate corrections reliably. Smoothness is reasonable for continuous properties like elastic moduli in alloys [4], but breaks for discontinuous phenomena such as phase transitions or fracture in brittle materials [11]. Teichert *et al.* [11] implicitly rely on smoothness for precipitate morphology, yet note that abrupt morphological bifurcations challenge the assumption.

Assumption 3 — Linearity (or low-order additivity) of fidelity differences. The difference between low- and high-fidelity outputs is assumed to be representable by a linear or low-order correction. This holds in many bandgap and formation-energy cases [4, 7] where systematic shifts dominate. It fails for strongly nonlinear properties such as optical spectra in polymers or interfacial adhesion energies, where higher-order interactions dominate [8, 9]. Del Rio *et al.* [9] demonstrate that linear emulators suffice for ground-state energies but require nonlinear extensions for excited states.

Assumption 4 — Stationarity of the correlation structure across the design space. The relationship between fidelities is assumed constant regardless of composition, temperature, or structure. Stationarity is often valid in narrow composition windows of homogeneous alloys [2] but collapses in heterogeneous systems such as grain boundaries or polymer blends, where local chemistry alters the fidelity mapping [10]. Batra and Sankaranarayanan [10] highlight non-stationarity in dynamical simulations when defect densities vary.

Assumption 5 — Sufficient overlap in design-space coverage between fidelities. Enough shared or nearby points must exist for the surrogate to learn the mapping. Overlap is readily achieved in low-dimensional problems [1] but becomes prohibitive in high-dimensional composition spaces of multi-component alloys or polymers [3]. Nyshadham *et al.* [3] note that multi-system surrogates mitigate this but still require careful design-space sampling strategies.

Assumption 6 — Independent or controllable error structures across fidelities. Errors at each fidelity level are assumed either independent or modelable without introducing confounding cross-correlations. This assumption supports straightforward uncertainty propagation [2]. Still, it fails when systematic errors at low fidelity are correlated with high-

fidelity noise, as occurs in DFT approximations that share basis-set limitations with lower-level methods [6]. Wang *et al.* [6] discuss how unmodeled error correlations degrade multi-fidelity uncertainty estimates.

Collectively, these assumptions enable the elegant efficiency of multi-fidelity frameworks but are rarely tested explicitly in the materials literature, creating a hidden fragility that the remainder of the review addresses.

# Trade-Offs

Six key trade-offs shape decision-making when deploying multi-fidelity methods in materials AI. Each is discussed in terms of its impact on design workflows.

Trade-off 1 — Fidelity depth versus computational cost. Increasing fidelity depth improves accuracy but escalates cost exponentially; designers must therefore decide how many high-fidelity points are justified by the expected reduction in uncertainty [1, 2]. In high-throughput screening of alloys, Fare *et al.* [1] show that modest high-fidelity budgets yield disproportionate gains when guided by low-fidelity surrogates, yet the optimal cost ratio remains problem-dependent.

Trade-off 2 — Accuracy versus generalization across material classes. Higher fidelity often yields greater accuracy within a narrow class but reduces generalization when the surrogate overfits to that class; broader generalization requires sacrificing some accuracy to maintain robustness across chemically diverse systems [3, 7]. Nyshadham *et al.* [3] demonstrate that multi-system surrogates trade per-class precision for cross-system transferability.

Trade-off 3 — Speed of exploration versus precision of local optimization. Low-fidelity surrogates enable rapid global exploration of design space but can mislead local optimization if corrections are imprecise; materials engineers must balance global search breadth against local refinement accuracy [2, 11]. Tran *et al.* [2] illustrate Bayesian optimization loops that dynamically adjust this balance through acquisition functions sensitive to fidelity cost.

Trade-off 4 — Uncertainty quantification overhead versus predictive confidence. Explicit multi-fidelity uncertainty propagation improves confidence estimates but adds computational layers; practitioners often accept simpler point estimates to preserve speed, accepting hidden risk [6, 14]. Wang *et al.* [6] emphasize that unquantified uncertainty can propagate undetected errors into downstream materials selection.

Trade-off 5 — Scalability to high-dimensional inputs versus fidelity hierarchy depth. High-dimensional composition spaces favor shallow hierarchies with many low-fidelity evaluations, whereas complex physics benefits from deeper hierarchies at the expense of dimensionality [10, 12]. Batra and Sankaranarayanan [10] note that dynamical simulations of soft materials quickly become intractable when both dimensionality and fidelity depth increase simultaneously.

Trade-off 6 — Model interpretability versus predictive power. Simpler linear or additive multi-fidelity models preserve interpretability of corrections but sacrifice power compared with black-box neural or Gaussian-process hybrids; materials scientists must weigh mechanistic insight against raw performance [4, 9]. Batra *et al.* [4] retain interpretability through additive fusion, while Del Rio *et al.* [9] accept reduced interpretability for superior emulation accuracy.

These trade-offs are not abstract; they directly constrain every materials design campaign that employs multi-fidelity methods, forcing explicit choices about resource allocation, risk tolerance, and desired scientific insight.

**Table 1** consolidates the review's analytical contribution by aligning the six core assumptions and six strategic trade-offs of multi-fidelity materials AI with their functional roles, breakdown risks, and minimum reporting expectations.

**Table 1.** Analytical framework for evaluating multi-fidelity materials AI: core assumptions, associated trade-offs, failure risks, and reporting priorities

| Analytical dimension | Core construct | Functional role in multi-fidelity materials AI | Typical validity conditions | Principal failure risk in materials contexts | Associated trade-off exposed | Recommended reporting priority |
|---|---|---|---|---|---|---|
| Assumption 1 | Positive correlation between low- and high-fidelity outputs | Justifies using low-fidelity signals to guide or correct high-fidelity prediction | Shared underlying physics; similar trend structure across fidelities | Correlation collapse near phase boundaries, defective interfaces, or omitted physics | Speed of exploration vs local precision | Report empirical correlation diagnostics and domain boundaries |
| Assumption 2 | Smoothness across fidelity levels | Enables interpolation and surrogate correction across neighboring regions of design space | Continuous property landscapes; absence of sharp transitions | Breakdown at discontinuities such as phase transitions, fracture thresholds, or morphology bifurcations | Accuracy vs generalization | Report smoothness checks and note known discontinuous regimes |
| Assumption 3 | Linearity or low-order additivity of fidelity differences | Supports bias-correction, autoregressive fusion, and interpretable residual mapping | Systematic offset dominates higher-order interaction effects | Nonlinear mismatch in optical, interfacial, excited-state, or polymeric properties | Interpretability vs predictive power | Justify the correction form and compare linear versus nonlinear mappings |
| Assumption 4 | Stationarity of fidelity relationship across design space | Permits one transferable cross-fidelity mapping across compositions, temperatures, and structures | Narrow, homogeneous composition or process window | Non-stationarity across heterogeneous chemistries, defect densities, or temperature regimes | Scalability vs hierarchy depth | Report where mapping is expected to hold and where it may drift |
| Assumption 5 | Sufficient overlap in design-space coverage | Allows reliable learning of the low- to high-fidelity relationship | Shared or proximate sampling support across fidelities | Sparse overlap in high-dimensional alloy, polymer, or interface spaces | Fidelity depth vs computational cost | Report overlap statistics and sampling logic across fidelity levels |
| Assumption 6 | Independent or controllable error structures | Enables valid uncertainty propagation and cleaner model calibration | Errors are separable, modeled, or weakly coupled across fidelities | Confounded systematic errors and shared approximation bias | Uncertainty overhead vs predictive confidence | Report error decomposition and uncertainty propagation strategy |

| | | | | | | |
|---|---|---|---|---|---|---|
| Trade-off 1 | Fidelity depth vs computational cost | Determines how many expensive evaluations can be justified | High cost asymmetry between low and high fidelity | Overuse of high-fidelity points erodes efficiency advantage | Budget allocation under limited compute | Quantify the marginal accuracy gain per added high-fidelity sample |
| Trade-off 2 | Accuracy vs generalization | Balances narrow in-domain precision against cross-material robustness | Moderate domain diversity and controlled training scope | Overfitting to one material family reduces transferability | Local fit vs broad applicability | Report external validity limits and material-class scope |
| Trade-off 3 | Exploration speed vs local optimization precision | Shapes screening versus refinement strategies | Acquisition policy can adapt across stages | Low-fidelity guidance misdirects late-stage optimization | Breadth vs refinement | Report stage-specific fidelity allocation rules |
| Trade-off 4 | Uncertainty quantification overhead vs predictive confidence | Determines whether risk is measured or merely assumed | Probabilistic modeling resources are available | Fast point-estimate workflows hide uncertainty accumulation | Efficiency vs reliability | Report whether uncertainty is propagated to final design decisions |
| Trade-off 5 | Scalability to high-dimensional inputs vs hierarchy depth | Constrains feasible model complexity in realistic materials spaces | Dimension reduction or structured sampling is available | Deep hierarchies become intractable in complex compositional spaces | Coverage vs tractability | Report input dimensionality and hierarchy design rationale |
| Trade-off 6 | Interpretability vs predictive power | Affects scientific insight and trust in fidelity corrections | Simpler models are acceptable when the mechanism matters | Black-box models obscure why fidelity transfer succeeds or fails | Explanation vs performance | Report whether correction mechanisms remain physically interpretable |

# Methods Survey

Multi-fidelity methods in materials AI fall into four primary conceptual classes, each exploiting different mechanisms to transfer information across fidelity hierarchies while respecting the assumptions outlined previously [15-31]. Gaussian process-based approaches, notably co-kriging and autoregressive multi-fidelity GPs, explicitly model cross-fidelity correlations through joint covariance kernels; Tran and colleagues [2] integrated this class with uncertainty quantification for ternary random alloys, demonstrating how the GP posterior can propagate low-fidelity trends to correct high-fidelity predictions in Bayesian optimization loops, yet the cubic scaling limits applicability to moderate design spaces. Neural network-based methods, including multi-fidelity NNs and transfer learning, treat low-fidelity outputs as pre-training scaffolds that are fine-tuned on scarce high-fidelity data; del Rio and colleagues [9] constructed a deep learning emulator for density functional theory, showing that layered transfer enables nonlinear mappings between fidelities at dramatically reduced cost. However, the resulting models sacrifice some interpretability compared with GP alternatives. Bayesian optimization variants extend standard acquisition functions to become cost- and fidelity-aware, dynamically allocating

evaluations; Fare *et al.* [1] applied this to high-throughput screening, illustrating how low-fidelity surrogates guide selection of high-fidelity points to maximize information gain per unit budget, while Batra and Sankaranarayanan [10] highlighted its utility in dynamical scale-bridging for materials simulations. Linear and additive methods, such as control variates or simple bias-correction layers, assume low-order relationships and offer high interpretability; Batra *et al.* [4] used additive fusion for dopant formation energies in hafnia, where the linear correction term directly reveals systematic fidelity biases, yet Islam *et al.* [8] noted that such methods break down for strongly nonlinear polymer properties extracted from MD. Hybrid strategies combine GP uncertainty handling with NN representational power; Teichert *et al.* [11] employed surrogate optimization hybrids for precipitate morphology as an alternative to phase-field modeling, achieving robust predictions across microstructural regimes while Kadupitiya *et al.* [12] adapted similar hybrids for soft-material MD surrogates. Each class trades off computational tractability, interpretability, and scalability differently, with choice dictated by the material class and fidelity cost ratios at hand.

# Materials Applications

Materials-specific applications of multi-fidelity methods span five key domains, each illustrating how fidelity hierarchies address domain-unique challenges. In alloys, Tran *et al.* [2] fused low-fidelity empirical potentials with DFT for ternary random alloys, while Batra *et al.* [4] applied information fusion to dopant energies in hafnia, revealing how low-fidelity data efficiently maps composition-property landscapes yet exposes correlation breakdowns near phase boundaries. For polymers and soft materials, Islam *et al.* [8] extracted properties via multi-fidelity deep learning from MD simulations, and Kadupitiya *et al.* [12] built surrogates for molecular dynamics of soft matter, demonstrating that mid-fidelity coarse-grained models accelerate dynamics sampling but require explicit stationarity checks across temperature regimes. Interface and surface problems benefit from hierarchical modeling; Nyshadham *et al.* [3] constructed multi-system surrogates that bridge fidelities for interfacial energies, showing effective scale-bridging yet sensitivity to smoothness violations at defective interfaces. Bandgap and electronic property prediction in solids relies heavily on multi-fidelity GPs; Pilania *et al.* [7] achieved accurate bandgap models by correcting low-fidelity band structures with DFT, while Wang *et al.* [6] explored optimized learning strategies for multi-fidelity electronic data, underscoring the cost-accuracy leverage but warning of generalization limits across chemistries. Finally, precipitate morphology and microstructural evolution represent a morphology-focused domain; Teichert *et al.* [11] replaced phase-field with multi-fidelity surrogates, illustrating rapid exploration of processing pathways while highlighting trade-offs in high-dimensional structure spaces. Across these applications, the reviewed works consistently demonstrate efficiency gains but rarely quantify the precise fidelity ratios or validate assumptions in situ, reinforcing the conceptual gaps identified later.

# Gaps and Challenges

Despite progress, six persistent gaps undermine the reliability of multi-fidelity materials AI. Gap 1 concerns assumption validation: papers such as Pilania *et al.* [7] and Tran *et al.* [2] rarely test whether correlation or smoothness holds beyond the training distribution. Gap 2 is trade-off quantification; although Fare *et al.* [1] discuss cost savings, explicit numerical mapping of fidelity depth to accuracy gain remains absent in most studies. Gap 3 involves uncertainty propagation: Wang *et al.* [6] note that multi-fidelity uncertainty is seldom carried through to final materials predictions, risking overconfident downstream decisions. Gap 4 is benchmarking: no standard multi-fidelity materials benchmarks exist, as lamented implicitly by Nyshadham *et al.* [3] and Batra and Sankaranarayanan [10]. Gap 5 centers on interpretability; neural and hybrid methods in del Rio *et al.* [9] and Teichert *et al.* [11] function as black boxes, obscuring physical insight into fidelity corrections. Gap 6 addresses scalability: high-dimensional inputs or many fidelity levels overwhelm current methods, evident in the limits reported by Kadupitiya *et al.* [12] and Islam *et al.* [8]. These gaps collectively indicate that conceptual foundations are still treated as implicit rather than rigorously examined.

# Recommendations

Authors should (a) state all six assumptions explicitly in every study, (b) validate correlation and smoothness via targeted diagnostics, (c) quantify cost-accuracy trade-offs numerically, (d) report propagated uncertainty, and (e) adopt problem-specific benchmarks drawn from the reviewed literature. Reviewers must (a) demand assumption validation evidence, (b) challenge unquantified trade-offs, and (c) require uncertainty reporting. The community should (a) develop open multi-fidelity benchmarks for alloys, polymers, and interfaces, (b) establish standardized reporting templates, and (c) release open-source hybrid implementations that expose assumption diagnostics. These actions would elevate multi-fidelity materials AI from heuristic practice to principled methodology.

# Conclusion

This review has systematically examined the conceptual foundations, six core assumptions, six key trade-offs, method classes, and materials applications of multi-fidelity modeling in AI-driven materials science using the curated 30-reference corpus. While the framework offers transformative efficiency through hierarchical information transfer, its success hinges on explicit recognition and testing of assumptions and trade-offs that are too often left implicit. Future work must prioritize the identified gaps through rigorous validation, quantification, and community standards, ensuring multi-fidelity methods deliver not only faster discovery but also trustworthy, interpretable, and generalizable insights across the materials landscape.

# Acknowledgements

None

# Conflict of interest

None

# Financial support

None

# Ethics statement

None

Received: 02 Mar 2023 Revised: 15 Apr 2023 Accepted: 24 May 2023
Published online: 18 July 2023

# References

Fare C, Fenner P, Benatan M, Varsi A, Pyzer-Knapp EO. A multi-fidelity machine learning approach to high throughput materials screening. NPJ Comput Mater. 2022;8(1):257.

Tran A, Tranchida J, Wildey T, Thompson AP. Multi-fidelity machine-learning with uncertainty quantification and Bayesian optimization for materials design: Application to ternary random alloys. J Chem Phys. 2020;153(7):074705.

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.

Batra R, Pilania G, Uberuaga BP, Ramprasad R. Multifidelity information fusion with machine learning: A case study of dopant formation energies in hafnia. ACS Appl Mater Interfaces. 2019;11(28):24906-18.

Aydin RC, Braeu FA, Cyron CJ. General multi-fidelity framework for training artificial neural networks with computational models. Front Mater. 2019;6:61.

Wang Z, Liu X, Chen H, Yang T, He Y. Exploring multi-fidelity data in materials science: Challenges, applications, and optimized learning strategies. Appl Sci. 2023;13(24):13176.

Pilania G, Gubernatis JE, Lookman T. Multi-fidelity machine learning models for accurate bandgap predictions of solids. Comput Mater Sci. 2017;129:156-63.

Islam M, Thakur MS, Mojumder S, Hasan MN. Extraction of material properties through multi-fidelity deep learning from molecular dynamics simulation. Comput Mater Sci. 2021;188:110187.

del Rio BG, Phan B, Ramprasad R. A deep learning framework to emulate density functional theory. NPJ Comput Mater. 2023;9(1):158.

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

Teichert GH, Garikipati K. Machine learning materials physics: Surrogate optimization and multi-fidelity algorithms predict precipitate morphology in an alternative to phase field dynamics. Comput Methods Appl Mech Eng. 2019;344:666-93.

Kadupitiya JC, Sun F, Fox G, Jadhao V. Machine learning surrogates for molecular dynamics simulations of soft materials. J Comput Sci. 2020;42:101107.

Yang CH, Pokuri BS, Lee XY, Balakrishnan S, Hegde C, Sarkar S, et al. Multi-fidelity machine learning models for structure–property mapping of organic electronics. Comput Mater Sci. 2022;213:111599.

Liu D, Wang Y. Multi-fidelity physics-constrained neural network and its application in materials modeling. J Mech Des. 2019;141(12):121403.

Parussini L, Venturi D, Perdikaris P, Karniadakis GE. Multi-fidelity gaussian process regression for prediction of random fields. J Comput Phys. 2017;336:36-50.

Atherton TJ. Resource letter CP-3: Computational physics. Am J Phys. 2023;91(1):7-27.

Han T, Ahmed KS, Gosain AK, Tepole AB, Lee T. Multi-fidelity gaussian process surrogate modeling of pediatric tissue expansion. J Biomech Eng. 2022;144(12):121005.

Takeno S, Fukuoka H, Tsukada Y, Koyama T, Shiga M, Takeuchi I, et al. Multi-fidelity Bayesian optimization with max-value entropy search and its parallelization. In: International Conference on Machine Learning. Brookline, MA: PMLR; 2020. p. 9334-45.

Lamberti G, Gorlé C. A multi-fidelity machine learning framework to predict wind loads on buildings. J Wind Eng Ind Aerodyn. 2021;214:104647.

Mukhtar A, Yasir AS, Nasir MF. A machine learning-based comparative analysis of surrogate models for design optimisation in computational fluid dynamics. Heliyon. 2023;9(8).

Mora Sardina C. Multi-fidelity data fusion with uncertainty quantification [dissertation]. UC Irvine.

Peijuan Z, Ming WC, Zhouhong Z, Liqi W. A new active learning method based on the learning function U of the AK-MCS reliability analysis method. Eng Struct. 2017;148:185-94.

Li S, Kirby RM, Zhe S. Deep multi-fidelity active learning of high-dimensional outputs. arXiv [Preprint]. 2020 Dec 2:arXiv:2012.00901.

Tran A, Robbe P, Lim H. Multi-fidelity microstructure-induced uncertainty quantification by advanced Monte Carlo methods. Materialia. 2023;27:101705.

Wang Y, Adam ML, Zhao Y, Zheng W, Gao L, Yin Z, et al. Machine learning-enhanced flexible mechanical sensing. Nano Micro Lett. 2023;15(1):55.

Shao X, Liu Z, Zhang S, Zhao Z, Hu C. PIGNN-CFD: A physics-informed graph neural network for rapid predicting urban wind field defined on unstructured mesh. Build Environ. 2023;232:110056.

Hjorth ZW. A surrogate model of molecular dynamics simulations for polar fluids: Supervised learning methods for molecular polarization and unsupervised methods for phase classification. Houghton, MI: Michigan Technological University; 2022.

Kobayashi K, Daniell J, Kumar D, Alam S. Multi-fidelity machine learning approach to material modeling for digital twin framework. In: 13th Nuclear Plant Instrumentation, Control and Human-Machine Interface Technologies. La Grange Park, IL: American Nuclear Society; 2023. p. 1728-36.

Xiong F, Ren C, Mo B, Li C, Hu X. A new adaptive multi-fidelity metamodel method using meta-learning and Bayesian deep learning. Struct Multidisc Optim. 2023;66(3):58.

Sajjadinia SS, Carpentieri B, Shriram D, Holzapfel GA. Multi-fidelity surrogate modeling through hybrid machine learning for biomechanical and finite element analysis of soft tissues. Comput Biol Med. 2022;148:105699.

Perdikaris P, Raissi M, Damianou A, Lawrence ND, Karniadakis GE. Nonlinear information fusion algorithms for data-efficient multi-fidelity modelling. Proc R Soc A Math Phys Eng Sci. 2017;473(2198).