REVIEW Open access

Physics-Integrated Machine Learning for Materials Science — Conceptual Taxonomies and Open Questions

Daniel Brooks[1*], Amelia Carter[2], Ethan Moore[1]

Abstract

The integration of physical principles into machine learning (ML) frameworks has emerged as a transformative approach in materials science, addressing the limitations of purely data-driven models by incorporating domain knowledge to enhance predictive accuracy, generalizability, and interpretability. This narrative review explores the conceptual taxonomies of physics-integrated ML methods, their applications in materials discovery and design, and the associated challenges in data bias and ethical considerations. Drawing on recent peer-reviewed literature, we classify physics-integration strategies such as physics-informed neural networks (PINNs), hybrid models combining ML with physical simulations, and constraint-based learning, and highlight their roles in solving complex problems such as material property prediction, microstructure analysis, and phase stability. We also examine how data biases in training datasets can propagate errors and inequities in model outputs, and discuss the ethical values underpinning the use of AI in scientific research, including transparency, accountability, and societal impact. The review underscores the potential of these methods to accelerate innovation in materials science while emphasizing the need for rigorous validation and interdisciplinary collaboration. By synthesizing current advancements, this article aims to provide a foundational understanding for researchers and practitioners, paving the way for future developments in this interdisciplinary field.

Keywords Materials informatics, Data bias, Materials science, Physics-informed machine learning, Conceptual taxonomies, Ethical considerations

*Correspondence:
Daniel Brooks
daniel.brooks@gmail.com
[1] Department of Materials Informatics, Faculty of Engineering, University of Manchester, Manchester, United Kingdom
[2] Department of Artificial Intelligence Systems, Faculty of Computer Science, University of Birmingham, Birmingham, United Kingdom

Introduction

The discovery and optimization of advanced materials underpin technological progress across a wide range of domains, including energy conversion and storage, microelectronics, aerospace engineering, and biomedical applications. Traditionally, materials science has relied on a triad of empirical experimentation, theoretical analysis, and computational modeling to understand structure–property–processing relationships and guide materials design. While these approaches have yielded substantial breakthroughs, they are increasingly strained by the sheer complexity of modern materials systems. The multidimensional design space—spanning chemical composition, crystallographic and microstructural features, synthesis and processing routes, and operational environments—renders exhaustive exploration by conventional trial-and-error methods impractical, costly, and time-intensive [1, 2].

In recent years, machine learning (ML) has emerged as a transformative tool in this context, offering data-driven frameworks that can extract hidden correlations from large, heterogeneous datasets. ML models have demonstrated

remarkable success in predicting material properties, screening candidate compounds, optimizing processing parameters, and accelerating the materials discovery cycle [3]. By leveraging high-throughput computational databases and experimental repositories, ML-driven approaches promise to reduce development timelines and resource consumption substantially. Nevertheless, despite their growing adoption, purely data-centric ML models exhibit notable limitations when applied to materials science problems. These include limited extrapolation beyond the training domain, susceptibility to overfitting—especially in regimes with sparse, noisy, or biased data—and a general lack of physical interpretability, which undermines trust and limits scientific insight [4].

To address these challenges, physics-integrated machine learning has gained increasing attention as a paradigm that synergistically combines data-driven learning with established physical principles. By embedding governing equations, conservation laws, symmetries, and known material constraints directly into ML architectures, these hybrid models ensure consistency with fundamental physics while retaining the flexibility of statistical learning [5]. Such integration enhances model robustness, improves generalization to unseen conditions, and yields predictions that are both physically plausible and scientifically interpretable—an essential requirement in materials science, where adherence to physical laws is non-negotiable [6].

Among the most prominent examples of this approach are physics-informed neural networks (PINNs), which incorporate partial differential equations into the training objective, enabling the solution of forward and inverse problems without reliance on large labeled datasets. PINNs have been successfully applied to simulate diffusion, phase transformations, fracture mechanics, and thermo-mechanical behavior in materials systems, offering a powerful alternative to traditional numerical solvers [7]. Beyond PINNs, other physics-integrated strategies include constraint-based learning, symmetry-preserving architectures, and hybrid models that couple ML with first-principles or continuum simulations, further expanding the methodological toolkit available to materials researchers.

Parallel to these developments, materials informatics has emerged as a unifying framework that integrates data science, ML, and domain knowledge to manage and exploit materials data for accelerated discovery systematically [8]. By organizing experimental and computational data into structured databases and enabling predictive analytics at scale, materials informatics has catalyzed progress across alloy design, functional materials, and soft matter systems. However, the effectiveness of informatics-driven approaches is not without limitations. Data bias—arising from the overrepresentation of well-studied material classes or favorable experimental conditions—can distort model predictions and hinder transferability to unexplored regions of the materials space [9]. Additionally, the increasing reliance on AI-driven decision-making raises ethical and societal considerations, including transparency, reproducibility, accountability, and the broader implications of algorithmic influence on scientific research and innovation pathways [10].

Against this backdrop, the present review aims to provide a critical and comprehensive examination of physics-integrated machine learning in materials science. Specifically, the objectives are to: (1) survey state-of-the-art methodologies that integrate physical knowledge with ML; (2) propose conceptual taxonomies of integration strategies; (3) highlight representative applications and case studies across materials domains; (4) analyze the impact of data bias and dataset limitations in materials informatics; and (5) discuss ethical values and responsible AI practices in the context of AI-driven materials research. By synthesizing peer-reviewed literature, this review seeks to offer a balanced perspective on the transformative potential and critical challenges of this rapidly evolving field, while identifying open questions and future research directions.

Main Text

Fundamentals of physics-integrated machine learning

Physics-integrated machine learning (PIML) constitutes a fundamental departure from conventional data-centric machine learning paradigms by explicitly embedding physical domain knowledge into learning algorithms. Rather than treating physical systems as black boxes, PIML seeks to harmonize the inductive capabilities of ML with the deductive structure of

physical laws, thereby producing models that are not only accurate but also interpretable, data-efficient, and robust under extrapolation [11]. This integration introduces meaningful inductive biases that constrain the hypothesis space of ML models, enabling them to generalize more reliably in regimes where data are scarce, noisy, or incomplete—conditions that are typical in materials science.

At a foundational level, physics integration can occur at multiple stages of the ML pipeline, including model architecture, training objectives, and inference strategies. Common mechanisms include incorporating physics-based loss functions that penalize violations of governing equations, developing hybrid frameworks that couple ML surrogates with classical numerical solvers, and designing equivariant or invariant neural networks that explicitly preserve physical symmetries, such as translational, rotational, or permutational invariance [12]. These strategies ensure consistency with conservation laws and symmetry principles, which are essential for modeling material behavior across length and time scales.

Among the most influential methodologies in this domain are physics-informed neural networks (PINNs), originally proposed to address forward and inverse problems described by partial differential equations (PDEs) [13]. In the PINN framework, neural networks are trained using a composite loss function that combines data-driven terms with physics-based residuals derived from the governing equations. This formulation enforces adherence to fundamental principles such as conservation of mass, momentum, and energy, even in the absence of dense labeled datasets. In materials science, PINNs have been successfully applied to model heat conduction in heterogeneous composites, phase evolution during solidification, and nonlinear stress–strain relationships in polymers and soft materials [14]. Recent extensions, including variational PINNs and Bayesian formulations, enable uncertainty quantification, while adaptive sampling and multi-domain decomposition techniques improve scalability and performance for multiscale and multiphysics problems [15].

Beyond PINNs, physics-integrated ML plays a critical role in enhancing and accelerating traditional computational materials science methods. First-principles approaches, such as density functional theory (DFT), and atomistic simulations, such as molecular dynamics (MD), provide high-fidelity insights into material behavior but are often prohibitively expensive for large systems or long time scales. ML models trained on outputs from these simulations and constrained by physical priors can serve as efficient surrogate models that dramatically reduce computational cost while maintaining acceptable accuracy [16]. Techniques such as Gaussian process regression with physics-based kernels, neural network interatomic potentials, and force-matching strategies exemplify this synergy between physics-based simulations and data-driven learning.

Collectively, these foundational concepts establish a coherent framework for categorizing physics-integrated ML approaches based on the extent and manner of their incorporation of physical knowledge. Existing methodologies span a spectrum from weakly constrained models, which impose physical consistency through post-processing or regularization, to strongly constrained models, in which governing equations and physical laws are explicitly embedded in the learning architecture. This spectrum forms the basis for systematic taxonomies that clarify methodological trade-offs and guide the selection of appropriate integration strategies for specific materials science applications.

Conceptual taxonomies of physics-integration approaches

Conceptual taxonomies are essential for systematically organizing the diverse and rapidly evolving approaches to physics-integrated machine learning (PIML) in materials science. As integration strategies vary widely in terms of methodology, assumptions, and intended outcomes, taxonomies provide a coherent framework for comparing models, identifying methodological trade-offs, and guiding the selection of appropriate approaches for specific scientific and engineering problems. In this work, physics-integration strategies are classified along three complementary dimensions: the degree of physics integration within the learning pipeline, the nature of the physical knowledge incorporated, and the application domain to which the model is applied [17].

From the perspective of integration level, PIML approaches span a continuum from weak to strong coupling between physical principles and machine learning models. At one end of this spectrum are physics-constrained machine learning methods, in which physical laws are introduced as soft constraints, most commonly through regularization terms in the training loss function. Physics-informed neural networks exemplify this category, as they enforce governing equations, such as conservation of mass or momentum, through residual-based penalties while retaining the flexibility to fit observational data. This strategy has been successfully applied to viscoelastic material modeling, where constitutive relationships are incorporated to improve predictive consistency under nonlinear deformation regimes [18]. A moderate level of integration is achieved in physics-guided machine learning, where physical knowledge is not directly embedded in the model structure. Instead, it informs data generation, feature engineering, or parameter initialization. This approach is widely used in metamaterials research, where high-fidelity numerical simulations provide training datasets that enable ML models to efficiently explore large design spaces without explicitly encoding the underlying physics [19]. At the strongest level of integration are physics-embedded machine learning methods, in which physical operators, constraints, or invariances are hard-coded into the model architecture itself. Examples include neural networks whose convolutional structures are designed to mimic wave propagation or diffusion operators in photonic and acoustic materials, ensuring intrinsic compliance with governing physical laws [20].

A complementary taxonomy is based on the type of physical knowledge incorporated into the learning framework. Some approaches rely on explicit physical formulations, such as partial differential equations describing diffusion processes in alloys, elasticity equations governing mechanical response, or thermodynamic relations defining phase stability [21]. Other methods embed physics implicitly by enforcing symmetry preservation, invariance, or equivariance through specialized representations and architectures, such as graph neural networks that respect crystallographic symmetries. Increasingly, hybrid models that combine explicit and implicit physics have gained prominence. Machine-learning interatomic potentials are a representative example, as they integrate energy conservation and symmetry constraints while learning force fields from quantum-mechanical data, enabling atomistic simulations with near–first-principles accuracy at substantially reduced computational cost [22]. The major classes of physics-integration strategies and their defining characteristics are summarized in **Table 1**.

Table 1. Conceptual taxonomy of physics-integrated machine learning (PIML) approaches

Integration level	Mode of physics incorporation	Representative methods	Conceptual strengths	Key limitations
Weak (physics-constrained)	Soft constraints via regularization or loss penalties	PINNs (residual losses), constraint-augmented regression	Data efficiency; physical plausibility under sparse data	Sensitive to loss balancing; limited multiscale scalability
Moderate (physics-guided)	Physics informs data generation, features, or initialization	ML trained on DFT/MD outputs; physics-guided descriptors	Computational acceleration; flexible deployment	Indirect enforcement of physical laws; biased inheritance
Strong (physics-embedded)	Physical operators and symmetries encoded in architecture	Equivariant networks; operator-learning models	Guaranteed physical consistency; improved extrapolation	Architectural rigidity; higher implementation complexity

A third taxonomy emerges when PIML approaches are classified by their application objectives in materials science. In forward modeling and property prediction tasks, physics-integrated models are used to estimate material properties such as elastic moduli, thermal conductivity, or electronic bandgaps, often leveraging physically informed descriptors or constraints to improve generalization. In contrast, inverse design and generative modeling focus on producing material compositions, crystal structures, or microstructures that satisfy predefined target properties. In such cases, generative models are often coupled with physical feasibility checks, such as thermodynamic stability or manufacturability constraints, to ensure that generated candidates are physically realizable [23]. Collectively, these application-oriented

distinctions highlight how physics integration enhances both predictive accuracy and design reliability, while introducing trade-offs in computational complexity and model scalability.

Applications in materials science

Physics-integrated machine learning has enabled significant advances across a broad range of materials science applications, spanning property prediction, process optimization, and materials discovery. In the context of material property prediction, models that integrate ML with physical constraints have demonstrated improved accuracy and robustness compared to purely data-driven approaches, particularly when experimental data are limited. For polymers and composite materials, physics-informed models have been used to predict mechanical behavior under complex loading conditions, substantially reducing reliance on exhaustive experimental testing [24]. Notably, physics-informed neural networks have been applied to model viscoelastic and hyperelastic responses in large-deformation solids by embedding constitutive laws directly into the learning process, allowing nonlinear material behavior to be captured with fewer training samples [18].

In manufacturing-related applications, physics-integrated ML has shown particular promise in additive manufacturing, where complex thermal and mechanical interactions govern defect formation and microstructure evolution. By coupling ML models with thermal transport equations and process physics, researchers have developed real-time monitoring and defect prediction systems capable of identifying anomalies during fabrication, thereby improving process reliability and product quality [15]. In the field of energy materials, hybrid ML–physics frameworks have been employed to optimize battery electrolytes and electrode materials by enforcing electrochemical equilibria and transport constraints, thereby improving performance and stability predictions [16]. Similarly, the design of mechanical, acoustic, and photonic metamaterials has benefited from generative ML models constrained by physical principles, enabling the discovery of structures with tailored and often non-intuitive properties that would be difficult to obtain through conventional design methods [19]. Representative materials science applications enabled by physics-integrated ML are categorized in **Table 2**.

Table 2. Application domains of physics-integrated ML in materials science

Application domain	Target task	Integrated physics	Conceptual benefit	Persistent challenge
Property prediction	Elasticity, thermal/electronic properties	Conservation laws; constitutive relations	Improved generalization under limited data	Sensitivity to noisy experimental inputs
Process modeling	Additive manufacturing, defect prediction	Heat transfer; phase evolution equations	Real-time monitoring; defect mitigation	Multiphysics coupling complexity
Energy materials	Electrodes, electrolytes, transport modeling	Electrochemical equilibria; diffusion laws	Physically consistent performance prediction	Dataset imbalance across chemistries
Metamaterials and inverse design	Structure generation under constraints	Stability and manufacturability conditions	Physically realizable generative outputs	Computational expense of feasibility checks

Case studies further illustrate the impact of physics integration on scalability and reliability. Recent reviews highlight the application of PINNs in polymer science for simulating chain dynamics, phase separation, and thermomechanical transitions, demonstrating their ability to bridge molecular- and continuum-scale descriptions [17]. In high-pressure and extreme-condition materials research, machine-learning interatomic potentials have accelerated the discovery of novel phases by enabling large-scale atomistic simulations that would otherwise be computationally prohibitive [18]. Despite

these successes, challenges remain in integrating multi-fidelity data sources and ensuring consistent performance across disparate material systems.

Challenges in data bias in materials informatics

Data bias represents a persistent and critical challenge in materials informatics, with significant implications for the reliability and generalizability of machine learning models [19]. Biases commonly arise from the uneven representation of material classes, experimental constraints, and historical research priorities that favor well-studied or thermodynamically stable systems. As a result, ML models trained on such datasets often exhibit degraded performance when applied to underrepresented materials, such as rare-earth compounds or metastable phases, thereby limiting their utility for exploratory discovery [9].

Empirical studies have demonstrated that data bias can substantially distort model predictions in both synthesizability assessments and property estimation tasks. For example, ML models trained on imbalanced crystallographic datasets tend to overestimate the feasibility of common structural types while failing to accurately predict the stability of less-represented configurations [20]. Similar effects have been observed in electronic property prediction, where bandgap estimation models trained predominantly on inorganic materials perform poorly when extended to organic or hybrid systems [21]. These biases not only propagate systematic errors but also reinforce existing research blind spots.

To mitigate these issues, several bias-reduction strategies have been proposed, including entropy-driven active learning schemes that prioritize sampling of underexplored regions of materials space and improve dataset diversity [22]. The development of large-scale, high-quality materials databases has also been advocated to enhance model robustness and generalization [23]. Nevertheless, challenges associated with dataset shift, evolving data distributions, and hidden correlations persist, underscoring the need for routine bias audits and transparent evaluation protocols in materials informatics workflows [24].

Ethical and value considerations in AI for materials science

The deployment of AI in materials science raises ethical concerns that intersect with scientific integrity, societal impact, and responsible innovation [12]. Key issues include transparency in model decision-making, where “black-box” ML can obscure biases or errors, undermining trust in research outcomes. Explainable AI techniques, such as feature importance analysis, are essential for materials applications to ensure accountability [21].

Values in science, such as objectivity and reproducibility, are challenged by AI’s reliance on potentially biased data, risking perpetuation of inequities in material access or environmental impacts [21]. Ethical frameworks advocate for stakeholder engagement, including communities affected by new materials (e.g., in sustainable energy), to address fairness and inclusivity [2].

Broader concerns include intellectual property rights for AI-generated designs and the environmental footprint of training large models [10]. Guidelines for ethical AI use in research emphasize disclosure of AI contributions and education in responsible conduct [21]. Integrating these values ensures that physics-integrated ML advances materials science in an equitable and sustainable way.

Results and Discussion

The integration of physics into machine learning frameworks has profoundly reshaped materials science by establishing a principled connection between empirical data-driven approaches and theory-based physical modeling. Physics-integrated machine learning (PIML) addresses long-standing limitations of purely statistical models by constraining learning with governing laws, conservation principles, and symmetry considerations. Despite these advances, the current state of the

field reveals persistent challenges that limit the robustness, scalability, and generalizability of such approaches, particularly when applied to complex and multiscale material systems [1].

A central limitation lies in the treatment of multiscale phenomena that span atomistic, mesoscopic, and continuum regimes. While physics-informed neural networks have demonstrated strong performance in enforcing physical constraints for continuum-level problems, their extension to atomistic or quantum-mechanical scales remains restricted by computational cost and sensitivity to boundary conditions [2, 22]. This limitation highlights a fundamental tension between physical fidelity and computational efficiency. Hybrid frameworks that combine machine learning surrogates with density functional theory or molecular dynamics simulations attempt to mitigate this trade-off; however, inconsistencies in data resolution, uncertainty propagation, and scale coupling often introduce artifacts that compromise predictive reliability [3]. These issues underscore the need for more rigorous multi-fidelity modeling strategies that explicitly account for uncertainty and information transfer across scales.

Data bias further compounds these technical challenges. As discussed earlier, materials informatics datasets are often skewed toward well-characterized, industrially relevant material classes, such as metals and ceramics. In contrast, organic materials, low-dimensional systems, and nanomaterials remain underrepresented [4, 23]. Such imbalances not only degrade predictive accuracy for minority material classes but also reinforce inequities in research focus and funding allocation. Recent studies indicate that biased training data can systematically distort predictions of synthesizability and stability, leading to overconfidence in conventional systems while overlooking unconventional or emerging material candidates [5]. Although mitigation strategies such as active learning, entropy-based sampling, and targeted data acquisition have shown promise, their effectiveness depends critically on the availability of diverse, high-quality datasets, which remain limited in many application domains [6].

Ethical considerations introduce an additional layer of complexity to the deployment of physics-integrated ML in materials science. As these models increasingly inform decisions in high-stakes applications—such as biomedical implants, structural materials, and energy storage technologies—transparency and accountability become essential [7, 24]. Black-box models, even when physically constrained, risk eroding trust if their predictions cannot be interrogated or validated by domain experts [8]. Explainable artificial intelligence techniques, including saliency analysis and counterfactual reasoning, are emerging as tools to improve interpretability and facilitate model validation. However, their adoption within materials science remains nascent, and standardized protocols for explainability in physics-integrated contexts have yet to be established [9].

Addressing these challenges necessitates sustained interdisciplinary collaboration. Effective progress in PIML requires coordinated efforts among physicists, materials scientists, computer scientists, and ethicists to ensure that models are not only accurate but also fair, transparent, and aligned with societal needs [10]. Incorporating stakeholder perspectives during model development can help identify biases embedded in historical data collection and experimental practices [11]. Moreover, the environmental cost of training large-scale physics-integrated models—particularly those coupled with extensive simulations—raises sustainability concerns. The carbon footprint of high-performance computing workloads underscores the need for energy-efficient algorithms and responsible computational practices [25]. Key unresolved technical and ethical challenges in physics-integrated materials AI are synthesized in **Table 3**.

Table 3. Open challenges and ethical dimensions in physics-integrated materials AI

Challenge category	Core issue	Underlying cause	Conceptual implication
Multiscale modeling	Inconsistent scale coupling	Heterogeneous data fidelities	Limits the robustness of cross-scale inference
Data bias	Overrepresentation of stable materials	Historical and experimental biases	Reinforces blind spots in discovery

Interpretability	Opaque decision pathways	Complex hybrid architectures	Erodes trust and scientific accountability
Sustainability and ethics	Computational and societal cost	Energy-intensive training; IP ambiguity	Necessitates responsible AI governance

When synthesizing the proposed taxonomies and reviewing applications, it becomes evident that the effectiveness of physics-integration strategies is strongly problem-dependent. Constraint-based learning approaches perform well in forward property prediction tasks but often struggle in inverse design scenarios, where multiple physically admissible solutions may exist [14]. Case studies in additive manufacturing demonstrate tangible successes in process monitoring and optimization, yet failures in defect prediction reveal the importance of rigorous uncertainty quantification and validation under real-world operating conditions [26]. Collectively, these observations suggest that physics-integrated machine learning is transitioning from an exploratory phase to a stage of methodological consolidation, in which future advances must be accompanied by systematic validation, awareness of uncertainty, and ethical oversight.

Figure 1 conceptualizes the interactions among physical knowledge, data characteristics, and value considerations that collectively shape the behavior and reliability of physics-integrated machine learning systems in materials science.

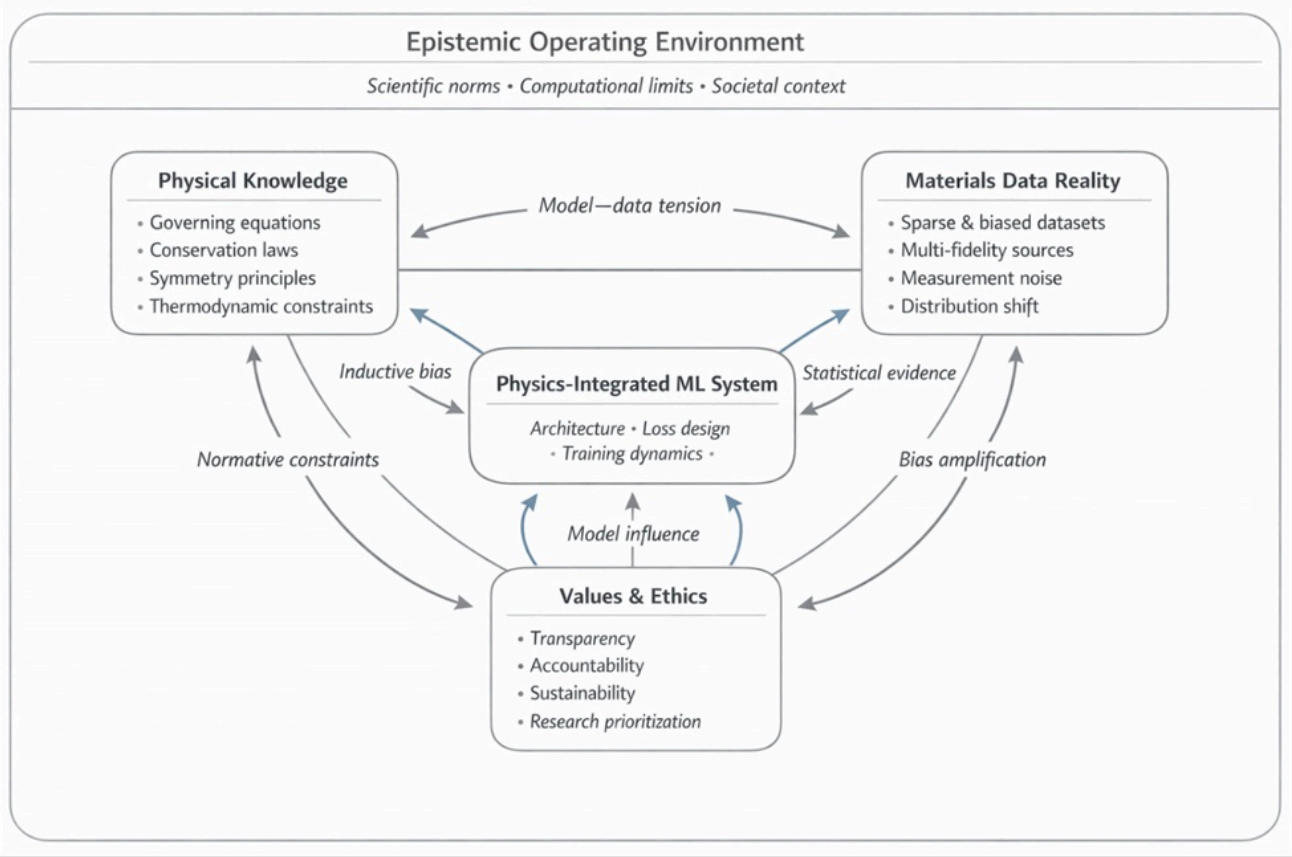

Figure 1. Interaction dynamics of physics, data, and values in materials machine learning

Conclusion

Physics-integrated machine learning represents a significant methodological advance in materials science, enabling more accurate, interpretable, and physically consistent solutions to longstanding challenges in materials discovery, characterization, and design. By embedding governing laws, symmetry principles, and domain knowledge directly into

machine learning frameworks, these approaches transcend the limitations of purely data-driven models and offer a robust pathway to accelerated, trustworthy materials innovation. The conceptual taxonomies presented in this review provide a structured lens through which diverse integration strategies can be understood and selected. At the same time, the examined applications highlight both the field's transformative potential and its current limitations.

Despite notable progress, critical challenges remain unresolved. Persistent data bias continues to constrain generalizability and equitable discovery across materials classes, while the scalability of physics-integrated models to multiscale and multiphysics systems remains computationally demanding. Ethical considerations, particularly those related to transparency, accountability, and sustainability, must also be addressed more systematically as AI-driven methods increasingly influence high-impact materials decisions.

Future research should prioritize the development of robust physics-integrated frameworks that can adapt to distribution shifts in real-world data, potentially through online and adaptive learning strategies. Efforts to reduce bias in materials informatics must focus on creating diverse, standardized datasets and on incorporating bias-aware learning mechanisms. Equally important is the deeper integration of ethical values into model design and deployment, supported by interdisciplinary guidelines that promote interpretability, inclusivity, and responsible use. Advances in multi-fidelity and multiscale modeling offer promising routes to bridge atomistic and continuum descriptions, while emerging computational paradigms may further enhance the efficiency and reach of physics-integrated approaches. Ultimately, sustained collaboration between academia, industry, and policymakers will be essential to ensure that the benefits of physics-integrated machine learning are realized broadly and responsibly across the materials science ecosystem.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 29 Jul 2025 Revised: 19 Aug 2025 Accepted: 29 Sep 2025
Published online: 18 January 2026

References

Wu Y, Li Y, Wang X, et al. Physics-informed machine learning: A comprehensive review on applications in anomaly detection and condition monitoring. Expert Syst Appl. 2024;255:124678.

Perdikaris P. Probabilistic data fusion and physics-informed machine learning: A new paradigm for modeling under uncertainty, and its application to accelerating the discovery of new materials. Univ Pennsylvania. 2024.

Yang S, Peng S, Guo J, Wang F. A review on physics-informed machine learning for monitoring metal additive manufacturing process. Adv Manuf. 2024;1(2):1-2.

Merchant A, Batzner S, Schoenholz SS, et al. Advancing materials science through next-generation machine learning. PMC. 2024;PMC11285097.

Oviedo F, Ferres L, Ren Z, et al. Explainable machine learning in materials science. npj Comput Mater. 2022;8:204.

Zohdi TI. Preface to the theme issue ‘physics-informed machine learning and its structural integrity applications. Philos Trans A Math Phys Eng Sci. 2023;381(2264):20230176.

Kalidindi SR, Brough DB, et al. Scientific AI in materials science: A path to a sustainable and scalable paradigm. Mach Learn Sci Technol. 2020;1(3):033001.

Wang Y, Zhang L, Liu Z. Physics-informed machine learning enabled virtual experimentation for 3D printed thermoplastic. Mater Horiz. 2024;11:6028-39.

Ghimire S, Mishra A, Matos R. Towards physics-informed explainable machine learning and causal models for materials research. Comput Mater Sci. 2024;233:112740.

Mortazavi B, Novikov IS, Podryabinkin EV, Roche S, Rabczuk T, Shapeev AV, et al. Exploring phononic properties of two-dimensional materials using machine learning interatomic potentials. Appl Mater Today. 2020;20:100685.

Hermann E, Good B, et al. Machine intelligence in metamaterials design: A review. Oxf Open Mater Sci. 2024;4(1):itae001.

Abbasiantaeb Z, Ghodrati M, et al. Revolutionizing physics: A comprehensive survey of machine learning applications. Front Phys. 2024;12:1322162.

Götz M, Field JG. Data sharing and data integrity. In: Data, methods and theory in the organizational sciences. Routledge; 2022. p. 49-72.

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

Ghimire S, et al. Roadmap on data-centric materials science. Model Simul Mater Sci Eng. 2024;32(6):065001.

Zhang Y, et al. Mitigating bias in scientific data: A materials science case study. NeurIPS. 2023.

Resnik DB, Elliott KC, Miller AK, et al. The ethics of using artificial intelligence in scientific research: New guidance needed for a new tool. AI Ethics. 2024.
https://doi.org/10.1007/s43681-024-00493-8.

Oberoi A, et al. The sociodemographic biases in machine learning algorithms: A biomedical informatics perspective.

O'Connor TC, et al. Lessons from 'a general-purpose machine learning framework for predicting properties of inorganic materials'. Digit Discov. 2024;3:199.

NITRD. 2020–2024 progress report: Advancing trustworthy artificial intelligence R&D. NITRD.gov. 2024.

Li C, Zheng K. Methods, progresses, and opportunities of materials informatics. InfoMat. 2023;5(8):e12425.

Schiff D. Out of the laboratory and into the classroom: The future of artificial intelligence in education. AI Soc. 2022;36:331-48.

Abuhassna H, Awae F, Adnan MA, Daud M, Almheiri AS. The information age for education via artificial intelligence and machine learning: A bibliometric and systematic literature analysis. Int J Inf Educ Technol. 2024;14(5):700.

Crompton H, Burke D. Artificial intelligence in higher education: The state of the field. Int J Educ Technol High Educ. 2023;20(1):22.

Bearman M, Ryan J, Ajjawi R. Discourses of artificial intelligence in higher education: A critical literature review. High Educ. 2023;86(2):369-85.

Akinwalere SN, Ivanov V. Artificial intelligence in higher education: Challenges and opportunities. Border Cross. 2022;12(1):1.