REVIEW **Open access**

Governance and Responsible Use of Materials AI — Standards, Transparency, and Risk Management

Fatima Zahra Amrani[1*], Youssef Benali[1], Samira El-Haddad[2], Karim Idrissi[1]

Abstract

The integration of artificial intelligence (AI) into materials science, often referred to as Materials AI, has revolutionized the field by enabling accelerated discovery, design, and optimization of new materials. This narrative review explores the governance and responsible use of Materials AI, focusing on standards, transparency, and risk management. Drawing on peer-reviewed literature, we examine the evolution of AI applications in materials science, ethical considerations, and the need for robust frameworks to ensure accountable deployment. Key themes include the ethical implications of data bias and intellectual property in AI-driven materials discovery; the development of standards for model validation and interoperability; mechanisms to enhance transparency in black-box AI models; and strategies to identify and mitigate risks, such as model unreliability and societal impacts. The review highlights how autonomous experimentation systems and machine learning techniques have transformed materials research, while underscoring the importance of reflexive governance to address potential harms. Objectives include synthesizing current practices, identifying gaps in responsible AI adoption, and proposing pathways for sustainable integration. By fostering transparency and risk-aware approaches, Materials AI can contribute to societal benefits, such as advancing energy materials and sustainable manufacturing, while minimizing ethical pitfalls. This work emphasizes the interdisciplinary nature of responsible Materials AI and calls for collaboration among scientists, policymakers, and ethicists to establish trustworthy systems.

Keywords Materials AI, Responsible AI, Governance, Transparency, Risk management, Ethical standards

*Correspondence:
Fatima Zahra Amrani
fatima.amrani@gmail.com
[1] Department of Computational Materials Science, Faculty of Engineering, Mohammed V University, Rabat, Morocco
[2] Department of Artificial Intelligence in Engineering Systems, Faculty of Engineering, University of Fez, Fez, Morocco

Introduction

Materials science has historically functioned as a foundational driver of technological progress, enabling advances across energy systems, electronics, healthcare, aerospace, and environmental sustainability. Breakthroughs in material composition, structure, and processing have consistently shaped industrial capabilities and societal infrastructure. In recent years, the integration of artificial intelligence (AI) into materials research has catalyzed a paradigm shift in how materials are discovered, characterized, and optimized. By leveraging data-driven inference, AI systems now enable the prediction of material properties, the exploration of vast compositional spaces, and the acceleration of discovery workflows at temporal and spatial scales previously unattainable with conventional experimental or computational approaches [1, 2].

The term Materials AI broadly encompasses the application of machine learning (ML), deep learning, and generative modeling techniques to materials-centric tasks such as property prediction, structure–property mapping, inverse design, and high-throughput screening [3, 4]. These methods have demonstrated substantial promise across diverse material classes, including alloys, polymers, ceramics, nanomaterials, and functional composites. By integrating heterogeneous datasets derived from experiments, simulations, and curated repositories, AI-driven frameworks have accelerated the identification of candidate materials for batteries, catalysts, photovoltaics, and biomedical applications [5, 6]. In this sense, AI is increasingly positioned not merely as a computational aid, but as an epistemic instrument that reshapes how material knowledge is generated and operationalized.

Despite these transformative capabilities, the rapid proliferation of Materials AI raises critical concerns regarding its responsible development and deployment. Many state-of-the-art AI models—particularly deep neural networks and generative architectures—operate as opaque or partially interpretable systems, producing high-accuracy predictions without transparent reasoning pathways [7]. This “black-box” characteristic poses challenges for scientific trust, reproducibility, and validation, especially in safety-critical or resource-intensive materials applications. Opaque decision-making can obscure latent biases embedded in training data, amplify uncertainties, and hinder the identification of failure modes that may only manifest under extrapolative conditions [8].

Beyond technical opacity, the governance landscape surrounding Materials AI remains fragmented and underdeveloped. Unlike mature engineering disciplines with well-established validation standards and regulatory oversight, Materials AI currently lacks harmonized protocols for model verification, data provenance, uncertainty reporting, and ethical accountability [9]. Emerging risks extend beyond model performance to encompass broader societal and environmental implications. These include the optimization of computationally efficient but environmentally unsustainable materials, unresolved intellectual property challenges associated with shared datasets and collaborative platforms, and inequities in access to AI infrastructure that may reinforce global disparities in scientific capability [10, 11]. Collectively, these concerns underscore the need to situate Materials AI within a broader framework of responsible innovation.

Against this backdrop, the present review pursues three interrelated objectives. First, it provides a structured overview of the evolution and current role of AI in materials science, emphasizing not only technical capabilities but also epistemic and ethical dimensions. Second, it critically examines existing approaches to transparency, validation, and risk management within Materials AI, identifying conceptual gaps and unresolved tensions. Third, it synthesizes these insights to propose governance-oriented recommendations to promote ethical, sustainable, and socially responsible innovation. By integrating perspectives from materials science, AI ethics, and science policy, this review seeks to inform researchers, policymakers, and industry stakeholders navigating the rapidly evolving landscape of Materials AI.

The evolution of AI in materials science

The application of artificial intelligence in materials science has undergone a marked evolution, progressing from narrowly scoped predictive models to increasingly autonomous and integrated discovery systems. Early implementations predominantly relied on supervised learning techniques to establish correlations between material descriptors and target properties, such as mechanical strength, thermal stability, or electronic behavior. Neural networks and regression-based models were commonly trained on compositional or structural features to approximate structure–property relationships, offering modest gains in predictive efficiency over traditional empirical approaches [12].

By the late 2010s and early 2020s, advances in deep learning architectures and representation learning enabled a shift toward more expressive models capable of handling complex, high-dimensional materials data. These developments facilitated the emergence of generative design paradigms, in which AI systems do not merely evaluate existing candidates but actively propose novel material structures and compositions [13]. Techniques such as generative adversarial networks (GANs) and variational autoencoders have been deployed to explore expansive chemical and structural design spaces, identifying candidate materials for high-performance batteries, catalysts, and functional coatings that would be infeasible to enumerate exhaustively using conventional methods [14, 15].

A pivotal driver of this transformation has been the convergence of AI with high-throughput computation, automated synthesis, and robotic experimentation. This integration has given rise to so-called “self-driving laboratories,” where AI systems orchestrate closed-loop workflows encompassing hypothesis generation, material synthesis, characterization, and iterative model refinement [16]. By minimizing human intervention in routine experimental cycles, these platforms have demonstrated the potential to compress discovery timelines from years to months, fundamentally altering the tempo of materials innovation [17].

Concurrently, methodological advances have addressed long-standing challenges associated with data scarcity in materials research. Unlike domains with massive labeled datasets, materials science often works with limited, heterogeneous, and noisy data. To mitigate these constraints, researchers have increasingly adopted physics-informed and hybrid modeling strategies that embed domain knowledge, conservation laws, or mechanistic constraints into learning architectures [18]. Looking forward, the anticipated convergence of AI with quantum computing and advanced simulation techniques holds promise for further enhancing the fidelity of atomistic and electronic structure predictions, potentially expanding the scope of AI-enabled materials design [19].

Nevertheless, the pace of technical innovation has significantly outstripped the development of corresponding governance and oversight mechanisms. As Materials AI systems assume more autonomous and influential roles in scientific decision-making, questions of responsibility, accountability, and ethical alignment become increasingly salient. Addressing these challenges is essential to ensuring that the accelerating capabilities of Materials AI translate into durable scientific and societal benefits rather than unintended or inequitable outcomes [1].

Ethical considerations in materials AI

Ethical challenges in Materials AI arise not only from the general risks associated with artificial intelligence, but also from domain-specific characteristics of materials research, including data scarcity, high economic and environmental stakes, and long downstream impact horizons. A central concern is the amplification of bias through training data that are historically contingent, incomplete, or institutionally skewed. Materials datasets often originate from proprietary industrial sources, legacy experimental programs, or simulation campaigns optimized for narrow performance objectives. As a result, AI models trained on such data may systematically privilege well-studied material classes—such as conventional alloys or petrochemical polymers—while underrepresenting emerging, sustainable, or locally sourced alternatives [7].

This imbalance has concrete ethical implications. If environmentally benign or low-resource materials are sparsely represented in training corpora, AI-driven optimization may inadvertently favor high-performance but environmentally harmful materials, reinforcing unsustainable design trajectories [20]. Unlike bias in consumer-facing AI systems, these distortions may remain latent until materials are deployed at scale, at which point environmental, economic, or health consequences become difficult to reverse. Ethical evaluation in Materials AI must therefore account for downstream material lifecycles, not merely algorithmic fairness at the point of prediction.

Intellectual property (IP) considerations further complicate the ethical landscape. Materials AI increasingly relies on collaborative platforms that aggregate datasets from academia, industry, and public repositories. While such openness accelerates innovation, it also raises concerns regarding unauthorized data reuse, ambiguous ownership of AI-generated materials designs, and inequitable benefit distribution among contributors [8]. These issues are particularly salient when AI systems generate novel compositions or structures derived from proprietary inputs, blurring the boundary between legitimate inference and unlicensed derivation.

The dual-use potential of Materials AI introduces an additional ethical dimension. Techniques developed for benign applications—such as biodegradable polymers, energy storage materials, or biomedical implants—may be readily repurposed for harmful ends, including advanced weaponry, surveillance-enabling materials, or environmentally destructive technologies [10]. Unlike many software-only AI systems, materials innovations are physically instantiated, making their misuse persistent and difficult to contain. This reality underscores the inadequacy of generic AI ethics frameworks when applied without modification to materials research contexts.

Recent literature emphasizes the need for ethical guidelines that are scientifically situated, drawing inspiration from broader AI ethics principles while adapting them to the epistemic norms, risk profiles, and temporal scales of materials science [7]. In this view, ethics is not an external constraint imposed post hoc, but an integral component of materials AI system design and deployment.

Responsible research and innovation (RRI) frameworks offer a promising foundation for this integration. RRI emphasizes anticipatory governance, reflexivity, and responsiveness, advocating ethical reflection early in the research and development pipeline rather than after technological lock-in [21]. In Materials AI, this entails embedding ethical assessment at stages such as data curation, objective function selection, and model validation. Stakeholder engagement —encompassing materials scientists, ethicists, policymakers, industry actors, and affected communities—plays a critical role in surfacing value assumptions that might otherwise remain implicit in model design choices [22]. Such engagement helps align AI-driven materials innovation with societal priorities, including sustainability, equity, and long-term resilience.

Standards and frameworks for responsible materials AI

Establishing shared standards is central to translating ethical principles into operational practice within Materials AI. Standards serve not only as technical benchmarks but also as governance instruments that shape norms of accountability, transparency, and interoperability across institutions and sectors. In recent years, the materials community has begun developing domain-specific benchmarking efforts, such as standardized evaluation suites for machine learning models applied to crystal structure prediction and materials stability assessment [23]. These benchmarks enable systematic comparison of models, reducing the risk of overfitting to idiosyncratic datasets and promoting reproducibility across research groups.

Beyond performance evaluation, standards play a critical role in facilitating model transferability. Materials AI models are often deployed beyond their original training context, applied to new chemistries, processing conditions, or scales. Without standardized reporting of data provenance, uncertainty, and applicability domains, such transfers risk producing misleading or unsafe predictions [4]. Harmonized standards, therefore, function as safeguards, enabling informed reuse while delineating the limits of model validity.

At the international level, broader responsible AI frameworks—such as those articulated by intergovernmental organizations—have articulated high-level principles emphasizing human-centered values, transparency, accountability, and robustness [21]. While these principles provide an important ethical compass, their direct application to materials science requires contextual adaptation. In response, materials-focused initiatives have begun emphasizing data quality standards that mandate explicit metadata on experimental conditions, simulation assumptions, uncertainty quantification, and lifecycle relevance [8]. Such requirements are particularly important in materials contexts, where subtle variations in processing or measurement can significantly affect outcomes.

Emerging certification and auditing schemes for AI tools, inspired by established quality management systems, aim to verify compliance with ethical and technical norms before deployment [21]. In Materials AI, certification could encompass criteria such as dataset traceability, model interpretability thresholds, environmental impact assessments, and documentation of design intent. However, the proliferation of standards also raises concerns about accessibility and inclusivity.

Implementation gaps remain pronounced, especially in resource-limited settings where access to curated datasets, computational infrastructure, and certification processes may be constrained [12]. If standards are designed without consideration of global disparities, they risk entrenching existing inequities by privileging well-resourced institutions and regions. Responsible Materials AI, therefore, requires inclusive standard-setting, ensuring that governance mechanisms support broad participation and capacity-building rather than acting as barriers to entry.

Transparency in AI models for materials discovery

Transparency occupies a central position in the ethical and epistemic evaluation of Materials AI systems. In scientific contexts, trust in AI-generated insights depends not only on predictive accuracy but also on researchers' ability to interrogate, interpret, and contextualize model outputs. Opaque AI systems hinder this process, limiting their acceptance as legitimate contributors to scientific reasoning and materials design. **Figure 1** conceptualizes responsible Materials AI as a layered governance architecture, illustrating how transparency, standards, and risk management interact across the AI lifecycle to transform predictive systems into accountable scientific instruments.

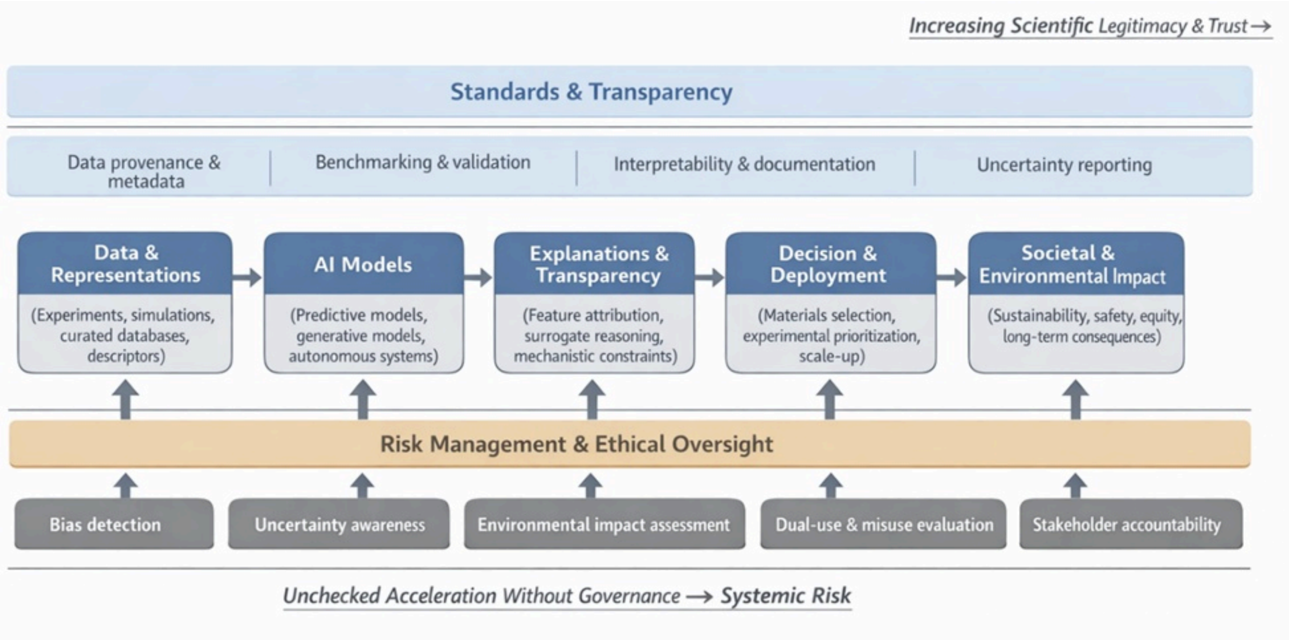

Figure 1. Governance architecture for responsible and risk-aware materials AI

Explainable AI (XAI) techniques have emerged as key instruments for enhancing transparency in materials discovery workflows. Methods such as feature attribution, sensitivity analysis, and surrogate modeling provide insights into how input variables—such as compositional features, structural descriptors, or processing parameters—influence model predictions [7]. In applications like bandgap prediction or phase stability assessment, XAI has revealed which atomic or electronic features dominate model behavior, enabling researchers to connect AI outputs with established materials knowledge and guiding rational design strategies [24].

However, transparency introduces inherent trade-offs. Highly interpretable models often rely on simplified representations or constrained architectures, which may compromise predictive performance in complex materials systems [3]. This tension between accuracy and interpretability is particularly acute in domains characterized by nonlinear, multiscale interactions. Recent advances in mechanistic AI seek to mitigate this trade-off by embedding physical laws, conservation principles, or symbolic constraints directly into learning architectures [25]. Such approaches enhance interpretability while preserving expressive power, aligning AI predictions more closely with scientific reasoning.

Transparency must also extend beyond model internals to encompass the entire data and computational pipeline. Open-source datasets, transparent preprocessing workflows, and accessible model documentation enable independent scrutiny, replication, and cumulative knowledge building [8]. In the absence of such openness, even interpretable models risk becoming epistemically fragile, as their conclusions cannot be independently validated or contested. To translate high-level ethical principles into operational practice, governance mechanisms in Materials AI can be systematically organized across the AI lifecycle, as summarized in Table 1.

Table 1. Governance dimensions for responsible Materials AI across the lifecycle

Governance dimension	Core objective	Key practices in materials AI	Risks addressed	Representative instruments
Data governance	Ensure traceability, representativeness, and integrity of materials data	Dataset documentation, provenance tracking, FAIR compliance, metadata on synthesis and processing conditions	Dataset bias, irreproducibility, hidden assumptions	FAIR data standards, curated materials repositories, data cards
Model validation and benchmarking	Establish the reliability and comparability of AI models	Standardized benchmarks, cross-dataset evaluation, and domain-of-applicability reporting	Overfitting, misleading performance claims	Community benchmarks (e.g., MatBench), validation protocols
Transparency and interpretability	Enable scientific scrutiny and trust in AI outputs	Feature attribution, surrogate models, physics-informed constraints	Black-box opacity, epistemic fragility	Explainable AI (XAI), interpretable architectures
Risk and uncertainty management	Support risk-aware decision-making	Uncertainty quantification, probabilistic predictions, risk registers	Unsafe extrapolation, overconfident recommendations	Bayesian ML, ensemble methods, uncertainty reporting
Ethical and societal oversight	Align AI-driven materials innovation with societal values	Lifecycle impact assessment, dual-use evaluation, stakeholder engagement	Environmental harm, misuse, and inequitable benefits	RRI frameworks, ethical review boards
Institutional and policy alignment	Ensure accountability beyond individual models	Certification, auditing, regulatory engagement	Governance gaps, fragmented responsibility	AI audits, emerging AI certification schemes

Taken together, transparency in Materials AI is not a singular technical feature but a systemic property emerging from the interaction of data practices, model design, and governance structures. Its realization is essential to transforming AI from a powerful predictive tool into a trustworthy, ethically aligned component of materials science.

Risk assessment and management in materials AI

Risk assessment in Materials AI extends beyond conventional concerns of model accuracy to encompass a broader spectrum of technical, epistemic, environmental, and societal vulnerabilities. At the technical level, common failure modes include model overfitting, data leakage, and inappropriate extrapolation beyond the domain of applicability. Such failures are particularly consequential in materials contexts, where AI-generated recommendations may inform costly experiments, industrial-scale production, or safety-critical deployments. Extrapolative errors—where models are applied to chemistries, processing regimes, or length scales that are insufficiently represented in the training data—pose a significant risk of generating misleading or unsafe material recommendations [1].

To address these uncertainties, probabilistic and uncertainty-aware approaches have gained prominence. Bayesian machine learning, ensemble methods, and probabilistic surrogate models explicitly quantify epistemic and aleatory uncertainty, enabling risk-informed decision-making rather than relying solely on point predictions [12]. In materials discovery workflows, uncertainty estimates can guide experimental prioritization, flag high-risk predictions, and support human oversight by distinguishing robust insights from speculative extrapolations. Importantly, uncertainty quantification also serves an epistemic function, delineating the boundaries of model knowledge and preventing the overconfidence that often accompanies high-performing black-box systems.

Beyond technical reliability, Materials AI introduces broader environmental and societal risks that require systematic assessment. AI-driven optimization may inadvertently favor solutions that maximize performance or efficiency metrics while externalizing environmental costs. For example, AI-optimized mining or materials-processing strategies could increase extraction rates, energy consumption, or waste generation, intensifying ecological degradation if sustainability constraints are not explicitly embedded in the optimization objectives [5]. Similarly, AI systems may accelerate the deployment of materials with uncertain long-term environmental or health impacts, outpacing regulatory evaluation cycles.

In response to these multifaceted risks, emerging governance strategies advocate integrating structured risk management instruments directly into Materials AI workflows. These include risk registers that catalog potential failure modes across technical, environmental, and social dimensions, as well as impact assessments conducted at multiple stages of the AI lifecycle—from data curation and model design to deployment and scale-up [7]. Such instruments shift risk management from a reactive posture to a proactive, anticipatory practice, aligning with the principles of responsible innovation.

Effective mitigation further requires multi-stakeholder oversight mechanisms that distribute responsibility across researchers, institutions, industry actors, and regulatory bodies. Regulatory engagement is particularly important in contexts where AI-driven materials decisions intersect with public safety, environmental protection, or national security considerations. Oversight frameworks that clarify accountability, mandate documentation of model assumptions, and enforce compliance with ethical and environmental standards are essential for ensuring that the accelerating capabilities of Materials AI do not outpace societal safeguards [26].

Results and Discussion

The reviewed literature collectively indicates that Materials AI has evolved from an auxiliary computational aid into an integral component of contemporary materials discovery and design pipelines. AI systems now influence not only prediction and screening tasks but also experimental strategy, hypothesis generation, and decision-making at multiple stages of the research lifecycle [1, 2, 12, 18]. Autonomous laboratories, generative models, and physics-informed learning architectures exemplify this transition, offering unprecedented acceleration and scope in materials innovation [3, 13, 16]. Yet, this rapid technical maturation has not been matched by commensurate advances in governance, oversight, and ethical integration.

A persistent barrier to broader trust and adoption lies in the opacity of many AI systems. Despite impressive predictive performance, black-box models challenge scientific norms of explanation, validation, and reproducibility [7, 21]. While explainable AI techniques and physics-informed approaches have made important strides in mitigating opacity, tensions between interpretability and predictive power remain unresolved, particularly for complex, multiscale materials systems [3, 24]. These tensions are not merely technical but epistemic, shaping how AI-generated outputs are interpreted, trusted, and acted upon within scientific practice.

Ethical frameworks adapted from general AI governance—emphasizing fairness, accountability, and non-maleficence—provide a valuable starting point but require substantial tailoring for material contexts. Bias in materials datasets tends to privilege well-characterized material classes, such as oxides or conventional alloys, while marginalizing emerging or sustainable alternatives, thereby shaping innovation trajectories in subtle but consequential ways [7, 20]. Unlike consumer-facing AI, these biases may manifest over long time horizons through cumulative environmental and industrial effects, underscoring the need for lifecycle-aware ethical assessment.

Standardization initiatives, including benchmarking efforts such as MatBench, represent important progress toward reproducibility and methodological rigor [4, 23]. However, adoption of interoperability protocols, uncertainty reporting standards, and data provenance requirements remains uneven across the field [8, 12]. Without widespread harmonization, Materials AI risks fragmenting into siloed practices that hinder cumulative knowledge building and exacerbate disparities between well-resourced and resource-limited research environments.

Transparency-oriented strategies—ranging from XAI methods to mechanistic model integration—have demonstrated their capacity to transform AI outputs into scientifically meaningful insights [3, 7, 24]. Nevertheless, scalability challenges persist, particularly as models grow in complexity and autonomy. Transparency must therefore be understood as a systemic property supported by data governance, documentation practices, and institutional norms, rather than as a feature of individual algorithms alone.

Risk management emerges as a unifying lens for addressing these challenges. Probabilistic modeling, impact assessments, and governance instruments offer pathways to mitigate both technical unreliability and broader harms, including environmental overexploitation and dual-use risks [1, 5, 7]. Yet significant gaps remain. These include limited global inclusivity in standards development, insufficient attention to long-term societal and environmental consequences, and fragmented governance across academic, industrial, and policy domains [1, 9, 10, 12]. Interdisciplinary collaboration is widely acknowledged as essential, but entrenched disciplinary silos continue to impede the development of holistic frameworks [22, 27].

Looking forward, progress in Materials AI governance will depend on reflexive, adaptive approaches that evolve alongside technical capabilities. Rather than constraining innovation, such approaches can steer AI-enabled materials discovery toward societal priorities, including sustainability, resilience, and equitable access to technological benefits. Aligning Materials AI with long-term sustainability goals—such as clean energy transitions and responsible resource use—represents both an ethical imperative and an opportunity to redefine AI as a genuinely integrative scientific instrument rather than a narrowly optimized predictive engine [14, 15, 19].

Conclusion

Artificial intelligence has emerged as a transformative force in materials science, reshaping how materials are discovered, evaluated, and deployed across critical technological domains. By enabling accelerated exploration of complex design spaces and the integration of heterogeneous data sources, Materials AI offers unprecedented opportunities to address global challenges in energy sustainability, environmental resilience, healthcare, and advanced manufacturing. Yet, as this review has demonstrated, the realization of these benefits is contingent not only on technical sophistication but on the establishment of governance structures that ensure ethical alignment, scientific legitimacy, and societal trust.

Responsible governance in Materials AI must be understood as a foundational requirement rather than an auxiliary consideration. Robust standards, meaningful transparency, and proactive risk management are indispensable for translating AI-driven acceleration into durable and equitable innovation. Without such safeguards, Materials AI risks reinforcing existing biases, obscuring scientific reasoning, and externalizing environmental and societal costs, thereby undermining its long-term value.

Drawing on the reviewed literature, several strategic priorities emerge. First, there is a clear need for community-endorsed standards governing data provenance, model validation, uncertainty quantification, and reporting practices. Building on existing benchmarking initiatives, these standards should explicitly define domains of applicability and promote reproducibility across institutional and geographic boundaries. Second, interpretability must be elevated as a scientific objective, not merely a technical add-on. Prioritizing explainable AI approaches and hybrid physics–machine learning models can enhance mechanistic insight, support hypothesis generation, and align AI outputs with established materials reasoning practices. Third, comprehensive risk governance frameworks should be embedded throughout the Materials AI lifecycle. Such frameworks must extend beyond technical failure mitigation to incorporate environmental lifecycle assessments, dual-use considerations, and structured stakeholder engagement.

Fourth, open and FAIR-compliant infrastructures are essential for democratizing access to Materials AI capabilities and mitigating systemic biases arising from uneven data availability or computational resources. Open datasets, transparent algorithms, and shared evaluation protocols can foster inclusivity while enabling independent scrutiny and cumulative knowledge building. Finally, interdisciplinary education and sustained policy dialogue are critical for aligning Materials AI development with societal values. Training programs that bridge materials science, data science, ethics, and policy can cultivate a generation of practitioners equipped to navigate both technical complexity and normative responsibility.

Looking ahead, future research and governance efforts should focus on developing autonomous yet ethically constrained laboratories in which AI systems operate within explicitly defined safety, sustainability, and accountability boundaries. Integrating environmental and social sustainability metrics directly into AI objective functions represents a promising direction for aligning optimization with long-term societal goals. Equally important is the pursuit of international harmonization of governance frameworks, which can prevent regulatory fragmentation and ensure that Materials AI evolves as a globally responsible scientific enterprise rather than a patchwork of incompatible practices.

In sum, the future of Materials AI depends on its evolution from a powerful predictive engine into a trustworthy scientific paradigm. By embedding responsibility from model design and data curation through deployment and scale-up, Materials AI can fulfill its promise as a catalyst for innovation that not only accelerates discovery but also safeguards human well-being and environmental integrity.

Acknowledgements
None

Conflict of interest
None

Financial support
None

Ethics statement
None

Received: 10 May 2024 Revised: 22 Jun 2024 Accepted: 20 Jul 2024
Published online: 18 January 2025

Rights and permissions

References

DeCost BL, Hattrick-Simpers JR, Trautt Z, Kusne AG, Campo E, Green ML. Scientific AI in materials science: A path to a sustainable and scalable paradigm. Mach Learn Sci Technol. 2020;1(1):013001.

Batra R, Song L, Ramprasad R. Emerging materials intelligence ecosystems propelled by machine learning. Nat Rev Mater. 2021;6(8):655-78.

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

Chen C, Ong SP, Persson K. Machine learning for materials scientists: An introductory guide toward best practices. Chem Mater. 2020;32(14):5933-45.

Merchant A, Batzner S, Schoenholz SS, Aykol M, Cheon G, Cubuk ED. Scaling deep learning for materials discovery. Nature. 2023;624(7990):80-5.

Zhang Y, Ling C. A strategy to apply machine learning to small datasets in materials science. npj Comput Mater. 2018;4(1):25.

Oviedo F, Ferres JL, Buonassisi T, Butler KT. Interpretable and explainable machine learning for materials science and chemistry. Acc Mater Res. 2022;3(5):597-607.

Himanen L, Geilhufe M, Clark RM. Materials data science and how FAIR data management can enable it. Adv Theory Simul. 2021;4(5):2000243.

Himanen L, Geilhufe M, Clark RM. Data-driven materials science: status, challenges, and perspectives. Adv Sci. 2019;6(21):1900808.

Ramprasad R, Batra R, Pilania G, Mannodi-Kanakkithodi A, Kim C. Machine learning in materials science. npj Comput Mater. 2017;3(1):1-13.

Jain A, Ong SP, Hautier G, Chen W, Richards WD, Dacek S, et al. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. APL Mater. 2013;1(1):011002.

Wang AY, Murdock R, Kauwe S, Oliynyk AO, Gurlo A, Brgoch J, et al. Machine learning for materials scientists: An introductory guide toward best practices. Chem Mater. 2020;32(12):4954-65.

Dan Y, Zhao Y, Li X, Li S, Hu M, Hu J. Generative deep learning for inorganic materials discovery. Matter. 2020;3(4):1074-92.

Xie T, Grossman JC. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. Phys Rev Lett. 2018;120(14):145301.

Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T. The rise of deep learning in drug discovery. Drug Discov Today. 2018;23(6):1241-50.

MacLeod BP, Parlane FGL, Morrissey TD, Häse F, Roch LM, Dettelbach KE, et al. Self-driving laboratory for accelerated discovery of thin-film materials. Sci Adv. 2020;6(20):eaaz8867.

https://doi.org/10.1126/sciadv.aaz8867.

Burger B, Maffettone PM, Gusev VV, Aitchison CM, Bai Y, Wang X, et al. A mobile robotic chemist. Nature. 2020;583(7815):237-41. https://doi.org/10.1038/s41586-020-2442-2.

Butler KT, Frost JM, Skelton JM. Interpretable machine learning models for predicting the electronic structure of high-entropy alloys. Phys Rev Mater. 2021;5(2):025002.

Unke OT, Chmiela S, Gastegger M, Schütt KT, Sauceda HE, Müller KR. SpookyNet: Learning force fields with electronic degrees of freedom and nonlocal effects. Nat Commun. 2021;12(1):7273. https://doi.org/10.1038/s41467-021-27504-0.

Wilary DM, Butler KT. Interpretable machine learning models for predicting the electronic structure of high-entropy alloys. Phys Rev Mater. 2021;5(2):025002.

OECD. OECD Principles on Artificial Intelligence. 2019.

Owen R, Macnaghten P, Stilgoe J. Responsible research and innovation: From science in society to science for society, with society. Sci Public Policy. 2012;39(6):751-60.

Dunn A, Wang Q, Jain A. Matbench 1.0: An updated materials benchmark for machine learning. Comput Mater Sci. 2020;184:109849.

ISO/IEC. ISO/IEC TR 24028:2020 Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence. 2020.

Jha D, Ward L, Paul A, Liao W, Choudhary A, Chris Wolverton C, et al. ElemNet: Deep learning the chemistry of materials from only elemental composition. Sci Rep. 2018;8(1):17593.

Karniadakis GE, Kevrekidis IG, Lu L, et al. Physics-informed machine learning. Nat Rev Phys. 2021;3(6):422-40.

Stilgoe J, Owen R, Macnaghten P. Developing a framework for responsible innovation. Res Policy. 2013;42(9):1568-80.