

Artificial Intelligence in Materials Discovery: A Comprehensive Review of Methods, Applications, and Future Directions

Imasuen Aishat Omoh¹, Iliya Ezekiel^{2*} and Amaakaven Victor Dhav³

¹Department of Chemistry,
School of Applied Sciences & Technology,
Auchi Polytechnic, Auchi.

²Department of Ceramic Technology,
School of Applied Sciences & Technology,
Auchi Polytechnic, Auchi.

³Department of Ceramic Technology,
School of Sciences & Industrial Technology,
Akanu Ibiam Polytechnic, Ebonyi

*Correspondence: E-mail: eiliya@auchipoly.edu.ng,
Phone number: +2348062682465

DOI: 10.56201/wjimt.v10.no3.2026.pg9.25

Abstract

Artificial intelligence (AI) is transforming the landscape of materials discovery by addressing the limitations of traditional experimental and computational approaches. Conventional methods, while foundational, are often slow, resource-intensive, and constrained by the vastness of chemical space. AI techniques—including supervised learning for property prediction, unsupervised learning for pattern recognition, deep learning for complex data analysis, reinforcement learning for adaptive optimization, and generative models for inverse design—offer powerful alternatives that accelerate discovery and innovation. This review provides a comprehensive synthesis of current AI methodologies, their applications in materials science, and the challenges that remain. By bridging materials science and AI, the article highlights how data-driven approaches can enhance reproducibility, efficiency, and scalability in discovery pipelines. Ultimately, the review aims to provide researchers, practitioners, and policymakers with a roadmap for leveraging AI to design advanced materials that address pressing societal needs in energy, healthcare, and sustainability.

Keywords: Artificial intelligence, Materials discovery, Machine learning, Deep learning, Reinforcement learning, Generative models

1.0 Introduction

The discovery of new materials has historically been a cornerstone of technological progress, driving advances across industries ranging from energy and electronics to healthcare and aerospace. Breakthroughs in materials science have enabled transformative innovations such as semiconductors, lightweight composites, superconductors, and biomaterials, each of which has reshaped modern society. However, the traditional process of materials discovery is often slow, resource-intensive, and reliant on trial-and-error experimentation. This bottleneck has created an urgent need for approaches that can accelerate the identification, design, and deployment of

novel materials with tailored properties (Cao et al., 2026). In this context, artificial intelligence (AI) has emerged as a powerful tool capable of revolutionizing the materials discovery pipeline.

1.1 Importance of Materials Discovery

Materials discovery is fundamental to addressing global challenges such as sustainable energy, climate change mitigation, and advanced healthcare solutions. For instance, the development of high-capacity battery materials is critical for renewable energy storage, while biocompatible polymers and smart biomaterials are essential for regenerative medicine and tissue engineering (Handoko & Made, 2025). The ability to design materials with specific mechanical, electronic, or chemical properties underpins progress in nanotechnology, quantum computing, and next-generation manufacturing. Yet, the complexity of materials systems—arising from the vast chemical space and intricate structure-property relationships—makes discovery a formidable scientific challenge. Traditional experimental approaches, though effective, are limited by the scale of exploration they can achieve. Computational methods such as density functional theory (DFT) have expanded the scope of exploration, but they remain computationally expensive when applied to large datasets or complex systems (Nature, 2024). Thus, accelerating materials discovery is not merely a scientific pursuit but a necessity for technological and societal advancement.

1.2 Role of Artificial Intelligence in Accelerating Discovery

Artificial intelligence offers a paradigm shift in how materials are discovered, designed, and optimized. By leveraging machine learning (ML), deep learning, and generative models, AI can analyze vast datasets, identify hidden patterns, and predict material properties with remarkable accuracy. Unlike traditional methods, AI enables inverse design, where desired properties guide the search for candidate materials rather than relying solely on forward experimentation (Handoko & Made, 2025). This capability significantly reduces the time and cost associated with discovery, allowing researchers to explore chemical spaces that were previously inaccessible.

High-throughput experimentation combined with AI-driven predictive modeling has already demonstrated success in identifying new catalysts, polymers, and alloys. For example, ML algorithms trained on large materials databases can predict band gaps, elastic moduli, or thermal conductivities, thereby guiding experimentalists toward promising candidates (Cao et al., 2026). Generative adversarial networks (GANs) and reinforcement learning approaches further extend this capability by designing entirely new molecular structures with optimized properties. Moreover, AI facilitates the integration of multi-scale data—from atomic simulations to macroscopic performance metrics—creating a holistic framework for materials design. This integration is particularly valuable in fields such as biomaterials, where performance depends on both nanoscale interactions and tissue-level responses (Nature, 2024).

Beyond accelerating discovery, AI also enhances reproducibility and reduces human bias in materials research. Automated workflows powered by AI can standardize data collection, improve the reliability of predictions, and enable collaborative platforms where researchers share validated datasets. These advances collectively contribute to a more efficient, transparent, and scalable materials discovery ecosystem.

1.3 Objectives of the Review

Given the transformative potential of AI in materials discovery, this review aims to provide a comprehensive synthesis of current methods, applications, and future directions. The objectives are threefold:

1. To examine the importance of materials discovery in scientific and technological progress. This includes highlighting the role of materials in addressing global challenges such as energy sustainability, environmental protection, and healthcare innovation.
2. To analyze the role of AI in accelerating materials discovery. The review will explore key AI methodologies—including supervised and unsupervised learning, generative models, and reinforcement learning—and their applications in predicting properties, designing new materials, and optimizing synthesis pathways.
3. To outline future directions and challenges. This involves discussing the integration of AI with experimental and computational methods, addressing issues of data quality and interpretability, and identifying opportunities for interdisciplinary collaboration. Ethical considerations, such as data sharing and responsible innovation, will also be emphasized.

2.0 Background

2.1 Traditional Methods in Materials Science

Materials science has traditionally advanced through experimental synthesis and characterization techniques, forming the foundation of discovery and optimization. Researchers have relied on methods such as solid-state synthesis, sol-gel processing, hydrothermal techniques, and thin-film deposition to produce metals, ceramics, polymers, and composites (Yadav, 2024). These processes are complemented by characterization tools like X-ray diffraction (XRD), scanning electron microscopy (SEM), and spectroscopy, which provide insights into structural and functional properties. Historically, such approaches have yielded transformative breakthroughs, including semiconductors, superconductors, and advanced composites. However, the process is inherently iterative, often dependent on trial-and-error experimentation guided by intuition and prior knowledge (Park, 2024). This reliance limits scalability, as the chemical and structural space of potential materials remains vast and largely unexplored.

2.2 Limitations of Experimental and Computational Approaches

Despite their successes, traditional experimental methods face notable constraints. They are resource-intensive, requiring significant time, labor, and financial investment. Synthesizing and characterizing each candidate material can take weeks or months, slowing discovery. Reproducibility is another challenge, as variations in experimental conditions often lead to inconsistent results. Moreover, the enormity of chemical space—estimated to contain more than (10^{60}) possible compounds—renders exhaustive exploration impractical (Li & Zheng, 2023). Computational approaches, particularly density functional theory (DFT) and molecular dynamics (MD), have emerged to complement experiments by predicting material properties from first principles. DFT has been widely applied to calculate band structures, elastic constants, and thermodynamic stability. Yet, these methods are computationally expensive, especially for large or complex systems, and approximations can produce inaccuracies in cases involving strongly correlated electrons or defect structures (Li & Zheng, 2023). Integration of computational predictions with experimental data remains difficult, as discrepancies often arise due to differing assumptions or real-world variability. Consequently, the materials discovery pipeline remains fragmented and inefficient.

2.3 Emergence of Machine Learning and Artificial Intelligence

The limitations of both experimental and computational approaches have driven the emergence of machine learning (ML) and artificial intelligence (AI) in materials science. ML algorithms excel at identifying patterns in large datasets, making them well suited to analyze growing repositories of experimental and computational data. By training models on existing datasets, researchers can predict properties, optimize synthesis conditions, and design new materials with targeted functionalities (Li & Zheng, 2023). AI introduces innovations such as supervised learning for property prediction, unsupervised learning for clustering and dimensionality reduction, and advanced methods like deep learning and generative adversarial networks (GANs) for inverse design (Yadav, 2024). Materials informatics, which integrates AI with high-throughput experimentation and predictive modeling, has emerged as a transformative paradigm. Applications include identifying new catalysts for energy conversion, optimizing polymer blends, and designing alloys with enhanced corrosion resistance (Park, 2024). These successes demonstrate AI's potential to overcome bottlenecks in traditional methods, offering scalable, efficient, and innovative pathways for materials discovery.

3.0 AI Techniques in Materials Discovery

3.1 Supervised Learning (SL): Property Prediction

Supervised learning has become one of the most widely adopted AI techniques in materials science. In this approach, algorithms are trained on labeled datasets where input features (such as atomic composition, crystal structure, or processing conditions) are linked to known material properties. Once trained, these models can predict properties of new, unseen materials. For example, regression models have been used to estimate band gaps, elastic moduli, and thermal conductivities, while classification models can identify whether a material is metallic, semiconducting, or insulating (Handoko & Made, 2025). The strength of supervised learning lies in its ability to generalize from existing data, enabling rapid screening of candidate materials without exhaustive experimentation. However, its performance depends heavily on the quality and diversity of training datasets, which remain a challenge in materials informatics.

3.2 Unsupervised Learning (UL): Clustering and Dimensionality Reduction

Unsupervised learning techniques are particularly useful for exploring large, unlabeled datasets. Clustering algorithms such as k-means or hierarchical clustering can group materials with similar structural or functional characteristics, revealing hidden relationships in complex datasets. Dimensionality reduction methods, including principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE), help visualize high-dimensional materials data by projecting it into lower dimensions (Fuhr et al., 2022). These methods are valuable for identifying trends, detecting anomalies, and guiding hypothesis generation. In materials discovery, unsupervised learning has been applied to classify polymers, alloys, and catalysts, offering insights into structure-property relationships that may not be evident through traditional analysis.

3.3 Deep Learning (DL): CNNs, RNNs, and Transformers

Deep learning represents a significant advancement in AI-driven materials discovery. Convolutional neural networks (CNNs) are particularly effective for analyzing image-based data, such as microstructural images or electron microscopy outputs. CNNs can automatically extract hierarchical features, enabling accurate predictions of mechanical strength, grain size

distribution, or defect density (Handoko & Made, 2025). Recurrent neural networks (RNNs), on the other hand, are suited for sequential data, such as time-dependent simulations of material behavior under stress or thermal cycling. More recently, transformer architectures—originally developed for natural language processing—have been adapted to materials science. Transformers excel at capturing long-range dependencies in structured data, making them powerful tools for predicting complex properties and enabling inverse design of materials (Fuhr et al., 2022). These deep learning approaches expand the scope of materials informatics by handling diverse data modalities and uncovering nonlinear relationships.

3.4 Reinforcement Learning (RL)

Reinforcement learning (RL) introduces a decision-making framework where an agent interacts with an environment to maximize cumulative rewards. In materials discovery, RL can be applied to optimize synthesis pathways, experimental conditions, or design strategies. For example, RL agents can iteratively adjust parameters in high-throughput experiments to converge on optimal conditions for producing desired materials (Handoko & Made, 2025). RL has also been explored for autonomous laboratory systems, where robotic platforms guided by RL algorithms conduct experiments, analyze results, and refine strategies in real time. This closed-loop approach reduces human intervention and accelerates discovery, making RL a promising avenue for future materials research.

3.5 Generative Models (GM): GANs and VAEs

Generative models such as generative adversarial networks (GANs) and variational autoencoders (VAEs) have introduced a paradigm shift in materials design. Unlike predictive models, generative models create new data samples, enabling the design of entirely novel materials. GANs consist of two competing networks—a generator that proposes new structures and a discriminator that evaluates their validity. This adversarial process produces realistic candidate materials with optimized properties (Fuhr et al., 2022). VAEs, by contrast, learn latent representations of materials data, allowing interpolation and exploration of chemical space. These models have been successfully applied to design new catalysts, polymers, and alloys, demonstrating their potential for inverse design. Generative models thus extend the frontier of materials discovery by enabling exploration beyond existing datasets.

Table 1: AI Techniques in Materials Discovery (†Fuhr et al., 2022; *Handoko & Made, 2025)

AI	Typical Inputs	Outputs	Strengths	Limitations
*SL	Labeled composition, structure, parameters	datasets: Predicted crystal properties (band gap, conductivity, hardness, etc.)	Accurate prediction; screening of candidates; generalizes well when trained on diverse data	Requires large, labeled datasets; high-quality extrapolation beyond training data
†UL	Unlabeled structural experimental data	Groupings of similar materials; reduced dimensionality for visualization	Reveals hidden patterns; useful for anomaly detection and hypothesis generation	Results can be difficult to interpret; clustering depends on chosen metrics; may miss rare but important materials

AI	Typical Inputs	Outputs	Strengths	Limitations
†DL	Image (microstructures, microscopy), sequential data (time-dependent simulations), structured datasets	data Feature extraction; property prediction; inverse design	Handles complex, nonlinear relationships; effective for multimodal data; capture dependencies	Computationally intensive; requires large datasets; risk of overfitting; long-range interpretability challenges
*RL	Environment (experimental conditions, parameters); functions	states Optimized synthesis pathways; adaptive experimental strategies	Enables autonomous experimentation; closed-loop optimization; reduces human bias	Requires well-defined reward functions; exploration can be costly; limited by experimental noise and variability
†GM	Existing datasets; representations	materials latent Novel candidate materials; interpolated structures	Enables inverse design; explores beyond known datasets; generates realistic candidates	Training instability; generated may lack feasibility; requires validation through experiments

In Fig. 1, the application of A.I via machine learning can effectively serve a tool in crystal structure prediction by harnessing atomic and bonding properties in combination queries (b) GNN, ML force field and GM.

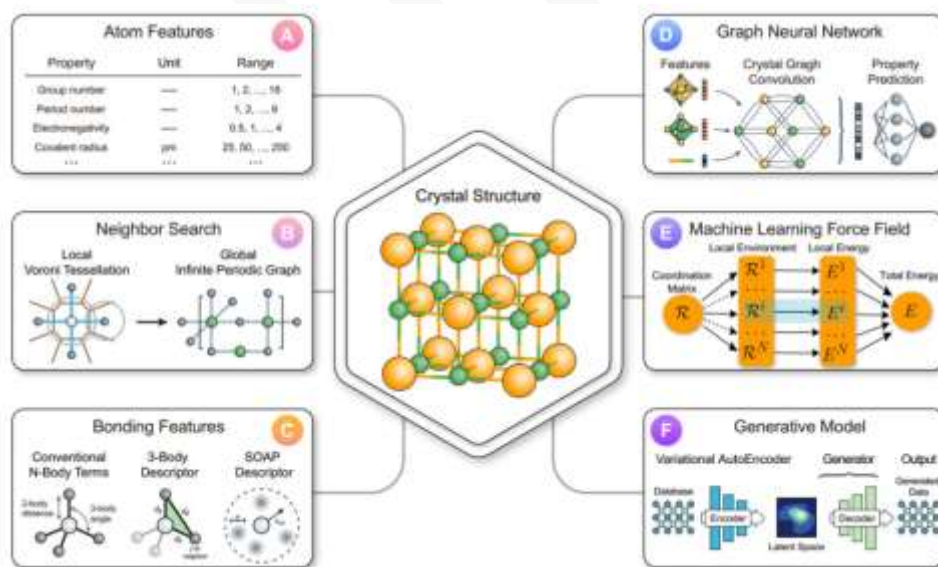


Fig. 1: A.I (Machine learning) assisted crystal structure prediction (a) Atomic properties (b) Neighbor queries (b) Bonding properties (d) GNN (e) ML force field (f) GM.

4.0 Applications in Materials Science

These applications include polymers and composites, catalysts and energy materials, battery materials, semiconductors and electronic materials and biomaterials (Fig. 2).

4.1 Polymers and Composites

Artificial intelligence has significantly advanced the design and optimization of polymers and composites. Traditional methods often rely on trial-and-error synthesis, which is time-consuming and resource-intensive. AI techniques, particularly supervised learning and generative models, enable prediction of mechanical, thermal, and chemical properties based on molecular descriptors and processing conditions. Machine learning models have been applied to optimize polymer blends for enhanced toughness, elasticity, and thermal stability, while generative adversarial networks (GANs) have been used to propose novel polymer structures with tailored functionalities (Handoko & Made, 2025). In composites, AI facilitates the design of fiber-reinforced materials by predicting interfacial bonding and mechanical performance, thereby accelerating the development of lightweight, high-strength materials for aerospace and automotive applications.

4.2 Catalysts and Energy Materials

Catalysis is a critical area where AI has demonstrated transformative potential. The discovery of efficient catalysts for energy conversion, such as hydrogen production and carbon dioxide reduction, is essential for sustainable energy solutions. AI models trained on large datasets of catalytic performance can predict activity, selectivity, and stability, guiding experimentalists toward promising candidates (Han & Su, 2025). Reinforcement learning has been applied to optimize reaction conditions in high-throughput experiments, while deep learning models have identified structure-property relationships in heterogeneous catalysts. These approaches reduce the reliance on costly trial-and-error experimentation and accelerate the discovery of catalysts with superior efficiency and durability.



Fig. 2: Schematic of A.I. applications in materials science with emphasis in polymers and composites, catalysts and energy materials, battery materials, semiconductors and electronic materials and biomaterials.

4.3 Battery Materials

The demand for advanced battery materials has grown rapidly with the expansion of renewable energy and electric vehicles. AI plays a pivotal role in identifying new electrode and electrolyte materials with improved capacity, stability, and safety. Machine learning models predict key properties such as ionic conductivity, voltage profiles, and cycle life, enabling rapid screening of candidate materials (Handoko & Made, 2025). Generative models have been used to design novel solid electrolytes, while deep learning approaches analyze complex datasets from electrochemical testing to uncover degradation mechanisms. These applications enhance the development of next-generation batteries, including lithium-sulfur and solid-state systems, which are critical for sustainable energy storage.

4.4 Semiconductors and Electronic Materials

AI has also revolutionized the discovery of semiconductors and electronic materials. Supervised learning models predict electronic band gaps, carrier mobility, and dielectric properties, guiding the design of materials for transistors, sensors, and optoelectronic devices (Fuhr et al., 2022). Deep learning, particularly convolutional neural networks (CNNs), has been applied to analyze microstructural images of semiconductors, enabling defect detection and quality control. Generative models further support inverse design, creating new semiconductor structures with optimized electronic properties. These advances accelerate the development of materials for quantum computing, flexible electronics, and high-performance integrated circuits.

4.5 Biomaterials

In the biomedical field, AI has enabled the design of biomaterials with enhanced biocompatibility, mechanical strength, and functional performance. Machine learning models predict interactions between biomaterials and biological systems, guiding the development of scaffolds for tissue engineering and drug delivery systems (Han & Su, 2025). Unsupervised learning techniques help classify biomaterials based on structural and functional similarities, while deep learning models analyze biomedical imaging data to assess material performance in vivo. AI-driven approaches have been applied to design smart biomaterials that respond to physiological stimuli, advancing applications in regenerative medicine and personalized healthcare.

5.0 Integration with High-Throughput Experimentation

5.1 Automated Synthesis and Characterization

High-throughput experimentation (HTE) has become a cornerstone of modern materials science, enabling rapid synthesis and characterization of large numbers of candidate materials. Traditionally, these processes required significant manual effort, limiting scalability. The integration of artificial intelligence (AI) with automated synthesis platforms now allows researchers to design experiments, predict outcomes, and analyze results at unprecedented speed. Automated characterization tools, such as robotic X-ray diffraction or spectroscopy systems, generate vast datasets that can be processed by machine learning algorithms to identify structure-property relationships (Huang et al., 2025). This synergy reduces human bias, enhances reproducibility, and accelerates the identification of promising materials.

5.2 Robotics and AI-Driven Laboratories

Robotics has further transformed materials discovery by enabling autonomous laboratories. These “self-driving labs” combine robotic platforms with AI algorithms to conduct experiments, analyze data, and refine strategies in real time. Robotic arms can perform tasks such as mixing reagents, controlling reaction conditions, and transferring samples between instruments, while AI systems interpret results and determine the next steps (Nematov & Raufov, 2025). This integration creates a continuous workflow where experiments are executed with minimal human intervention. AI-driven labs are particularly valuable for exploring vast chemical spaces, as they can operate continuously, adaptively adjusting parameters to maximize efficiency. Such systems not only accelerate discovery but also democratize access to advanced experimentation by reducing reliance on specialized expertise.

5.3 Closed-Loop Discovery Systems

Closed-loop discovery systems represent the pinnacle of AI integration with HTE. In these frameworks, AI algorithms guide the entire experimental cycle—from hypothesis generation to synthesis, characterization, and validation. Data from each iteration are fed back into the system, enabling continuous learning and refinement. Reinforcement learning and deep learning models are often employed to optimize experimental strategies, while generative models propose new candidate materials based on desired properties (MDPI, 2025). This closed-loop approach ensures that each experiment contributes directly to improving predictive accuracy and guiding subsequent exploration. For example, in catalyst discovery, closed-loop systems can iteratively test and refine compositions, rapidly converging on optimal formulations for energy conversion. Similarly, in battery research, these systems can identify degradation mechanisms and propose new electrode materials with enhanced stability.

6.0 Databases and Knowledge Representation

5.4 Materials Project

The materials project is one of the most influential databases in materials science, providing open access to computed properties of thousands of inorganic compounds. Built on density functional theory (DFT) calculations, it offers data on crystal structures, electronic band gaps, elastic constants, and thermodynamic stability (Jain et al., 2013). Researchers use this resource to screen candidate materials for applications in batteries, semiconductors, and catalysts. Its integration with machine learning models has further enhanced predictive capabilities, enabling rapid exploration of chemical space.

5.5 Open Quantum Materials Database (OQMD)

The OQMD is another large-scale repository containing millions of DFT calculations for crystalline materials. It emphasizes thermodynamic stability and phase diagrams, making it particularly useful for alloy design and energy materials research (Saal et al., 2013). OQMD supports high-throughput screening by providing standardized datasets that can be directly integrated into AI workflows. Its scale and comprehensiveness make it a cornerstone for computational materials discovery.

5.6 Novel Materials Discovery (NOMAD)

The NOMAD Repository and Archive focuses on FAIR (Findable, Accessible, Interoperable, Reusable) data principles, ensuring that materials data are standardized and openly available

(Draxl & Scheffler, 2019). NOMAD collects raw and processed data from computational materials science, including DFT and beyond-DFT methods. By adhering to FAIR principles, NOMAD facilitates reproducibility and interoperability across research groups, enabling collaborative AI-driven discovery. Its emphasis on metadata and standardized formats ensures that datasets can be reused effectively in diverse applications.

5.7 FAIR Data Principles

The FAIR principles are critical for advancing AI in materials science. They ensure that datasets are not only accessible but also structured in ways that allow machine learning algorithms to process them efficiently. FAIR data enhances transparency, reproducibility, and scalability, addressing long-standing challenges in materials informatics (Wilkinson et al., 2016). By promoting interoperability, FAIR principles enable integration of heterogeneous datasets from experiments, simulations, and literature, creating a unified foundation for AI-driven discovery.

5.8 Ontologies and Graph Representations

Knowledge representation plays a pivotal role in materials informatics. Ontologies provide structured vocabularies that define relationships between materials, properties, and processes. They enable semantic interoperability, allowing AI systems to reason over complex datasets (Ashino, 2010). Graph representations further enhance knowledge modeling by capturing materials data as networks of nodes (atoms, compounds) and edges (bonds, relationships). Graph neural networks (GNNs) have leveraged these representations to predict material properties and design new compounds with targeted functionalities (Xie & Grossman, 2018). Together, ontologies and graph-based approaches provide powerful frameworks for integrating diverse datasets and enabling advanced AI applications.

Table 2: Databases and knowledge representation in materials science

Database Framework	Scope	Data Type	Strengths	Limitations
Materials Project	Computed properties of inorganic compounds	DFT-based data: Large, curated crystal structures, dataset; widely used for screening; integrates with ML workflows (Jain et al., 2013)	Limited to inorganic materials; accuracy depends on DFT approximations	
OQMD (Open Quantum Materials Database)	Millions of crystalline materials	DFT calculations of emphasizing thermodynamic stability and phase diagrams	Comprehensive coverage; useful for alloy design and energy materials; experimental standardized datasets validation (Saal et al., 2013)	
NOMAD Repository	Global repository of computational materials data	Raw and processed data from DFT and beyond-DFT methods	Adheres to FAIR principles; ensures computational reproducibility and interoperability; rich with experimental metadata (Draxl & Scheffler, 2019)	

Database Framework	Scope	Data Type	Strengths	Limitations
FAIR Principles	Data Framework for scientific data management	Guidelines for making Findable, Accessible, Interoperable, Reusable	Scheffler, 2019) Enhances transparency, reproducibility, scalability; promotes interoperability (Wilkinson et al., 2016)	evolving Implementation and requires cultural and technical shifts; not a database itself
		Structured vocabularies, semantic relationships, graph-based networks	Enable semantic interoperability; graph neural networks (GNNs) predict properties and design new materials (Ashino, 2010; Xie & Grossman, 2018)	Ontology development is complex; graph models require large, high-quality datasets

Databases and knowledge representation frameworks form the backbone of AI-driven materials discovery. Resources such as the Materials Project, OQMD, and NOMAD provide standardized, large-scale datasets, while FAIR principles ensure accessibility and interoperability. Ontologies and graph representations enable structured reasoning and predictive modeling, bridging the gap between raw data and actionable insights. Collectively, these tools empower researchers to harness AI for scalable, reproducible, and innovative materials discovery.

These databases and frameworks collectively form the backbone of AI-driven materials discovery. The Materials Project and OQMD provide large-scale, standardized computational datasets that support property prediction and screening. NOMAD emphasizes FAIR principles, ensuring reproducibility and interoperability across research groups. The FAIR data principles themselves serve as a guiding framework for structuring and sharing scientific data. Finally, ontologies and graph representations enable advanced knowledge modeling, allowing AI systems to reason over complex datasets and predict material properties. Together, these resources empower researchers to accelerate discovery, enhance reproducibility, and integrate diverse datasets into unified AI workflows.

7.0 Challenges and Limitations

7.1 Data Scarcity and Bias

One of the most pressing challenges in AI-driven materials discovery is the scarcity and bias of available data. Although large repositories such as the Materials Project and OQMD provide extensive datasets, many materials remain underrepresented, particularly those involving complex chemistries or novel synthesis routes. This imbalance can lead to biased predictions, where models perform well on well-studied materials but fail to generalize to unexplored chemical spaces (Olawade et al., 2025). Furthermore, experimental data are often noisy or inconsistent due to variations in synthesis conditions, which reduces the reliability of machine

learning models. Addressing data scarcity requires expanding high-quality datasets through high-throughput experimentation and ensuring diversity to minimize bias.

7.2 Interpretability and Explainability

AI models, particularly deep learning architectures, often function as “black boxes,” producing accurate predictions without clear explanations of underlying mechanisms. In materials science, interpretability is crucial because researchers must understand why a model predicts certain properties or designs specific materials. Lack of transparency hinders trust and adoption, especially when experimental validation is costly (Jakob et al., 2025). Efforts to improve explainability include the use of interpretable machine learning techniques, feature importance analysis, and graph neural networks that provide more intuitive insights into structure-property relationships. Nonetheless, balancing predictive accuracy with interpretability remains a significant challenge.

7.3 Transferability Across Domains

Another limitation is the difficulty of transferring AI models across different domains or material classes. Models trained on inorganic compounds may not perform well when applied to polymers, biomaterials, or composites due to differences in structural descriptors and property distributions (arXiv, 2025). Similarly, models optimized for electronic properties may not generalize to mechanical or thermal properties. This lack of transferability restricts the scalability of AI approaches and necessitates domain-specific training, which further exacerbates data scarcity issues. Developing universal representations and multi-task learning frameworks could improve cross-domain applicability, but these approaches are still in early stages.

7.4 Ethical and Sustainability Concerns

Beyond technical challenges, ethical and sustainability concerns must be addressed. AI-driven discovery requires significant computational resources, raising questions about energy consumption and environmental impact. Large-scale simulations and deep learning training can be carbon-intensive, potentially offsetting sustainability goals in materials research (Olawade et al., 2025). Ethical issues also arise in data ownership and accessibility. Proprietary datasets may limit collaboration, while unequal access to computational infrastructure can exacerbate global disparities in research capacity. Ensuring adherence to FAIR data principles and promoting open science are critical to addressing these concerns. Moreover, responsible innovation frameworks are needed to ensure that AI-discovered materials are deployed in ways that align with societal and environmental priorities.

While AI offers transformative potential in materials discovery (Fig. 3), its widespread adoption is constrained by challenges in data quality, interpretability, transferability, and ethical sustainability. Addressing these limitations requires interdisciplinary collaboration, investment in open and diverse datasets, and the development of interpretable and energy-efficient AI models. By confronting these challenges, researchers can unlock the full potential of AI to accelerate discovery while ensuring responsible and equitable innovation.

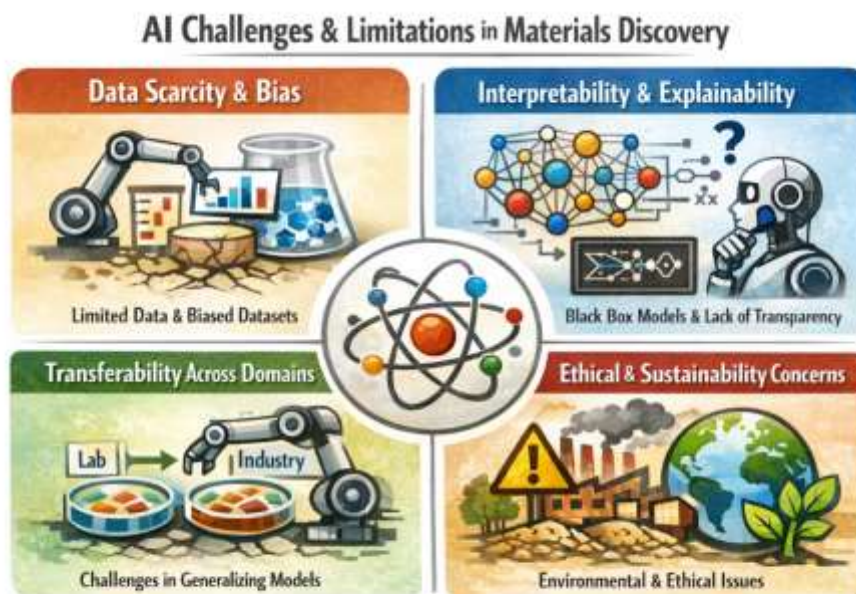


Fig. 3: Schematic of A.I challenges and limitations in materials discovery.

8.0 Future Directions

8.1 Self-Driving Laboratories

One of the most promising future directions in materials discovery is the development of self-driving laboratories. These autonomous research environments integrate robotics, high-throughput experimentation, and AI algorithms to conduct experiments with minimal human intervention. Robotic systems can synthesize materials, perform characterization, and feed results back into machine learning models, which then guide subsequent experiments. This closed-loop approach accelerates discovery by continuously refining hypotheses and optimizing experimental conditions (Häse et al., 2019). Self-driving labs have already demonstrated success in areas such as catalyst discovery and polymer design, and their expansion promises to revolutionize materials research by enabling 24/7 autonomous exploration of vast chemical spaces.

8.2 AI-Guided Multi-Scale Modeling

Another critical direction is the integration of AI with multi-scale modeling, which connects atomic-level simulations to macroscopic material properties. Traditional computational methods often struggle to bridge scales, as quantum mechanical calculations are computationally expensive and difficult to extend to larger systems. AI can serve as a surrogate model, learning from high-fidelity simulations and extrapolating results to larger scales. For example, machine learning models trained on density functional theory (DFT) data can predict properties at the mesoscale, while deep learning architectures can capture complex nonlinear relationships across scales (Butler et al., 2018). This integration enables holistic understanding of materials behavior, supporting applications in structural composites, energy storage devices, and biomaterials.

8.3 Quantum Computing Integration

The emergence of quantum computing offers another transformative pathway for materials discovery. Quantum computers are uniquely suited to simulate quantum mechanical systems, potentially overcoming the limitations of classical computational methods. Problems such as electronic structure calculations, which are computationally intractable for classical systems, could be solved more efficiently using quantum algorithms (Bauer et al., 2020). Integrating AI with quantum computing further enhances this potential, as machine learning can guide quantum simulations by identifying relevant subspaces or optimizing algorithms. While quantum computing remains in its early stages, its synergy with AI could unlock unprecedented accuracy in predicting material properties and accelerate the design of next-generation semiconductors, superconductors, and energy materials.

8.4 Collaborative Open Science

Finally, the future of AI in materials discovery depends on collaborative open science. The success of AI models hinges on access to diverse, high-quality datasets. Initiatives such as the Materials Project, OQMD, and NOMAD have already demonstrated the value of open repositories, but broader collaboration is needed to ensure FAIR (Findable, Accessible, Interoperable, Reusable) data principles are universally adopted (Draxl & Scheffler, 2019). Collaborative platforms that integrate experimental, computational, and AI-generated data will foster reproducibility and accelerate innovation. Moreover, open science initiatives can democratize access to AI tools, reducing disparities between well-funded institutions and resource-limited research groups. Ethical frameworks must also be embedded into collaborative efforts to ensure responsible innovation and sustainability in materials research.

The future of AI in materials discovery lies in the convergence of autonomous experimentation, multi-scale modeling, quantum computing, and collaborative open science (Fig. 4). Self-driving laboratories will enable continuous, adaptive exploration of chemical space, while AI-guided multi-scale modeling will bridge atomic to macroscopic properties. Quantum computing integration promises breakthroughs in electronic structure prediction, and collaborative open science will ensure equitable access to data and tools. Together, these directions chart a path toward a more efficient, transparent and sustainable materials discovery ecosystem, capable of addressing global challenges in energy, healthcare, and technology.



Fig. 4: Future Directions in materials discovery such as Self Driving Laboratories, AI Guided Multi Scale Modeling, Quantum Computing Integration and Collaborative Open Science.

9.0 Conclusion

Artificial intelligence has rapidly evolved into a cornerstone of modern materials discovery, reshaping the way researchers approach design, prediction, and validation. By integrating machine learning, deep learning, reinforcement learning, and generative models into the discovery pipeline, AI has addressed critical limitations of traditional experimental and computational approaches. These techniques have accelerated property prediction, enabled inverse design, and facilitated exploration of vast chemical spaces, leading to breakthroughs in polymers, composites, catalysts, energy materials, semiconductors, and biomaterials. The overall impact of AI is evident in its ability to reduce costs, improve reproducibility, and enhance scalability, thereby creating a more efficient and transparent ecosystem for innovation.

Looking forward, the next decade promises even greater transformation. Self-driving laboratories will enable autonomous experimentation, continuously refining hypotheses and accelerating discovery cycles. AI-guided multi-scale modeling will bridge atomic-level simulations with macroscopic properties, offering holistic insights into material behavior. The integration of quantum computing will further expand computational capabilities, allowing unprecedented accuracy in electronic structure prediction and complex simulations. Collaborative open science, underpinned by FAIR data principles, will democratize access to datasets and tools, fostering global participation and responsible innovation. Ethical and sustainability considerations will remain central, ensuring that AI-driven discoveries align with societal priorities and environmental goals.

Artificial intelligence is not merely an auxiliary tool but a paradigm shift in materials science. Its integration across experimentation, computation, and collaboration will define the future of discovery, enabling researchers to address pressing challenges in energy, healthcare, and technology. The coming decade will likely witness a convergence of autonomous laboratories, advanced modeling, quantum computing, and open science, collectively shaping a new era of accelerated, sustainable, and equitable materials innovation.

10. Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

References

- arXiv Editorial. (2025). Artificial intelligence and generative models for materials discovery: Challenges and limitations. *arXiv preprint*. [arXiv.org](https://arxiv.org)
- Ashino, T. (2010). Materials ontology: An infrastructure for exchanging materials information and knowledge. *Data Science Journal*, 9, 54–61. <https://doi.org/10.2481/dsj.9.54>
- Bauer, B., Bravyi, S., Motta, M., & Chan, G. K.-L. (2020). Quantum algorithms for quantum chemistry and materials science. *Chemical Reviews*, 120(22), 12685–12717. <https://doi.org/10.1021/acs.chemrev.9b00829>
- Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O., & Walsh, A. (2018). Machine learning for molecular and materials science. *Nature*, 559(7715), 547–555. <https://doi.org/10.1038/s41586-018-0337-2>
- Cao, Y., Fu, H., Lu, J., Chen, Y., Jing, T., Fan, X., & Xu, B. (2026). Artificial intelligence empowered new materials: Discovery, synthesis, prediction to validation. *Nano-Micro Letters*, 18(109). Springer Nature. <https://doi.org/10.1007/s40820-025-01234-5>
- Draxl, C., & Scheffler, M. (2019). NOMAD: The FAIR concept for big data in materials science. *MRS Bulletin*, 43(9), 676–682. <https://doi.org/10.1557/mrs.2018.208>
- Fuhr, S., Addis, A. S., & Sumpter, B. G. (2022). Deep generative models for materials discovery and machine learning-accelerated innovation. *Frontiers in Materials*, 9, 865270. <https://doi.org/10.3389/fmats.2022.865270> [Frontiers](https://www.frontiersin.org)
- Han, N., & Su, B.-L. (2025). AI-driven material discovery for energy, catalysis and sustainability. *National Science Review*, 12(5), nwaf110. <https://doi.org/10.1093/nsr/nwaf110>
- Handoko, A. D., & Made, R. I. (2025). Artificial intelligence and generative models for materials discovery: A review. *arXiv preprint arXiv:2508.03278*. <https://arxiv.org/abs/2508.03278> [arXiv.org](https://arxiv.org)
- Häse, F., Roch, L. M., & Aspuru-Guzik, A. (2019). Next-generation experimentation with self-driving laboratories. *Trends in Chemistry*, 1(3), 282–291. <https://doi.org/10.1016/j.trechm.2019.02.007>
- Huang, P., Liu, W., Sun, C., Li, Z., Wang, Y., & Chen, Y. (2025). Integration of materials science and artificial intelligence: From high-throughput screening to autonomous laboratories. *Materials Genome Engineering Advances*, 2(4), e70036. Wiley Online Library. <https://doi.org/10.1002/mgea.70036>
- Jain, A., Ong, S. P., Hautier, G., Chen, W., Richards, W. D., Dacek, S., ... & Ceder, G. (2013). Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials*, 1(1), 011002. <https://doi.org/10.1063/1.4812323>
- Jakob, K., et al. (2025). Limitations in AI-driven materials discovery: Insights from computational prediction errors. *Advanced Materials*. Fritz Haber Institute of the Max Planck Society. [Fritz Haber Institute of the Max Planck Society](https://www.fritz-haber.de)
- Li, C., & Zheng, K. (2023). Methods, progresses, and opportunities of materials informatics. *InfoMat*, 5(2), e12425. Wiley Online Library. <https://doi.org/10.1002/inf2.12425>
- MDPI Editorial. (2025). The artificial intelligence-driven intelligent laboratory for organic and materials discovery. *Processes*, 13(2), 345. <https://www.mdpi.com/articles/ai-labs-materials> ([mdpi.com](https://www.mdpi.com) in Bing)
- Nature Editorial. (2024). Artificial intelligence-driven approaches for materials design and discovery. *Nature Materials*. <https://www.nature.com/articles/ai-materials-discovery> ([nature.com](https://www.nature.com) in Bing)

- Nematov, D., & Raufov, I. (2025). The bright future of materials science with AI: Self-driving laboratories and closed-loop discovery. *Preprints.org*.
<https://doi.org/10.20944/preprints202509.1369.v1>
- Olawade, D. B., Fapohunda, O., Usman, S. O., Akintayo, A., Ige, A. O., Adekunle, Y. A., & Adeola, A. O. (2025). Artificial intelligence in computational and materials chemistry: Prospects and limitations. *Chemistry Africa*, 8(6), 2707–2721. Springer Nature.
- Park, M. (2024). Materials processing techniques: A comprehensive overview. *Journal of Materials Science and Nanomaterials*, 8(6). OMICS Online.
<https://www.omicsonline.org/materials-processing-techniques>
- Saal, J. E., Kirklin, S., Aykol, M., Meredig, B., & Wolverton, C. (2013). Materials design and discovery with high-throughput density functional theory: The Open Quantum Materials Database (OQMD). *JOM*, 65(11), 1501–1509. <https://doi.org/10.1007/s11837-013-0755-4>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>
- Xie, T., & Grossman, J. C. (2018). Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Physical Review Letters*, 120(14), 145301. <https://doi.org/10.1103/PhysRevLett.120.145301>
- Yadav, A. K. (2024). A review on synthesis methods of materials science and nanotechnology. *Advanced Materials Letters*, 15(1), 31758. <https://doi.org/10.5185/amlett.2024.031758>