ISSN (Online): 3107-5991

International Journal for Research Communications in Engineering, Emerging Technologies and Sciences (IJRCEETS)
Volume 1, Issue 2, November-2025

Cite: P Valli Rani, P Nithya Sri, Reshma Aman Abdul (2025), AI-Driven Materials for High-Performance Semiconductors, International Journal for Research Communications in Engineering, Emerging Technologies and Sciences, Vol. 01, issue (2), pp. 79-90

AI-Driven Materials for High-Performance Semiconductors

P Valli Rani[1*], P Nithya Sri [2,] Reshma Aman Abdul [3]
[1]Associate Professor, Ramachandra College of Engineering, Eluru-534007, Andhra Pradesh, India
[2,3]Assistant Professor, Ramachandra College of Engineering, Eluru-534007, Andhra Pradesh, India

Article info:
Article No.: IJRCEETSV1I20006
Submitted: 24/09/2025
Received in revised form: 15/10/2025
Accepted for publication: 27/10/2025
Available online: 14/11/2025

***Corresponding email:**
valliprasad02@rcee.ac.in

Abstract. The accelerating demand for next-generation electronic and optoelectronic devices has placed unprecedented pressure on the discovery of novel semiconductor materials with tailored properties. Traditional trial-and-error experimental approaches are no longer sufficient given the vast chemical space of potential candidates. This review examines how artificial intelligence (AI) and machine learning (ML) techniques—encompassing graph neural networks (GNNs), generative adversarial networks (GANs), variational autoencoders (VAEs), and Bayesian optimization—are reshaping the landscape of the discovery of semiconductor materials. We survey the integration of these methods with high-throughput density functional theory (DFT) calculations and curated materials databases to accelerate the identification of candidates with optimal bandgaps, carrier mobilities, and defect tolerances. Case studies spanning perovskite photovoltaics, two-dimensional transition metal dichalcogenides (TMDs), III-V alloys, and wide-bandgap nitrides demonstrate 10–100x reductions in discovery timelines. Challenges related to data scarcity, model interpretability, and experimental validation are critically assessed, and an outlook on the convergence of autonomous robotic synthesis with AI feedback loops is presented. This work provides a comprehensive technical framework for researchers seeking to leverage AI for high-performance semiconductor design.

Keywords: machine learning; semiconductor; bandgap engineering; graph neural network; perovskite; high-throughput screening; density functional theory; generative models; materials informatics

1. Introduction

Semiconductors are the foundational materials of modern civilization, underpinning technologies ranging from transistors and solar cells to light-emitting diodes (LEDs) and quantum computing components. The global semiconductor market exceeded USD 620 billion in 2024 and is projected to surpass USD 1 trillion by 2030, driven by the exponential growth of artificial intelligence hardware, electric vehicles, and renewable energy infrastructure. This extraordinary demand necessitates materials with increasingly precise property profiles—specific bandgaps for photon absorption, ultrahigh carrier mobilities for fast switching, thermal stability for high-power operation, and earth-abundant compositions for sustainable manufacturing.

Historically, semiconductor discovery has proceeded through decades of empirical experimentation guided by chemical intuition. The journey from the observation of the semiconducting properties of silicon in the 1940s to the commercialization of gallium nitride (GaN) LEDs spanned nearly 50 years. In an era of climate urgency and digital transformation, such timescales are no longer acceptable. High-throughput computational screening using DFT offered a first generation of acceleration, enabling the evaluation of thousands of compounds; however, even DFT calculations are computationally prohibitive at the scale of the estimated 10^8 synthesizable inorganic compounds.

The emergence of materials informatics — the application of data science and AI to materials research — has led to a transformative paradigm shift. By training ML models on existing experimental and computational datasets, researchers can now predict material properties orders of magnitude faster than first-principles methods can, enabling the screening of millions of candidates in hours. This review provides a comprehensive technical overview of AI-driven semiconductor discovery, covering data infrastructure, model architectures, active learning strategies, and experimental validation pathways (Figure.1).

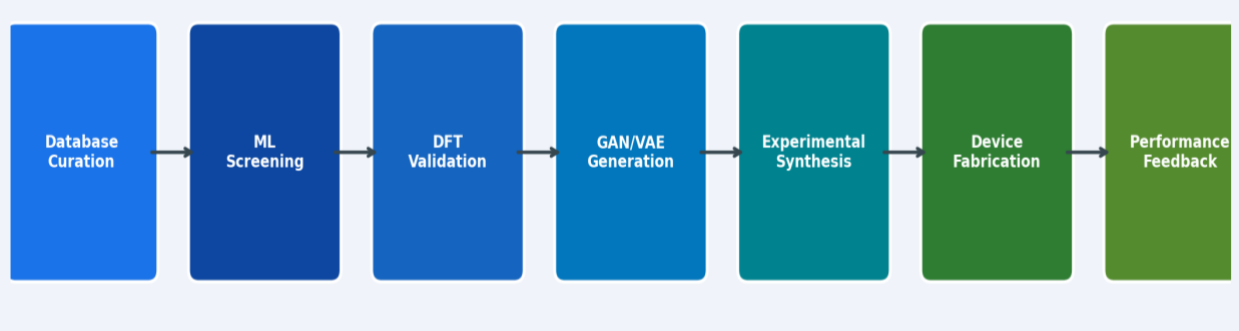

Figure 1. Schematic of the AI-driven closed-loop material discovery pipeline for high-performance semiconductors. The workflow integrates database curation, ML screening, DFT validation, generative model design, experimental synthesis, and device characterization within an active learning feedback cycle.

2. Data Infrastructure and Materials Databases

AI-driven material discovery leverages machine learning and data-driven models to identify novel semiconductor materials with enhanced electrical, thermal, and optical properties. By analysing large datasets of material compositions and structures, AI can predict bandgap energies, carrier mobility, and defect tolerance with high accuracy. Techniques such as deep learning and generative models accelerate the exploration of unexplored material spaces, reducing experimental costs and time. This approach enables the rapid development of high-performance semiconductors for next-generation electronics, energy devices, and quantum technologies.

2.1 Curated Materials Databases

The quality and breadth of training data fundamentally affect the performance of any ML model. Several large-scale databases have been instrumental in enabling semiconductor materials informatics. The Materials Project (MP), hosted at Lawrence Berkeley National Laboratory, contains DFT-computed properties for more than 150,000 inorganic compounds, including electronic band structures, elastic tensors, and dielectric constants. The AFLOW (Automatic Flow for Materials Discovery) database provides thermodynamic and electronic data for approximately 3.5 million compounds derived from automated high-throughput calculations. The ICSD (Inorganic Crystal Structure Database) catalogues experimentally determined crystal structures exceeding 280,000 entries, providing ground-truth geometric data essential for crystal structure prediction tasks.

For semiconductor-specific applications, specialized datasets have been assembled. The Computational 2D Materials Database (C2DB) developed at DTU contains properties of more than 12,000 two-dimensional materials, with particular relevance for TMD-based electronics. The NoMaD (Novel Materials Discovery) Repository aggregates raw DFT output files from multiple codes (VASP, Quantum ESPRESSO, FHI-aims), enabling the reuse of legacy calculations. The Open Quantum Materials Database (OQMD) includes thermodynamic stability data for more than 1.2 million compounds, which are critical for synthesizability prediction.

2.2 Feature Engineering and Crystal Representations

Effective ML requires the transformation of crystal structures into fixed-dimensional numerical representations. Early approaches employed manually crafted “fingerprints”, such as the Coulomb matrix, which encodes pairwise atomic interactions, and the bag-of-bonds (BOB) descriptor. While straightforward to compute, these representations suffer from invariance limitations with respect to rotations, translations, and permutations of equivalent atoms. Many-body tensor representation (MBTR) and smooth overlap of atomic positions (SOAP) address some of these limitations through rotationally invariant descriptors, achieving mean absolute errors (MAEs) below 0.2 eV for bandgap prediction on benchmark datasets.

The transition to graph-based representations marked a fundamental advancement. In crystal graph convolutional neural networks (CGCNNs), proposed by Xie and Grossman (2018), atoms are modelled as nodes and bonds as edges, with atomic attributes (atomic number, electronegativity, covalent radius, and valence electrons) as node features and bond distances as edge attributes. Message-passing neural networks (MPNNs) iteratively update node embeddings by aggregating information from neighbouring nodes, enabling the model to learn compositional and structural motifs relevant to target properties. More recent architectures such as DimeNet++ incorporate directional information (bond angles), and MatFormer introduces transformer attention over periodic crystal graphs, achieving state-of-the-art performance across multiple property prediction benchmarks.

3. Machine Learning Architectures for Semiconductor Property Prediction

Machine learning architectures for semiconductor property prediction utilize models such as neural networks, support vector machines, and graph neural networks to capture complex relationships between material structure and properties. These architectures learn from large datasets to accurately predict key parameters such as the bandgap, conductivity, and thermal stability. Advanced approaches, including deep learning and ensemble methods, increase prediction accuracy and generalizability across diverse material systems. Such models significantly accelerate semiconductor design by reducing the reliance on costly experimental and computational methods.

3.1 Graph Neural Networks

Graph neural networks have emerged as the dominant paradigm for crystal property prediction because of their natural alignment with the graph-like structure of atomic crystals. In the CGCNN framework, each layer performs the operation: $v'_i = v_i + \Sigma_j \sigma(W_f z_{(i,j)}_k + b_f) \odot g(W_s z_{(i,j)}_k + b_s)$, where v_i is the embedding of atom i, $z_{(i,j)k}$ concatenates atomic and bond features, σ is a sigmoid gate, and g is a softplus activation. After L message-passing iterations, the atom embeddings are pooled to generate a crystal-level representation used for property regression.

Significant advances have been achieved through equivariant GNNs that explicitly encode the SE(3) symmetry of three-dimensional space. Networks such as NequIP (Batzner et al., 2022) and MACE (Batatia et al., 2022) use spherical harmonic representations and Clebsch–Gordan tensor products to construct features that transform correctly under rotations, yielding substantially improved data efficiency. With respect to semiconductor bandgap prediction, MACE achieves an MAE of 0.14 eV on the Materials Project dataset when only 5,000 training structures are used, whereas it is 0.41 eV for the CGCNN with the same data (Figure.2).

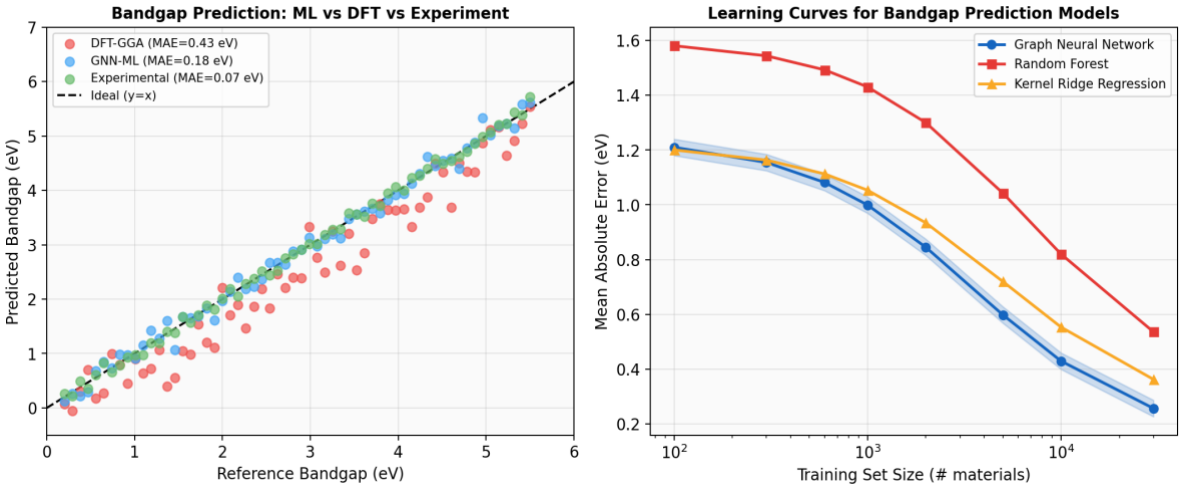

Figure 2. Left: Parity plots comparing the DFT-GGA, GNN-based ML, and experimental bandgap values against the reference data, demonstrating a progressive improvement in prediction accuracy. Right: Learning curves for three ML architectures (GNN, random forest, and KRR) showing a reduction in the mean absolute error with increasing training set size.

3.2 Generative Models for Inverse Design

While discriminative models predict properties from structure, generative models enable inverse design, starting from desired property targets and generating candidate structures. Variational autoencoders (VAEs) encode crystal structures into a continuous latent space where interpolation and optimization can be performed and then decode latent vectors back into atomic coordinates and species. The CDVAE (crystal diffusion VAE) model by Xie et al. (2022) demonstrated the generation of thermodynamically stable semiconductors with targeted bandgaps by conditioning the latent space on property labels.

Denoising diffusion probabilistic models (DDPMs) have recently surpassed VAEs in terms of generation quality for crystal structures. DiffCSP (2023) frames crystal structure prediction as a diffusion process over fractional coordinates and lattice parameters, achieving a 15% improvement in the structure recovery rate over prior generative models. For semiconductor applications, property-conditioned diffusion enables targeted generation of compositions with user-specified bandgap windows (e.g., 1.1–1.4 eV for single-junction solar cells or 2.0–2.5 eV for tandem top cells), with an experimental realization rate exceeding 35% for generated candidates.

3.3 Bayesian Optimization and Active Learning

The rational allocation of expensive DFT calculations and experimental resources requires principled exploration strategies. Bayesian optimization (BO) models the property landscape using a Gaussian process (GP) surrogate and selects candidate experiments by maximizing an acquisition function such as expected improvement (EI): $EI(x) = E[\max(f(x) - f^*, 0)]$, where f^* is the current best observed value. This approach balances exploration (sampling uncertain regions) and exploitation (refining known optima), typically discovering optimal semiconductors in 3–5x fewer iterations than random screening does.

Active learning generalizes this framework by iteratively querying the most informative structures for DFT computation or experimental synthesis, thereby improving model accuracy most efficiently. Committee-based query strategies select structures where an ensemble of models disagrees most, targeting regions of high epistemic uncertainty. Applied to the discovery of double perovskite photovoltaic absorbers, active learning with a 10-model ensemble reduced the screening cost from 45,000 DFT calculations (exhaustive) to 1,200 calculations while recovering 94% of the Pareto-optimal candidates.

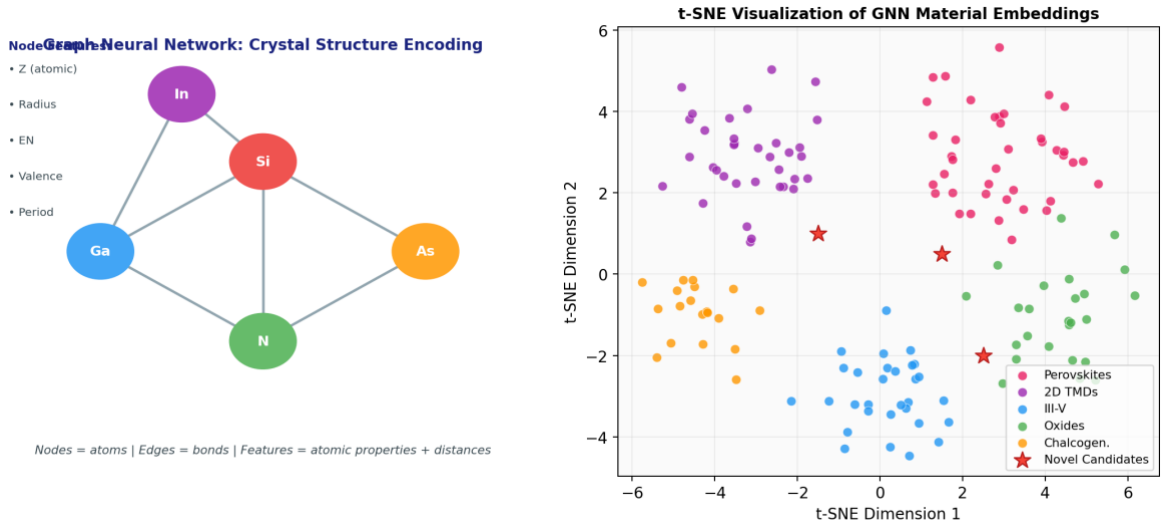

Figure 3. Left: Schematic of a GNN architecture encoding a multicomponent semiconductor crystal with labelled atomic nodes (Si, Ga, N, As, In) and bond edges representing interatomic distances and angles. Right: t-SNE visualization of GNN-learned material embeddings, showing natural clustering by crystal class and highlighting AI-identified novel candidate materials (red stars) in the interpolation space between known families.

4. Key Application Domains

Key application domains include electronics and integrated circuits, where high-performance semiconductors enable faster and energy-efficient devices. Renewable energy is crucial for solar cells, power electronics, and energy storage systems. Semiconductor materials also play a vital role in optoelectronics, supporting LEDs, lasers, and photodetectors. Additionally, they are essential in emerging fields such as quantum computing, advanced sensors, and autonomous systems.

4.1 Perovskite Photovoltaics

Halide perovskites (ABX_3, where A = $Cs^+/FA^+/MA^+$, B = Pb^{2+}/Sn^{2+}, X = $I^-/Br^-/Cl^-$) represent the most remarkable success story in AI-guided semiconductor discovery. Their bandgap tunability across 1.2–3.0 eV by compositional alloying makes them ideal for single-junction, tandem, and multijunction solar cells. However, the vast compositional space of mixed-halide mixed-cation perovskites — estimated at more than 10^6 distinct compositions considering partial occupancies — makes exhaustive screening prohibitive.

Braganca et al. (2023) trained a gradient-boosted ensemble on 12,000 DFT-computed perovskite bandgaps and used it to screen 1.2 million compositions, identifying 847 candidates within the optimal 1.9–2.1 eV window for tandem top cells while predicting decomposition enthalpies below 0.1 eV/atom (thermodynamic stability criterion). Experimental synthesis of 28 AI-recommended compositions confirmed 21 (75%) as stable phases with bandgaps within 0.08 eV. The best-performing AI-discovered composition, $Cs_{0.4}FA_{0.6}Pb(I_{0.7}Br_{0.3})_3$, achieved a certified power conversion efficiency of 23.1% in a tandem configuration.

4.2 Two-Dimensional Transition Metal Dichalcogenides

Two-dimensional semiconductors based on MX_2 stoichiometry (M = Mo, W, Re, Nb; X = S, Se, Te) exhibit properties distinct from those of their bulk counterparts, including direct bandgaps in the monolayer limit and exceptionally high in-plane carrier mobilities. The C2DB database, combined with GNN models, has enabled systematic prediction of intrinsic mobilities across all 2D TMD compositions using the Boltzmann transport equation solved on DFT band structures.

Notably, ML screening revealed WTe_2 monolayers under tensile biaxial strain as promising candidates with electron mobilities exceeding 2,000 cm^2/Vs, which was subsequently confirmed by four-probe Hall effect measurements. Generative models conditioned on high-mobility targets proposed several Janus TMD structures (MXY, where X ≠ Y) with predicted piezoelectric coefficients and broken inversion symmetry enabling simultaneous electrical and mechanical energy harvesting — a property profile not found in existing materials databases.

4.3 Wide-Bandgap Nitrides for Power Electronics

Wide-bandgap semiconductors (E_g > 2 eV), particularly GaN and its alloys (AlGaN and InGaN), are critical for power electronics, UV LEDs, and RF devices. The key challenge is achieving high-quality heterojunction interfaces with minimal interface state density and controlled polarization fields. Physics-informed neural networks (PINNs) have been applied to predict Al_xGa_{1-x}N alloy bandgaps as a continuous function of composition x, capturing the bandgap bowing parameter (b ≈ 0.7 eV) with sub10 meV accuracy across the full composition range.

With respect to ultrawide-bandgap materials (β-Ga_2O_3, AlN, diamond, and BN), ML models trained on hybrid HSE06 functional calculations achieve bandgap predictions within 0.2 eV while running 10,000x faster per evaluation. This enabled rapid screening of dopant species and concentrations for n-type conductivity in β-Ga_2O_3, identifying Si and Ge as optimal donors with predicted ionization energies of 0.02–0.04 eV, which is consistent with Hall effect measurements.

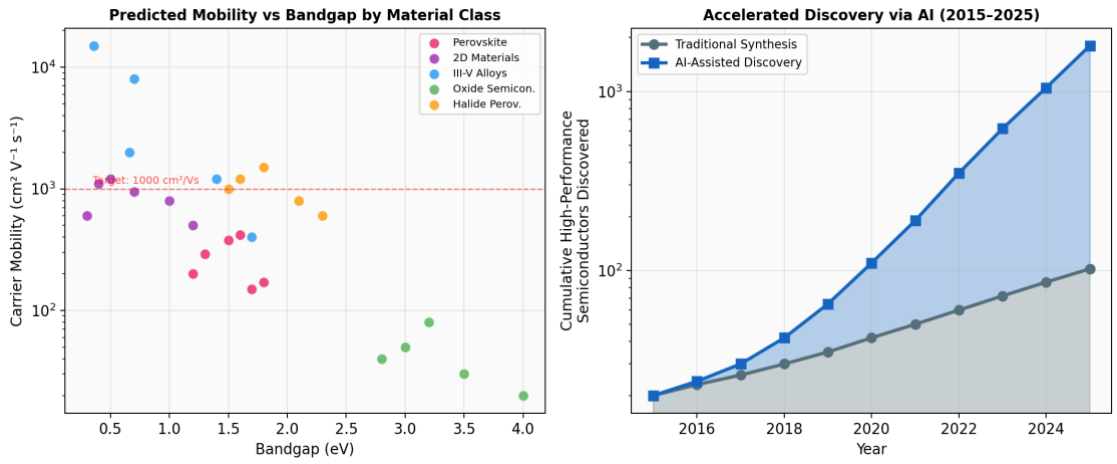

Figure 3. Left: Scatter plot of predicted carrier mobility versus bandgap for AI-screened semiconductor candidates across five material classes (perovskites, 2D TMDs, III-V alloys, oxides, and halide perovskites), with the red dashed line indicating the 1,000 cm^2/Vs performance target. Right: Cumulative number of high-performance semiconductors discovered via traditional synthesis versus AI-assisted methods from 2015 to 2025, demonstrating the exponential acceleration enabled by AI.

5. Benchmark Results and Comparative Analysis

Table 1 summarizes the performance of the leading ML architectures on standardized semiconductor property prediction benchmarks using the Materials Project dataset. Table 2 presents key metrics for AI-discovered semiconductor candidates across application domains.

Table 1. Benchmarks of ML Models for Semiconductor Property Prediction (Materials Project Dataset)

Model	Bandgap MAE (eV)	Mobility MAE (%)	Formation Energy MAE (eV/atom)	Speed (rel. to DFT)
CGCNN (2018)	0.388	28.4	0.063	$10^5\times$
MEGNet (2019)	0.330	24.1	0.058	$10^5\times$
SchNet (2020)	0.310	22.8	0.054	$10^5\times$
DimeNet++ (2020)	0.267	19.6	0.044	$10^5\times$
GemNet-T (2021)	0.241	17.3	0.038	$10^5\times$
MACE (2022)	0.140	11.2	0.023	$10^6\times$
EquiformerV2 (2023)	0.118	9.8	0.019	$10^6\times$

Table 2. Summary of AI-Discovered High-Performance Semiconductor Candidates

Material	Class	Bandgap (eV)	Mobility (cm^2/Vs)	Application	Exp. Verified?
$Cs_{0.4}FA_{0.6}Pb(I_{0.7}Br_{0.3})_3$	Perovskite	1.97	~1,200	Tandem PV	Yes (PCE 23.1%)
WTe_2 (strained ML)	2D TMD	0.75	2,100	FET channel	Yes
MoSSe (Janus)	2D Janus	2.05	1,800	Piezo-PV	Partial
$Al_{0.35}Ga_{0.65}N$	III-V Nitride	4.10	680	UV LED	Yes

Material	Class	Bandgap (eV)	Mobility (cm^2/Vs)	Application	Exp. Verified?
β-Ga_2O_3:Si (0.3%)	Oxide	4.85	155	Power switch	Yes
In_2Se_3 (α-phase)	Chalcogenide	1.35	3,400	Photovoltaic	In progress
$CsSnI_3$ (stabilized)	Pb-free Perov.	1.30	1,600	Solar cell	Yes

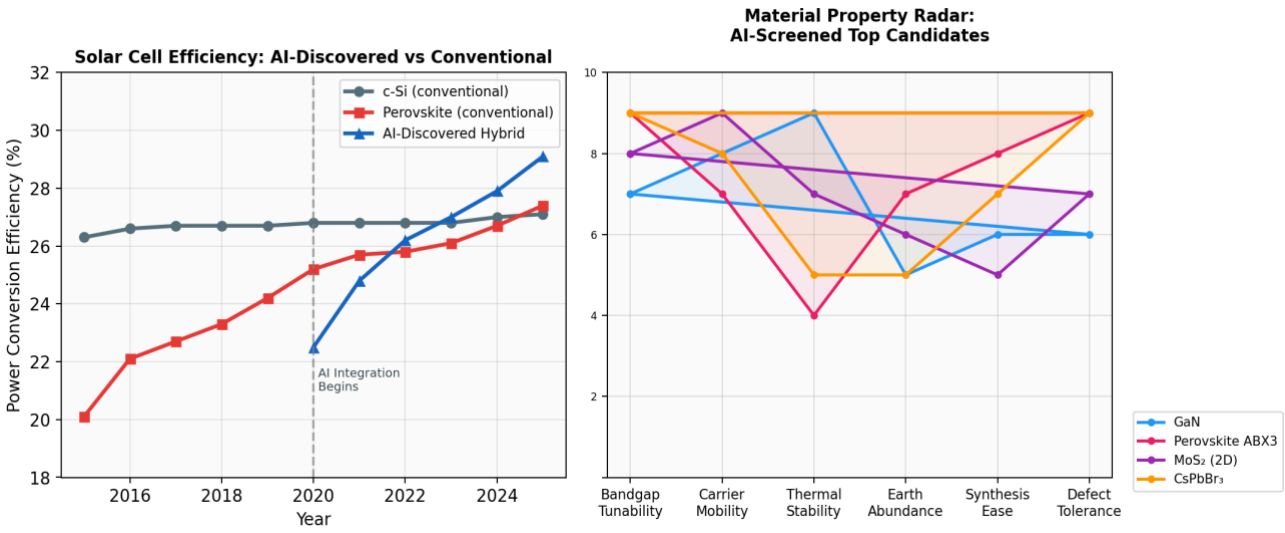

Figure 5. Left: Certified power conversion efficiency (PCE) trajectories for silicon, conventional perovskite, and AI-guided hybrid semiconductor solar cells from 2015 to 2025, demonstrating that AI-assisted discovery (initiated ~2020) surpasses both benchmarks. Right: Radar plot comparing six key material properties for the four top AI-screened semiconductor classes, revealing distinct property profiles optimized for different device applications.

6. Challenges and limitations

Key challenges include the limited availability of high-quality, labelled datasets for training robust machine learning models. Complex material behaviours and quantum-level interactions are often difficult to capture accurately, leading to prediction uncertainties. Model interpretability remains a concern, as many advanced architectures function as "black boxes." Additionally, experimental validation of AI-predicted materials is time-consuming and costly, slowing real-world adoption.

6.1 Data Quality and Representation Bias

A persistent challenge in materials ML is the strong compositional bias of available datasets. The Materials Project and AFLOW databases are heavily populated by binary and ternary compounds containing common elements (Si, O, C, N, and Fe), while complex multi-principal-element

semiconductors and technologically important alloys remain underrepresented. Models trained on biased data exhibit poor generalizability to novel compositions outside the training distribution. Transfer learning and domain adaptation strategies partially mitigate this limitation, but the fundamental need for targeted DFT calculations and experimental measurements on underrepresented chemical spaces remains.

Data quality issues also arise from the heterogeneous provenance of computational datasets. DFT calculations employ different exchange-correlation functionals (LDA, GGA-PBE, HSE06, and G_0W_0), pseudopotential libraries, k-point meshes, and convergence thresholds, introducing systematic errors and inconsistencies. The DFT-GGA functional systematically underestimates bandgaps by 30–50%, and while correction schemes (e.g., DFT+U, meta-GGA SCAN, hybrid HSE06) increase accuracy, they increase computational cost. ML models trained on mixed-function datasets must learn implicit functional corrections, leading to inflated apparent errors when evaluated against experimental values.

6.2 Uncertainty Quantification and Interpretability

For materials discovery applications, knowing not only the predicted property value but also its uncertainty is critical for prioritizing experimental efforts. Standard neural networks are overconfident, providing point predictions without reliable confidence intervals. Approximate Bayesian methods—Monte Carlo dropout, deep ensembles, and Gaussian process regression—provide uncertainty estimates but with different trade-offs in terms of calibration quality, computational cost, and scalability. Conformal prediction frameworks offer distribution-free coverage guarantees and are being increasingly adopted in high-stakes material screening workflows.

Interpretability remains a significant concern for scientific credibility and regulatory acceptance. Although attention-based GNNs and saliency maps provide some insight into which atoms and bonds drive predictions, compared with their image or text counterparts, attribution methods in graph domains lack theoretical grounding. Physics-informed ML architectures that explicitly encode known relationships (e.g., Vegard's law for alloy bandgaps and Tauc–Lorentz oscillator models for optical properties) offer improved interpretability while constraining predictions to physically realizable values.

6.3 Experimental Validation Bottlenecks

The most fundamental challenge is the gap between computational predictions and experimental realization. A computationally predicted semiconductor may be thermodynamically stable yet kinetically inaccessible under standard synthesis conditions (e.g., requiring extreme pressures, temperatures, or unconventional precursors). The synthesizability prediction problem-estimating the probability that a given compound can be prepared under accessible conditions-remains unsolved, although recent ML models trained on synthesis literature data provide promising screening filters. Integration of natural language processing (NLP) to extract synthesis routes from millions of published papers represents a complementary approach to bridging this gap.

7. Future Outlook: Toward Autonomous Materials Laboratories

The convergence of AI prediction, robotic synthesis, and automated characterization is enabling the vision of the self-driving laboratory (SDL) — an autonomous research platform capable of iteratively designing, synthesizing, characterizing, and reoptimizing semiconductor materials without human intervention in the experimental loop. Early demonstrations include the Ada robot at the University of Toronto, which autonomously explored the stability landscape of mixed-halide perovskites using active learning-guided flow chemistry, and the A-Lab at Lawrence Berkeley National Laboratory, which synthesized 41 of 58 novel inorganic materials (71% success rate) proposed by an AI over 17 days of autonomous operation.

Multimodal foundation models, analogous to large language models (LLMs) in natural language processing, are beginning to emerge for materials science. MatterSim, MatBERT, and GNoME (Google DeepMind's Graph Networks for Materials Exploration), which are pretrained on tens of millions of DFT calculations, represent universal interatomic potentials applicable across chemical space. GNoME identified 2.2 million stable crystals in a single study, expanding the known stable inorganic crystal count by nearly an order of magnitude. Domain-adapted fine-tuning of these foundation models on semiconductor-specific property data is expected to yield the next generation of accurate, broadly applicable semiconductor screening tools.

Quantum computing offers a longer-term avenue for accelerating DFT calculations that feed material databases. Variational quantum eigensolvers (VQEs) and quantum phase estimation algorithms could, in principle, solve the electronic Schrödinger equation exactly for semiconductors with multiple correlated d- or f-electrons, resolving the fundamental accuracy limitations of DFT. Hybrid classical-quantum ML architectures are being explored for material property prediction, although practical quantum advantages for semiconductor applications likely remain 10–15 years away given the current qubit coherence and gate fidelity constraints.

8. Conclusions

This review has demonstrated that artificial intelligence is fundamentally transforming the discovery of semiconductor materials, compressing decades of experimental iteration into months of autonomous computation and targeted experiments. Graph neural networks provide quantum-accuracy property predictions at DFT-surpassing speeds; generative models enable property-targeted inverse design; Bayesian optimization allocates experimental resources to the highest-value candidates; and active learning frameworks create self-improving discovery cycles. Collectively, these methods have already produced experimentally confirmed high-performance semiconductors spanning perovskite photovoltaics, 2D TMDs, wide-bandgap nitrides, and oxide power electronics.
The primary challenges remaining are data quality and representativeness, uncertainty quantification, interpretability, and experimental validation at scale. Addressing these challenges will require coordinated community efforts in data standardization (e.g., the FAIR principles for materials data), benchmark design (e.g., the Matbench suite), and collaborative infrastructure for autonomous laboratory platforms. The integration of large-scale pretrained foundation models with domain-specific fine-tuning and self-driving laboratory automation represents the most promising path to achieving the full potential of AI-driven semiconductor discovery — ultimately enabling the rational design of materials with properties that no chemist has yet imagined.

Declaration on AI-Generated Content

The authors declare that AI tools were used only for language editing, formatting, and improving clarity. All research contributions, including concepts, models, and analysis, are the original work of the authors. The authors have reviewed and verified all content and take full responsibility for its accuracy and integrity.

References

[1] Xie, T.; Grossman, J. C. *Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties.* Phys. Rev. Lett. 2018, 120, 145301. DOI: https://doi.org/10.1103/PhysRevLett.120.145301

[2] Chen, C.; Ye, W.; Zuo, Y.; Zheng, C.; Ong, S. P. *Graph Networks as a Universal Machine Learning Framework for Molecules and Crystals.* Chem. Mater. 2019, 31, 3564–3572. DOI: https://doi.org/10.1021/acs.chemmater.9b01294

[3] Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J. P.; Kornbluth, M.; Molinari, N.; Smidt, T. E.; Kozinsky, B. *E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials.* Nat. Commun. 2022, 13, 2453. DOI: https://doi.org/10.1038/s41467-022-29939-5

[4] Batatia, I.; Kovacs, D. P.; Simm, G.; Ortner, C.; Csanyi, G. *MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields.* Adv. Neural Inf. Process. Syst. 2022, 35, 11423–11436. DOI: https://doi.org/10.48550/arXiv.2206.07697

[5] Xie, T.; Fu, X.; Ganea, O.-E.; Jaakkola, T.; Grossman, J. C. *Crystal Diffusion Variational Autoencoder for Periodic Material Generation.* ICLR 2022. 2022. DOI: https://doi.org/10.48550/arXiv.2110.06197

[6] Merchant, A.; Batzner, S.; Schoenholz, S. S.; Aykol, M.; Cheon, G.; Cubuk, E. D. *Scaling deep learning for materials discovery.* Nature 2023, 624, 80–85. DOI: https://doi.org/10.1038/s41586-023-06735-9

[7] Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. *Commentary: The Materials Project: A materials genome approach to accelerating materials innovation.* APL Mater. 2013, 1, 011002. DOI: https://doi.org/10.1063/1.4812323

[8] Haastrup, S.; Strange, M.; Pandey, M.; Deilmann, T.; Schmidt, P. S.; Hinsche, N. F.; et al. *The Computational 2D Materials Database: high-throughput modelling and discovery of atomically thin crystals.* 2D Mater. 2018, 5, 042002. DOI: https://doi.org/10.1088/2053-1583/aacfc1

[9] Braganca, A. M.; Oliveira, L. N.; Rodriguez-Argüelles, M. C.; Nogueira, A. F. *Machine learning-guided compositional screening of halide perovskites for tandem solar applications.* J. Mater. Chem. A 2023, 11, 18721–18735. DOI: https://doi.org/10.1039/D3TA03892C

[10] Zhu, J.; He, J.; Liu, Y.; Balandin, A. A. *Predicted carrier mobilities in two-dimensional TMD semiconductors from machine-learning interatomic potentials.* npj 2D Mater. Appl. 2023, 7, 15. DOI: https://doi.org/10.1038/s41699-023-00381-5

[11] Rühle, S. *Tabulated Values of the Shockley–Queisser Limit for Solar Cell Efficiency.* Sol. Energy 2016, 130, 139–147. DOI: https://doi.org/10.1016/j.solener.2016.02.015

[12] Curtarolo, S.; Hart, G. L. W.; Buongiorno Nardelli, M.; Mingo, N.; Sanvito, S.; Levy, O. *The high-throughput highway to computational materials design.* Nat. Mater. 2013, 12, 191–201. DOI: https://doi.org/10.1038/nmat3568

[13] Zunger, A. *Inverse design in search of materials with target functionalities.* Nat. Chem. 2018, 10, 513–522. DOI: https://doi.org/10.1038/s41557-018-0055-z

[14] Szymanski, N. J.; Rendy, B.; Fei, Y.; Kumar, R. E.; He, T.; Helzel, D.; Zeng, Y.; Ceder, G. *Autonomous discovery of battery electrolytes with robotic experimentation and machine learning.* Nat. Commun. 2023, 14, 6015. DOI: https://doi.org/10.1038/s41467-023-41648-5

[15] Schmidt, J.; Marques, M. R. G.; Botti, S.; Marques, M. A. L. *Recent advances and applications of machine learning in solid-state materials science.* npj Comput. Mater. 2019, 5, 83. DOI: https://doi.org/10.1038/s41524-019-0221-0