

Digital Discovery

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: C. Ma, Y. WANG, D. Zhang, W. Du, Q. Gao, R. Su, K. Xu, H. Gu, L. Li, P. Ma and H. Li, *Digital Discovery*, 2026, DOI: 10.1039/D6DD00218H.



This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.

PERSPECTIVE

Closed-Loop Discovery of Energy Materials Empowered by Artificial Intelligence Models

Chenyao Ma,^a Yuhang Wang,^c Di Zhang,^{c,d} Wei Du,^{a,b} Qiang Gao,^a Rui Su,^a Kan Xu,^a Huan Gu,^a Limin Li,^{*a,b} Piao Ma,^{*a,b} and Hao Li^{*c}Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Energy materials underpin global energy transformation and low-carbon development, and their innovation and performance optimization are essential to removing bottlenecks in energy conversion and storage. Conventional research and development modes, including empirical, theoretical, and computational methods, are restricted by low efficiency or inherent conflicts between accuracy and system scale, and can hardly satisfy the requirements of modern high-precision, high-throughput material investigations. The 4^{th+} paradigm, an extension of the data-driven paradigm (4th paradigm) empowered by advanced artificial intelligence (AI) and data science, takes material databases, universal machine learning interatomic potentials (MLIPs), large language models (LLMs), intelligent agents, and full-process closed-loop systems as the core methodological framework of this perspective. It delivers atomic-scale modeling accuracy and efficient knowledge extraction capabilities to overcome these limitations, distinguished by its more sophisticated methodologies for generating, processing, and extracting knowledge from data. This *Perspective* summarizes the advances of this paradigm in energy materials research and clarifies the mechanisms by which MLIPs and LLMs break through traditional constraints. On this basis, we further construct a future-oriented research framework consisting of four modules and outline prospective development trends to advance intelligent research platforms and accelerate the discovery and industrial translation of high-performance energy materials.

Introduction

Energy materials form the bedrock of the global energy system and green low-carbon development.¹⁻³ Innovative exploration and performance optimization are crucial to breaking through bottlenecks in energy conversion, storage, and utilization. The progression of advanced energy materials research embodies a fundamental evolution across four distinct research paradigms: empirical science, theoretical science, computational science, and data-driven science (**Figure 1**).^{4,5} This evolution began with empirical trial-and-error experimentation, which yielded fragmented datasets yet laid foundational concepts for materials discovery. Theoretical approaches thereafter delivered atomic-level understanding based on intrinsic physical descriptors, and the advent of quantum chemistry ushered in computational science, correlating electronic configurations with macroscopic properties and accumulating fundamental research datasets. Limited by scale and efficiency, computational science paved the way for data-driven science, which rose to the forefront with exponential growth in data and computing power, driving big data analytics for accelerated materials discovery. Advanced AI tools extend conventional

data-driven paradigm into the upgraded 4^{th+} paradigm.⁵ Unlike conventional data-driven research, its core advantage lies in enlarged data scale and refined pipelines for data generation, processing and knowledge mining. Further empowered by large-scale AI models including universal machine learning interatomic potentials (MLIPs)⁶ and large language models (LLMs),⁴ the 4^{th+} paradigm facilitates the rational design and discovery of high-performance energy materials (**Figure 1**).^{7,8}

Each paradigm overcomes the limitations of its predecessor while giving rise to new research and development challenges. The evolution of energy materials R&D has spanned empirical, theoretical, and computational paradigms. As the core research direction of this evolutionary trend, AI-driven discovery focuses on designing, optimizing and predicting properties of energy materials via ML, data mining and high-throughput computation to assist conventional experiments. Many traditional trial-and-error experiments are inefficient, costly, and stochastic, generating fragmented data that impedes rational materials design. Theoretical advances, including quantum chemistry and density functional theory (DFT), offer atomic-level insights into electronic and crystal structures.⁵ Nevertheless, DFT exhibits inherent limitations: an accuracy-

^a Suzhou MatSource Technology Co., Ltd., Suzhou 215000, Jiangsu, China.^b Gusu Laboratory of Materials, Suzhou 215000, Jiangsu, China^c Advanced Institute for Materials Research (WPI-AIMR), Tohoku University, Sendai 980-8577, Japan^d Frontier Research Institute for Interdisciplinary Sciences (FRIS), Tohoku University, Sendai, 980-8577, Japan

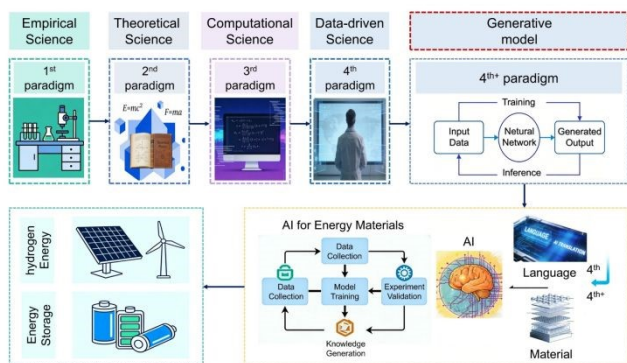


Figure 1 The evolution of scientific paradigms and application in AI-driven energy materials. This figure depicts the sequential evolution of scientific paradigms from empirical, theoretical, computational to data-driven science, culminating in the 4^{th+} generative paradigm empowered by AI and data science, and showcases their applications in energy materials research.

scale trade-off, poor adaptability to complex interfaces and disordered systems, and low efficiency in long-time dynamic simulations, which restrict the high-precision, high-throughput development of energy materials. The universal MLIPs bridge the gap between quantum-mechanical precision and large-scale exploration of energy materials, thereby overcoming the computational limitations of conventional DFT-based methods. The data-driven (4th) paradigm has ushered in a new era for energy materials research.⁹ In this Perspective, we use the term "4^{th+} paradigm" to emphasize that current AI-driven materials discovery remains fundamentally data-dependent, while being substantially enhanced by MLIPs, LLMs, and autonomous agents. In contrast, conventional high-throughput screening (HTS) and self-driving labs (SDLs) have greatly accelerated materials exploration, yet most existing implementations remain task-specific and fail to achieve full integration with standardized databases, physics-aware models, literature-mined knowledge, and long-term feedback infrastructures. Differently, the 4^{th+} paradigm organically unifies standardized databases, MLIP calculation and LLM knowledge mining to build a complete closed-loop workflow spanning data input, computational prediction and autonomous experimental validation. It integrates universal MLIPs with advanced natural language processing tools to address longstanding experimental and theoretical bottlenecks in the field.^{10, 11} In particular, MLIPs can achieve near DFT-level computational accuracy and boost the efficiency of large-scale simulations by orders of magnitude compared with traditional methods, enabling atomic-level investigations of complex systems. Furthermore, advanced text mining techniques can extract and synthesize knowledge from massive volumes of literature and datasets. The deep integration of computational modeling and knowledge synthesis underpins a data-centric research framework, shortens the R&D cycle, and drives the high-precision prediction and rational design of energy materials. Against this backdrop, energy materials research supported by advanced computational intelligence has entered a new stage of opportunity.

Drawing on high-quality databases, ML regression models (particularly MLIPs), LLMs, and full-cycle closed-loop research systems, this *Perspective* focuses on the data-decisive 4^{th+} paradigm to systematically review the applications of AI models in the rational design and discovery of energy materials. It analyzes how modern computational methods break through traditional research bottlenecks, aiming to provide support for building standardized, intelligent research platforms and advancing the efficient discovery and industrial translation of high-performance energy materials.^{12, 13}

Materials databases and digital platforms

Databases in energy materials science are undergoing a fundamental role transformation. They are no longer merely places to store data, but rather the underlying operating system of modern digital materials science.^{14, 15} This new role is rooted in a clear evolutionary trajectory of databases themselves.^{16, 17}

Early materials databases were centered on experimental data (**Figure 2a**). Extensive explorations have been carried out by researchers across various fields, as exemplified by several specialized databases: PolyInfo,¹⁸ which focuses on polymer structure–property relationships; Starrdata2,¹⁹ dedicated to functional inorganic materials (e.g., thermoelectric and magnetic materials); the Crystallography Open Database (COD), which compiles small-molecule crystal structures; and the

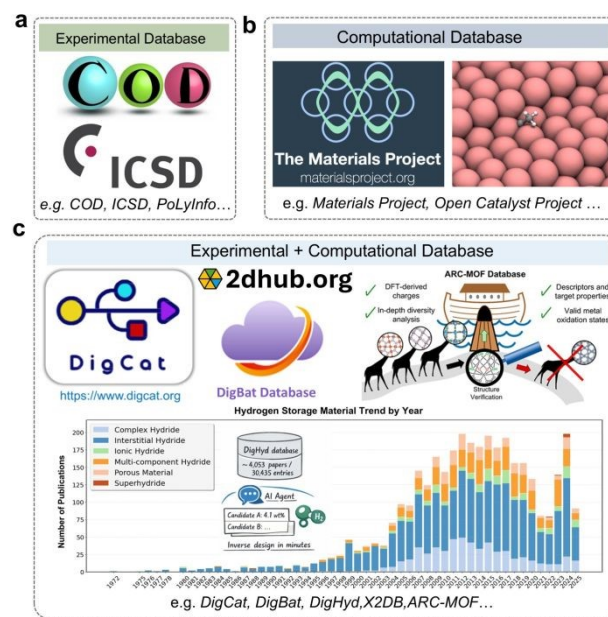


Figure 2 Evolution of materials databases. (a) Experiment-driven databases. Representative repositories for small-molecule and inorganic crystal structures (e.g., COD, ICSD, PoLyInfo). (b) Computation-driven databases. High-throughput first-principles databases (e.g., Materials Project, Catalysis-Hub). (c) Integrated experiment-theory platforms combining computational and experimental data, enabling AI-driven applications in heterogeneous catalysis, solid-state electrolytes, and hydrogen storage (e.g., DigCat, DigBat, DigHyd, X2DB, ARC-MOF).



Inorganic Crystal Structure Database (ICSD),²⁰ which stores experimentally determined inorganic crystal structures. For example, the COD is an open-access repository containing approximately 150,000 small-molecule crystal structures, supporting global data sharing, curation, and reuse of crystallographic knowledge.²¹ In parallel, the ICSD is a curated database of experimentally determined inorganic crystal structures, providing reliable structural data to support materials characterization, comparison, and discovery.²⁰ However, experiment-based databases face a key challenge: material performance strongly depends on testing conditions, process parameters, and sample preparation details.²² The same material may yield significantly different data under different laboratories and testing conditions. Without standardized recording and extraction of their metadata, the reference value of a single performance value is limited.

With the rapid development of computational materials science, especially the development of DFT and high-throughput computing, computational materials databases have emerged (Figure 2b). Through large-scale first-principles calculations, such databases provide systematic data for the stability, electronic structure, and thermodynamic properties of materials. A typical example is the Materials Project database, which opens crystal structure, energetics, and other data of inorganic materials and supports application programming interface (API) calls.²³ Other databases with similar functions include the Catalysis-Hub²⁴ and Open Catalyst.²⁵ The value of these databases lies in their ability to quickly screen candidate materials before experiments, transforming materials search from empirical exploration into a quantifiable computational process. However, computational databases also have inherent limitations. The vast majority of computational data is based on idealized assumptions, such as perfect crystals and ideal surface structures. While these assumptions ensure the feasibility and repeatability of the calculations, they also make it difficult for the results to directly reflect the complexities of actual operating conditions. Many real-world factors (e.g., surface reconstruction, defect sites, solvent effects, electrode potentials, and interfacial interactions) are often simplified or neglected.

Based on the respective advantages and disadvantages of the two types of databases mentioned above, a new database format has emerged: an integrated platform that combines computational results and experimental metrics. Figure 2c summarizes representative platforms sorted by five material categories: catalysis, solid electrolytes, hydrogen storage, porous MOFs and two-dimensional materials. For catalytic research, the Digital Catalysis Platform (*DigCat*: www.digcat.org),²⁶ integrates abundant computational surface libraries and standardized experimental datasets. *Via* embedded functional interfaces, the platform correlates theoretical adsorption energies with practical device performance and enables instant access to both computational predictions and experimental validations. For solid electrolyte development, Digital Battery Platform (*DigBat*) focuses on solid-state electrolytes,^{27, 28} emphasizing the standardization of core parameters such as ionic conductivity and migration activation energy; while the Digital Hydrogen Platform (*DigHyd*) extracts structured data from literature images through a multi-agent workflow to build a high-quality database;^{29, 30} in the porous material category, the ARC-MOF database contains around 280,000 experimentally reported and computationally

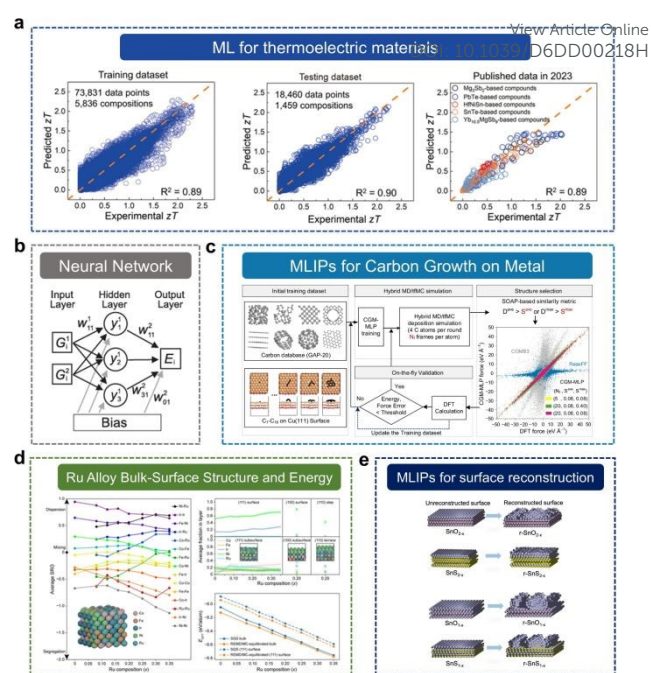


Figure 3 ML for materials research. (a) Evaluation of ML models and their predictions for new materials. (b) Example of a standard neural network employed for fitting potential-energy surfaces. (c) MLIPs training for surface reconstruction analyses. (d) MLIPs for the bulk-surface structure and energy of Ru-based high-entropy alloys (e) MLIPs for surface reconstruction analyses. Adapted with permission from: (a) ref. 19 © 2024 Springer Nature (b) ref. 35 © 2007 American Physical Society (c) ref. 40 © 2024 Springer Nature (d) ref. 41 © 2025 American Chemical Society (e) ref. 42 © 2025 Wiley-VCH

generated MOF structures with DFT-derived atomic and adsorption parameters,³¹ These data have been used to support comparative evaluation of CO_2 capture performance across different MOF materials.³² For low-dimensional materials, X2DB collects 370 experimentally synthesized two-dimensional materials and links experimental findings with computational databases for cross-scale property characterization.³³ These integrated platforms share a common feature. They connect computational descriptors, experimental records, contextual metadata, and tool interfaces. This transforms material data from simple aggregation into a dynamic infrastructure that supports AI access, experimental validation, and process feedback.

However, current materials databases still face an important challenge. Publication bias causes successful results to be widely reported, while “failed experiments” and negative results are rarely included. This can make AI models overly optimistic about the synthesizability of materials. In the future, negative results should be systematically collected and standardized, with clear categories for synthesis failures and performance failures.



Machine Learning (ML) Regression Models

Conventional machine learning regression lays the theoretical basis, while MLIPs constitute the core focus of this section. In the research and development of energy materials, DFT stands as the gold standard for atomic-scale property calculations, yet its exorbitant computational cost renders it inapplicable to long-time dynamic simulations or large-scale high-throughput screening. In contrast, traditional property prediction models, despite their high computational efficiency, rely on manually designed low-dimensional descriptors, suffer from a lack of universality, and fail to accurately capture the complex interatomic interactions. Data-driven ML methods provide a viable solution to the above bottlenecks. Serving as an effective bridge between DFT calculations and rapid performance evaluation, conventional machine learning models take structural features as basic inputs and mine hidden structure–performance relationships from limited datasets. These approaches maintain high prediction accuracy while greatly reducing computational consumption. Taking thermoelectric materials as a case study, a gradient boosting decision tree (GBDT) ML model was constructed to achieve high-precision prediction of material thermoelectric properties. Subsequently, DFT calculations verified the reliability of two potential high-performance thermoelectric materials predicted by the model, which intuitively reflects the practical value of the synergy between ML and DFT and also provides a feasible example for addressing big data challenges in the application of AI for materials science (Figure 3a).¹⁹

MLIPs represent a pivotal application of ML in materials research.³⁴ Boasting core advantages of high efficiency and precision, they play a crucial role in atomic-scale materials simulations and catalysis studies. The generalized neural-network representation of high-dimensional potential-energy surfaces proposed in the seminal paper published in 2007 is generally regarded as the prototype of current MLIPs, providing core ideas and a theoretical framework for the subsequent development and innovation of MLIP methods (Figure 3b). These later developments built on and broadened the foundational concepts introduced by Behler and other early MLIP pioneers.^{35–37} For example, researchers proposed the Gaussian Approximation Potential (GAP), Many-body Atomic Cluster Expansion (MACE) model, and the MACE-MP foundation model, which overcame critical bottlenecks in earlier MLIPs, enabled robust generalization across diverse chemical systems and length scales, and achieved landmark results in materials simulation, chemical calculation, and other fields. Based on this methodological framework, researchers further optimized and expanded the applicability of the GAP and MACE models, addressing key technical challenges in MLIP development for large-scale system simulations and complex reaction predictions, and promoting the transformation of MLIPs from theoretical research to practical application with large-scale implementation.^{38, 39} These works have greatly promoted the in-depth integration of MLIP technology with fields such as energy materials and catalytic chemistry, providing new and efficient tools for related research.

In practical applications, MLIPs integrated with molecular dynamics can accurately replicate graphene growth on Cu substrates and carbon deposition on various metal surfaces. As a representative class of advanced energy materials, such carbon-based functional materials play a vital role in

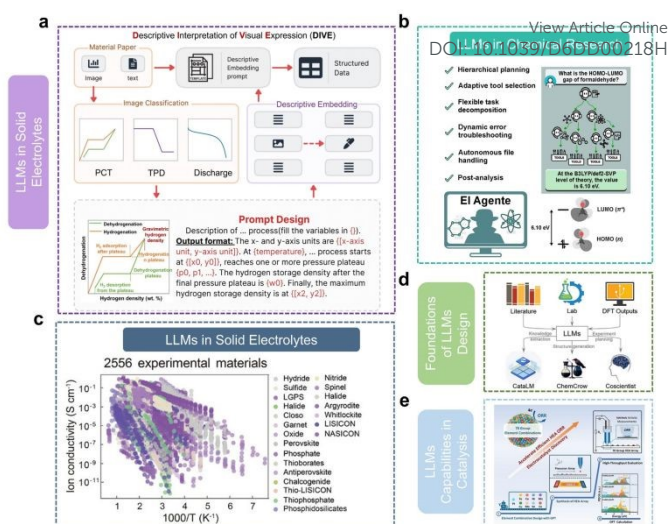


Figure 4 LLMs and AI agents for materials research. (a) Schematic diagram and evaluation methods of the DIVE workflow. (b) El Agente autonomous agent for quantum chemistry. (c) Big data-driven AI analysis of hydride SSEs. (d) Establishment of a large prediction model for catalyst design. (e) Accelerated discovery of ORR electrocatalysts in Pt-based high-entropy alloys. Adapted with permission from: (a) ref. 43 © 2026 Royal Society of Chemistry. (b) ref. 44 © 2025 Cell Press. (c) ref. 46 © 2025 Wiley-VCH (d) ref. 5 © 2026 Wiley-VCH (e) ref. 47 © 2025 American Chemical Society

electrochemical energy storage and conversion, and this work lays a solid foundation for their targeted development (Figure 3c).⁴⁰ For the acidic oxygen evolution reaction (OER), MLIPs combined with replica-exchange molecular dynamics resolve the bottlenecks plaguing Ru-based multicomponent energy materials, including insufficient stability, ambiguous atomic-scale mixing and phase formation pathways, and low-efficiency high-throughput screening routines. This integrated strategy characterizes the atomic-scale mixing behaviour of Ru_x(Ir,Fe,Co,Ni)_{1-x} and RuIr-based alloys, verifies homogeneous bulk face-centered cubic (fcc) phase mixing and reveals the formation of minor hexagonal close-packed (hcp) phase, and establishes a high-throughput simulation protocol to rapidly screen promising RuIr-containing quinary materials, delivering rigorous theoretical guidance for high-performance Ru-based OER energy components (Figure 3d).⁴¹ In the realm of electrocatalysis and sustainable carbon utilization, MLIPs greatly accelerate the simulation efficiency of CO₂ electroreduction on Sn-based substrates, accurately reproduce surface structural reconstruction, and when coupled with pH-coupled microkinetic models, uncover pH-dependent structure-activity relationships.⁴² This establishes a feasible route for the rational design of high-performance electroactive materials for CO₂ reduction (Figure 3e). Going forward, with the advancement of multi-fidelity training and active learning, and in-depth integration with LLMs and automated experimental platforms, MLIPs will further drive catalysis research toward the "data-driven, simulation-predicted, experimental-validated" closed-loop paradigm, fuelling the rational design and industrial application of high-efficiency catalysts and empowering innovations and industrialization in energy materials.



Nevertheless, MLIP techniques still suffer from inherent drawbacks including prediction uncertainty for configurations far outside training datasets, low reproducibility induced by differing training hyperparameters, and substantial experimental validation costs required to verify computational predictions.

Large language models (LLMs) and intelligent agents

LLMs (and AI agents developed based upon LLMs) serve as the core knowledge support and autonomous enabling tools for the intelligent R&D of energy materials, facilitating the implementation of the 4th research paradigm and driving the transformation of R&D models from experience-driven to knowledge-driven. The Descriptive Interpretation of Visual Expression (DIVE) multi-agent workflow is a typical practical application of such models in materials science.⁴³ Leveraging the visual analysis and textual reasoning capabilities of multimodal LLMs, and integrating prompt engineering, embedding model evaluation and ML validation technologies, this workflow assists in extracting and synthesizing literature resources in the field of solid-state hydrogen storage materials, and efficiently extracts multi-dimensional information from graphical experimental data in scientific literature. It effectively addresses the shortcomings of traditional databases, such as scattered information and insufficient standardization, and realizes cross-scenario and cross-scale knowledge integration and data activation in this field (Figure 4a).

El Agente⁴⁴ is a multi-agent system built on LLMs, facilitating the popularization of computational quantum chemistry. This system can efficiently accomplish core tasks such as molecular structure optimization and material property prediction, and shows great potential in transforming traditional research models in fields such as drug discovery and materials science; its future integration with SDLs will further accelerate the process of scientific discovery, making it a breakthrough tool supporting large-scale and inclusive quantum chemistry research and education, and opening up an avenue for the application of LLMs in chemical research (Figure 4b).⁴⁴

In terms of subsequent research on energy materials, LLMs facilitate research hypothesis generation, candidate system screening, and preparation optimization, guiding the rational design of high-performance energy materials.⁴⁵ In the development of hydride-based solid electrolytes, LLMs collaborate with comprehensive solid electrolyte databases and *ab initio* molecular dynamics simulations to break through the efficiency bottlenecks of conventional trial and error approaches. They rapidly identify promising candidates with low activation energies and provide key technical support for revealing complex ion migration mechanisms and developing high-performance solid electrolytes (Figure 4c).⁴⁶ Furthermore, the application of LLMs in energy catalysis continues to expand. For high efficiency oxygen reduction electrocatalysts based on Pt-containing quinary high entropy alloys, LLMs supply elemental libraries and combinatorial blueprints, laying the foundation for one-step fabrication, high throughput testing, activity screening, and theoretical validation, thus accelerating the discovery of high performance multi element catalytic materials. Future efforts will focus on improving decision-making reliability, addressing the limitation that AI cannot fully

replace researchers' professional judgment, and advancing the transition of AI agents from "auxiliary R&D" to "collaborative, human-supervised autonomous R&D", so as to provide intelligent impetus for the comprehensive advancement of the 4th paradigm (Figure 4d and 4e).^{5, 47} Meanwhile, existing LLMs face prominent challenges such as factual hallucination during literature reasoning, uncertain output reliability, inconsistent experimental reproducibility, and expensive practical validation when guiding material screening and experimental design.⁴⁸

Full-process closed-loop system

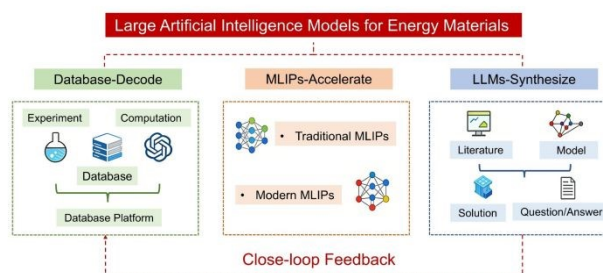


Figure 5 A Closed-loop framework of databases, MLIPs, and LLMs for energy materials discovery, presenting a closed-loop framework that integrates three core pillars of large AI models for energy materials: database-decode for fusing experimental and computational data *via* digital platforms, MLIPs-accelerate for advancing from traditional to modern MLIPs, and LLMs-synthesize for extracting knowledge from literature to guide material design and problem-solving.

The whole-process closed-loop system for intelligent energy materials research and development integrates experimental databases, ML models, and LLMs (AI agents).⁴⁹ The database supports data decoding, ML enables accelerated prediction, and LLMs accomplish comprehensive analysis. The deep integration of AI and experiments forms the core driving force of the system (Figure 5).

Within this framework, high-quality structured experimental data is essential for all intelligent algorithms and models. Material testing and performance testing produce raw experimental data, which is standardized and stored in the material database to form reusable data resources. The database integrates multi-source experimental information, extracts the inherent relationships among material structure, composition and performance, and supplies high-quality training samples and feature inputs for ML models. On this basis, ML models accurately predict high-potential candidate materials by learning mapping relationships in the data, providing clear guidance for experiments and reducing unnecessary trial and error. Experimental results obtained from validation are fed back to the database in real time to supplement performance data, calibrate model deviations and improve prediction reliability. LLMs further combine experimental data, model outputs and literature knowledge to generate new research hypotheses and experimental plans for continuous iteration. To quantify this closed-loop workflow, clear specifications are defined for each functional module: Structured experimental data and standardized computational



datasets serve as core module inputs, while optimized crystal structures, predicted material properties and refined preparation routes are defined as main outputs. Deviations between experimental measured values and model predictions exceeding preset error thresholds act as key feedback triggers to update database records and retrain ML/LLM models. The core decision criteria are defined by whether property prediction errors fall within pre-set acceptable tolerance ranges. Typical success metrics include reduced experimental trial counts, improved prediction accuracy of target performance and enhanced synthesis reproducibility of candidate materials. The system realizes highly integrated and autonomous operation of the entire research and development process, eliminates obstacles in traditional research modes, and establishes a positive cycle of data, model, experiment, and knowledge.⁴

Only through the deep integration of high-quality experimental data and AI can the new materials research paradigm be achieved, featuring intelligence-driven, high-efficiency R&D, autonomous closed-loop, paradigm shift, and precision innovation. Experimental data lays a solid foundation for AI models, while AI extracts structure and performance relationships to guide experimental design and screen promising material systems. This synergistic data-model-experiment mode overcomes the problems of fragmented processes and high costs in traditional research, promotes the transformation from experience-driven to intelligent closed-loop development, and finally realizes efficient and precise materials innovation.

Summary and Outlook

Focusing on the intelligent R&D of energy materials, this *Perspective* presents the functions and existing challenges of each core module based on the idea that **"databases are the core resource, ML models are the key method, LLMs are the knowledge linkage bridge, and full-process closed-loop systems are the integration and implementation carrier"**. Furthermore, it clarifies the core value of these components in breaking through the bottlenecks of traditional R&D and promoting the transformation of R&D models. Advances along these research routes will substantially cut experimental costs and shorten the R&D cycle of advanced energy materials. Such systematic upgrades also lay a solid foundation for the large-scale industrialization of next-generation energy materials. Nevertheless, several practical barriers restrain immediate large-scale deployment, including inconsistent experimental data standards among different laboratories, high computational overhead of multi-scale MLIP simulations, and prohibitive equipment cost restricting autonomous experimental setup popularization. To guide subsequent development, three-tier milestones are defined: unifying open material data specifications within two years, realizing fully automatic cross-module closed-loop coordination in 3–5 years, and establishing pilot intelligent synthesis production lines in 5–8 years. Accordingly, top implementation priorities cover dataset standardization rules, lightweight MLIP algorithm optimization, and low-cost combination of AI algorithms and bench-scale experimental equipment. Future research is

expected to focus on the standardized construction of databases, the optimization of model performance and the in-depth integration of multiple modules, improve the collaborative mechanism of the full-process closed-loop system, solve the "pain points" related to data and models, promote the transformation of AI agents from auxiliary R&D to collaborative, human-supervised autonomous R&D, and provide support for the efficient discovery and industrial transformation of high-performance energy materials.

Conflicts of interest

There is no conflict of interest to declare.

Data availability

No primary research data were generated in this Perspective. All information discussed is available from the cited literature.

Acknowledgements

This work was supported by JSPS KAKENHI (No. JP25H01508), Suzhou MatSource Technology Co., Ltd. (Suzhou, China), and the Gusu Laboratory of Materials (grant number Y2501).

Notes and references

1. S. Chu, Y. Cui and N. Liu, *Nature Materials*, 2017, 16, 16–22.
2. P. Bieuville, G. Majeau-Bettez and A. de Bortoli, *Nature Sustainability*, 2026, 9, 419–430.
3. A. Manthiram and Z. Cui, *Nature Energy*, 2026, <https://doi.org/10.1038/s41560-025-01963-x>.
4. D. Zhang, X. Jia, Y. Wang, H. Liu, Q. Wang, S.-H. Jang, D. Shah, S. Ye, H. B. Tran and H. Li, *Chemical Science*, 2026, <https://doi.org/10.1039/D5SC09229A>.
5. D. Zhang, Y. Chen, C. Liu, Y. Liu, H. Xin, J. Peng, P. Ou and H. Li, *Angewandte Chemie International Edition*, 2026, n/a, e26150.
6. J. Peng, D. Schwalbe-Koda, K. Akkiraju, T. Xie, L. Giordano, Y. Yu, C. J. Eom, J. R. Lunger, D. J. Zheng, R. R. Rao, S. Mui, J. C. Grossman, K. Reuter, R. Gómez-Bombarelli and Y. Shao-Horn, *Nature Reviews Materials*, 2022, 7, 991–1009.
7. L. Li, K. Xu, R. Su, H. Gu and P. Ma, *AI Agent*, 2025, 1, 8.
8. T. Yao, J. Huang, Y. Yan, Y. Yang, Z. Wang, X. Shao, Z. Gao and W. Yang, *AI Agent*, 2025, 1, 9.
9. Z. Lu, *Materials Reports: Energy*, 2021, 1, 100047.
10. Y. Mishin, *Acta Materialia*, 2021, 214, 116980.
11. S. S. Chowha, R. Alvi, S. S. Rahman, M. A. Rahman, M. A. K. Raiaan, M. R. Islam, M. Hussain and S. Azam, *Artificial Intelligence Review*, 2026, 59, 71.
12. Y. Su, X. Wang, Y. Ye, Y. Xie, Y. Xu, Y. Jiang and C. Wang, *Chemical Science*, 2024, 15, 12200–12233.
13. B. Kalita, H. Gokcan and O. Isayev, *Nature Computational Science*, 2025, 5, 1120–1132.



14. Y. Zhuang, X. Yang, C. Zhang, X. Jia, D. Zhang, M. Li, T. Yao, J. Peng, Z. Gao, W. Yang and H. Li, *Precision Chemistry*, 2026.
15. D. Zhang and H. Li, *Molecular Chemistry & Engineering*, 2025, 1, 100003.
16. Y. Wang, Q. Wang, S.-H. Jang, E. J. Cheng and H. Li, *Chemical Communications*, 2026, 62, 9536–9549.
17. Z. Cheng, X. Huang, H. Liu, J. Du, P. Ma, H. Yin, D. Zhang, H. Li and Y. Chen, *AI Agent*, 2026, 2, 8.
18. M. Ishii, T. Ito, H. Sado and I. Kuwajima, *Science and Technology of Advanced Materials: Methods*, 2024, 4, 2354649.
19. X. Jia, A. Aziz, Y. Hashimoto and H. Li, *Science China Materials*, 2024, 67, 1173–1182.
20. D. Zagorac, H. Müller, S. Ruehl, J. Zagorac and S. Rehme, *Journal of Applied Crystallography*, 2019, 52, 918–925.
21. S. Gražulis, A. Daškevič, A. Merkys, D. Chateigner, L. Lutterotti, M. Quirós, N. R. Serebryanaya, P. Moeck, R. T. Downs and A. Le Bail, *Nucleic Acids Research*, 2012, 40, D420–D427.
22. Y. Yu, S. Wu, L. Zhang, S. Xu, C. Dai, S. Gan, G. Xie, G. Feng and B. Z. Tang, *Biomaterials*, 2022, 280, 121255.
23. A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder and K. A. Persson, *APL Materials*, 2013, 1, 011002.
24. K. T. Winther, M. J. Hoffmann, J. R. Boes, O. Mamun, M. Bajdich and T. Bligaard, *Sci Data*, 2019, 6, 75.
25. R. Tran, J. Lan, M. Shuaibi, B. M. Wood, S. Goyal, A. Das, J. Heras-Domingo, A. Kolluru, A. Rizvi, N. Shoghi, A. Sriram, F. Therrien, J. Abed, O. Voznyy, E. H. Sargent, Z. Ulissi and C. L. Zitnick, *ACS Catalysis*, 2023, 13, 3066–3084.
26. D. Zhang, Z. Bao, Y. Chu, Z. Guo, X. Jia, Q. Jiang, H. Liu, T. Liu, T. Lu, Y. Lu, D. Devang Shah, Y. Wang, Y. Wang, Y. Wang, S. Ye, S. Ying, Z. Yu, L. Zhang, S. Zhao and H. Li, *ChemRxiv*, 2026.
27. F. Yang, E. Campos dos Santos, X. Jia, R. Sato, K. Kisu, Y. Hashimoto, S.-i. Orimo and H. Li, *Nano Materials Science*, 2024, 6, 256–262.
28. F. Yang, Q. Wang, E. J. Cheng, D. Zhang and H. Li, *Computers, Materials and Continua*, 2024, 81, 3413–3419.
29. S.-H. Jang, D. Zhang, H. B. Tran, X. Jia, K. Konno, R. Sato, S.-i. Orimo and H. Li, *Chemical Science*, 2025, 16, 23111–23120.
30. S.-H. Jang, D. Zhang, X. Jia, H. B. Tran, L. Zhang, R. Sato, Y. Hashimoto, T. Sato, K. Konno, S.-i. Orimo and H. Li, *undefined*, 2026, <https://doi.org/10.48550/ARXIV.2603.14139>.
31. J. Burner, J. Luo, A. White, A. Mirmiran, O. Kwon, P. Boyd, S. Maley, M. Gibaldi, S. Simrod, V. Ogden and T. Woo, *ARC-MOF: A Diverse Database of Metal-Organic Frameworks with DFT-Derived Partial Atomic Charges and Descriptors for Machine Learning*, 2022.
32. Y. Qiu, L. Wang, C. Liu, X. Zhang, Y. Tian and Z. Zhou, *Materials Today*, 2025, 91, 103–123.
33. M. A. Akhound, T. M. Boland, M. O. Sauer, M. Batzill, M. A. Bokinala, S. Canulescu, Y. Gogotsi, P. Hofmann, A. Kis, J. Lu, T. Michely, S. Raza, W. Ren, J. A. Robinson, Z. Sofer, J. H. Teng, S. Ulstrup, M. Zhao, X. Zhao, J. J. Mortensen, T. Olsen and K. S. Thygesen, *undefined*, 2026, <https://doi.org/10.48550/ARXIV.2603.05083>.
34. Y. Lin, P. Ou, Y. Lin and P. Ou, *AI Agent*, 2025, 1, 3.
35. J. Behler and M. Parrinello, *Physical Review Letters*, 2007, 98, 146401. VIEW ARTICLE ONLINE
DOI: 10.1039/D6DD00218H
36. J. P. Perdew, K. Burke and M. Ernzerhof, *Physical Review Letters*, 1996, 77, 3865–3868.
37. V. L. Deringer, A. P. Bartók, N. Bernstein, D. M. Wilkins, M. Ceriotti and G. Csányi, *Chem Rev*, 2021, 121, 10073–10141.
38. P. Rowe, G. Csányi, D. Alfè and A. Michaelides, *Physical Review B*, 2018, 97, 054303.
39. I. Batatia, D. a. P. e. Kovács, G. N. C. Simm, C. Ortner and G. Csányi, *ArXiv*, 2022, abs/2206.07697.
40. D. Zhang, P. Yi, X. Lai, L. Peng and H. Li, *Nature Communications*, 2024, 15, 344.
41. A. L. Maulana, S. Han, Y. Shan, P.-C. Chen, C. Lizandara-Pueyo, S. De, K. Schierle-Arndt and P. Yang, *Journal of the American Chemical Society*, 2025, 147, 10268–10278.
42. Y. Wang, Z. Wu, Y. Jiang, D. Zhang, Q. Wang, C. Wang, H. Li, X. Jia, J. Fan and H. Li, *Advanced Functional Materials*, 2025, 35, e06314.
43. D. Zhang, X. Jia, H. B. Tran, S. H. Jang, L. Zhang, R. Sato, Y. Hashimoto, T. Sato, K. Konno, S.-i. Orimo and H. Li, *Chemical Science*, 2026, 17, 3031–3042.
44. Y. Zou, A. H. Cheng, A. Aldossary, J. Bai, S. X. Leong, J. A. Campos-Gonzalez-Angulo, C. Choi, C. T. Ser, G. Tom, A. Wang, Z. Zhang, I. Yakavets, H. Hao, C. Crebolder, V. Bernales and A. Aspuru-Guzik, *Matter*, 2025, 8, 102263.
45. X. Jia, D. Zhang, Y. Lu, Q. Wang and H. Li, *AI Agent*, 2026, 2, 1.
46. Q. Wang, F. Yang, Y. Wang, D. Zhang, R. Sato, L. Zhang, E. J. Cheng, Y. Yan, Y. Chen, K. Kisu, S.-i. Orimo and H. Li, *Angewandte Chemie International Edition*, 2025, 64, e202506573.
47. D. Balcells, *ACS Catalysis*, 2025, 15, 16412–16420.
48. Y. Wang, Y. Wang and H. Li, *EES Catalysis*, 2026, <https://doi.org/10.1039/D6EY00079G>.
49. J. Peng, C. Liu, Y. Luo and K. Dandapat, *AI Agent*, 2025, 1, 5.



Data Statement

View Article Online
DOI: 10.1039/D6DD00218H

No primary research data were generated in this Perspective. All information discussed is available from the cited literature.

