Commentary | Open Access

Journal of Materials Informatics

Liu. *J. Mater. Inf.* **2026**, *6*, 43 **DOI:**10.20517/jmi.2026.36

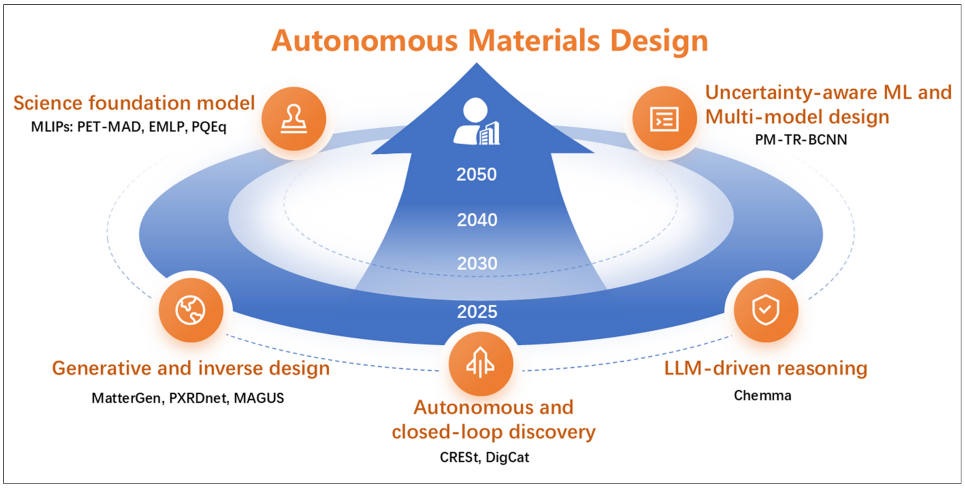
Commentary on ten selected 2025 papers in AI for Materials Science

Yi Liu*

Citation: Liu, Y. Commentary on ten selected 2025 papers in AI for Materials Science. *J. Mater. Inf.* **2026**, *6*, 43. https://dx.doi.org/10.20517/jmi.2026.36

Received: 28 May 2026
First Decision: 15 Jul 2026
Revised: 26 Jul 2026
Accepted: 3 Aug 2026
Published: 27 Aug 2026

Academic Editor: Xingjun Liu
Copy Editor: Pei-Yun Wang
Production Editor: Pei-Yun Wang



The integration of artificial intelligence into materials science has matured significantly in 2025, transitioning from surrogate property prediction to foundational atomistic models, generative inverse design, and autonomous experimental discovery. The following ten papers[1-10], published in high-impact journals, exemplify the current frontiers, methodological diversity, and remaining challenges of this interdisciplinary domain.

PET-MAD (NATURE COMMUNICATIONS[1])

PET-MAD[1] introduces a lightweight, universal interatomic potential based on the Point Edge Transformer, trained on the Massive Atomistic Diversity (MAD) dataset. Despite a compact training set, it achieves competitive accuracy across inorganic/organic solids, molecules, and surfaces. Its efficiency, stability, and built-in uncertainty quantification enable advanced simulations (e.g., ferroelectric phase transitions, ionic transport) with near-quantitative reliability and straightforward fine-tuning via low-rank adaptation (LoRA).

Significance: Demonstrates that structural/chemical diversity and internal consistency in training data can outweigh dataset size for developing general-purpose machine learning (ML) potentials.

Materials Genome Institute, Shanghai University, Shanghai 200444, China.

*Correspondence to: Prof. Yi Liu, Materials Genome Institute, Shanghai University, Shanghai 200444, China. E-mail: yiliu@shu.edu.cn

FOUNDATION ML POTENTIAL WITH POLARIZABLE LONG-RANGE INTERACTIONS (NATURE COMMUNICATIONS[2])

Gao *et al.* present a foundation ML interatomic potential combining an equivariant graph neural network with a polarizable charge equilibration (PQEq) scheme for explicit long-range electrostatics and polarization[2]. Trained across the periodic table (up to Pu) on MPtrj, it captures ionic/Coulomb interactions beyond cut-off, molecular response to electric fields, and polarization effects - outperforming non-long-range variants and baselines on charged systems. Applications include Li-ion diffusion in c-LLZO, $BaTiO_3$ phase transitions, and reactive molecular dynamics (MD) of solid electrolyte interphase (SEI) formation in solid-state batteries. Finetuning achieves *ab initio* accuracy for targeted systems.

Significance: Advances foundation ML potentials by rigorously and efficiently incorporating transferable long-range physical interactions, enabling accurate large-scale MD for ionic, polar, and interfacial materials problems previously intractable with local-only machine learning interatomic potentials (MLIPs).

EMLP (NATURE CATALYSIS[3])

The element-based machine learning potential (EMLP)[3] employs a random exploration via imaginary chemicals optimization (REICO) sampling strategy that focuses on diverse local atomic environments rather than system-specific structures. Trained on small, randomized configurations and their optimization trajectories, the Ag-Pd-C-H-O EMLP achieves density functional theory (DFT)-level accuracy for heterogeneous catalysis (e.g., CO oxidation, Fischer-Tropsch, solvent effects) and extends to organic reactions and liquid methanol.

Significance: Proposes a paradigm shift: ML potentials can be trained to learn transferable interatomic interactions, enabling reactivity and generality previously limited to system-specific models.

MATTERGEN (NATURE[4])

MatterGen[4] is a diffusion-based generative model for inorganic crystals, jointly denoising atom types, coordinates, and lattice. Fine-tuned via adapter modules, it conditions generation on chemistry, symmetry, and scalar properties (magnetic density, bandgap, bulk modulus). It generates stable, novel structures (> 75% within 0.1 eV/atom above hull) and rediscovers known ICSD entries. One generated $TaCr_2O_6$ was synthesized, with experimental bulk modulus within 20% of the target.

Significance: Provides a foundational step toward controllable "inverse materials design", though bias toward low-symmetry structures for larger cells remains a limitation.

MAGUS 2.0 (NATURE COMPUTATIONAL SCIENCE[5])

MAGUS 2.0[5] accelerates crystal structure prediction (CSP) by embedding the symmetry principle: a space group miner biases sampling toward supergroups of low-energy structures, while a graph-theory-based fragment reorganizer preserves favorable local environments. Symmetry-kept mutation maintains global symmetry. Benchmarks show up to 4× fewer structures needed to find ground states; it successfully predicts complex systems [e.g., violet phosphorus, Si(111)-(7 × 7) surface].

Significance: Highlights the value of incorporating physical priors (symmetry, local motifs) into evolutionary/search algorithms to navigate high-dimensional potential energy surfaces efficiently.

CHEMMA (NATURE MACHINE INTELLIGENCE[6])

Chemma[6], a fine-tuned LLaMA-2-7B model, functions as a generative assistant for organic synthesis. It handles forward/retrosynthesis, yield/selectivity prediction, and condition generation, outperforming prior art and GPT-4 on multiple benchmarks. Integrated into an active learning loop, it explores open reaction spaces, optimizing an unreported Suzuki-Miyaura coupling in only 15 wet experiments (67% yield).

Significance: Establishes large language models (LLMs) as actionable partners in synthetic chemistry, reducing reliance on DFT and enabling efficient exploration beyond predefined condition libraries.

CREST (NATURE[7])

CRESt (Copilot for Real-world Experimental Scientists)[7] is a multimodal robotic platform combining large multimodal models [chemical compositions, text, scanning electron microscopy (SEM) images], knowledge-assisted Bayesian optimization (KABO), and vision-language model (VLM) diagnostics. Applied to formate oxidation electrocatalysis, it explored > 900 compositions and 3,500 tests in 3 months, discovering an octonary alloy with 9.3× cost-specific performance improvement. VLM-driven anomaly diagnosis enhances reproducibility.

Significance: Exemplifies the closing loop between AI-driven hypothesis generation, automated synthesis/characterization, and self-correcting experimentation, moving toward autonomous materials discovery labs.

CLOSED-LOOP FRAMEWORK FOR BIFUNCTIONAL METAL OXIDE CATALYSTS (JACS[8])

This work presents a three-stage closed-loop[8]: (i) data mining (DigCat platform) + surface state analysis + microkinetic modeling for candidate selection; (ii) synthesis and electrochemical testing; (iii) advanced characterization [synchrotron, transmission electron microscopy (TEM), X-ray photoelectron spectroscopy (XPS)]. The loop identifies $RbSbWO_6$ as a stable, bifunctional [oxygen evolution reaction (OER)/hydrogen evolution reaction (HER)] acidic water-splitting catalyst, validated experimentally and fed back into the database.

Significance: Demonstrates a data-driven, theory-guided, experiment-validated workflow that systematically integrates computation and experimentation for electrocatalyst discovery.

UNCERTAINTY-INFORMED ML FOR CREEP-RESISTANT STEEL DESIGN (ACTA MATERIALIA[9])

Wang *et al.* propose a PM-TR-BCNN framework that integrates physical metallurgy (precipitate coarsening factor PF), transfer learning (short-time tensile → creep performance), and Bayesian convolutional neural networks (CNNs) to predict creep life and guide alloy design[9]. Unlike deterministic ML, it quantifies prediction uncertainty, enabling risk-aware optimization. Combined with a genetic algorithm, the model balances creep-life maximization and uncertainty minimization, yielding three new martensitic steels experimentally validated at 650 °C/140 MPa. The best design (D2 alloy) achieved ~562 h predicted *vs.* 540-616 h tested, with low uncertainty (± 0.27 log units)[9].

Significance: Demonstrates that embedding domain-informed features and uncertainty awareness into ML potentials enhances reliability, extrapolation, and practical alloy design - addressing a key gap in data-scarce materials informatics.

AB INITIO NANOCRYSTAL STRUCTURE SOLUTION FROM PXRD VIA DIFFUSION MODELS (NATURE MATERIALS[10])

Guo *et al.* introduce PXRDnet[10], a conditional diffusion model trained on 45,229 structures, to solve nanocrystal structures (≥ 10 Å) from broadened powder X-ray diffraction (PXRD) patterns and chemical formulas. It generates multiple candidate structures via Langevin dynamics, refines them with Rietveld, and succeeds on simulated nanocrystals across all seven crystal systems (average post-refinement R-factor ~7% for 100 Å cases). It also generalizes to 15 experimental PXRD patterns. An open benchmark (MP-20-PXRD) is released.

Significance: Provides the first end-to-end, uncertainty-aware, generative AI solution to the long-standing "nanostructure problem" in crystallography, enabling structure determination where traditional methods fail due to peak broadening and information loss.

SYNTHESIS AND OUTLOOK

Collectively, these 2025 contributions illustrate a rapidly consolidating "AI for Materials" ecosystem evolving along several convergent directions but with limitations as follows:

• **Foundation models and transferability:** Universal ML potentials (PET-MAD, EMLP, polarizable foundation potential)[1-3] and generative models (MatterGen, PXRDnet)[4,5,10] emphasize broad applicability via finetuning, LoRA, and physically grounded representations (symmetry, equivariance, long-range electrostatics). However, they face critical limits: local-cutoff message passing fails for disconnected fragments, fixed-parameter charge equilibration schemes struggle with redox-active systems, and pretraining biases (e.g., MPtrj) degrade out-of-distribution performance. Generative diffusion models overproduce low-symmetry structures, cannot guarantee synthesizability, and falter beyond their training distribution (e.g., > 20 atoms/unit cell). PXRDnet[10] requires known chemical formulas and clean experimental data, while transfer learning (PM-TR-BCNN[9]) suffers uncertainty collapse under extreme extrapolation. Collectively, these models suit in-distribution inorganic crystals, single-phase solids, and nanocrystals ≥ 10 Å, but remain unreliable for polymers, disordered alloys, high-pressure regimes, or safety-critical applications without rigorous experimental validation.

• **Generative and inverse design:** Diffusion-based generation (MatterGen, PXRDnet)[4,10] and symmetry-guided search (MAGUS)[5] enable inverse property-structure mapping but exhibit pronounced limitations: MatterGen[4] exhibits a bias toward low-symmetry structures and cannot guarantee thermodynamic or kinetic stability, with ~80% of generated candidates failing phonon stability checks in 2025 benchmarks. PXRDnet[10] requires known chemical formulas and struggles with experimental artifacts (e.g., container backgrounds) and structures exceeding 20 atoms per unit cell. MAGUS[5], while efficient, inherits the constraints of its underlying energy landscapes and may miss novel polymorphs outside seeded symmetry groups. Crucially, none embed synthesizability priors, often proposing structures unattainable under practical conditions. These methods are best suited for hypothesis generation within well-explored chemical spaces but remain unreliable for *de novo* discovery of synthesizable, defect-tolerant materials without tight integration of experimental feedback loops and stability validation.

• **LLM-driven reasoning:** Domain-adapted LLMs (Chemma)[6] demonstrate nascent capabilities in synthesis planning, condition interpretation, and interfacing with active learning and lab automation, yet remain constrained by intrinsic limitations: they frequently hallucinate reaction pathways or unrealistic conditions unsupported by chemical thermodynamics, require extensive domain-specific fine-tuning to surpass baseline performance, and lack true causal understanding of mechanistic steps. While effective within narrow,

well-curated reaction spaces, their reliability degrades sharply in unexplored chemical territories or multi-step syntheses involving air-sensitive intermediates. Furthermore, LLMs cannot autonomously validate hypotheses - experimental closed-loop validation remains essential to correct errors and ensure reproducibility. Consequently, these models serve best as assistive agents for ideation and protocol drafting in established domains, rather than as standalone decision-makers for novel, high-risk synthetic routes without human oversight and experimental verification.

• **Autonomous and closed-loop discovery:** Robotic platforms (CRESt)[7] and integrated computation–experiment workflows[8] demonstrate AI closing the loop from hypothesis generation to synthesis, characterization, and iterative refinement. Nevertheless, their operational scope is constrained by hardware versatility and data bottlenecks: CRESt's VLM[7] diagnostics excel at identifying macroscopic anomalies but struggle with subtle crystallographic defects or subsurface degradation, while its exploration remains confined to pre-defined compositional libraries (e.g., octonary alloys), limiting true *de novo* discovery. Similarly, the catalyst discovery workflow[8] relies heavily on existing databases (DigCat) and microkinetic models, inheriting their inherent biases and failing to capture complex, dynamic surface reconstructions under reaction conditions. Both approaches demand substantial upfront investment in standardized protocols and high-quality reference data; they perform optimally for incremental optimization within known material families but are less effective for discovering entirely novel structure–property relationships or handling highly air-sensitive syntheses without bespoke atmospheric controls. Consequently, these systems currently augment - rather than replace - expert intuition, serving as powerful accelerators for targeted optimization rather than fully autonomous explorers of uncharted chemical spaces.

• **Uncertainty-aware ML and robust design:** Creep-steel work[9] and the PM-TR-BCNN[9] framework exemplify the growing norm of quantifying prediction confidence to guide safe extrapolation, optimization, and experimental validation, particularly vital in data-scarce alloy design. However, this approach faces clear boundaries: epistemic uncertainty estimates collapse when extrapolating far beyond the training distribution - evidenced by alloy D3's severe overprediction of creep life due to excessive δ-ferrite formation and Laves phase precipitation unaccounted for in the precipitation factor (PF). Furthermore, current implementations often exclude critical long-term degradation mechanisms (e.g., Laves phase coarsening), restricting robust design to service regimes where dominant failure modes are captured by the embedded physical metallurgy descriptors. Consequently, while indispensable for ranking candidates and avoiding high-risk outliers within familiar chemical spaces, uncertainty-aware ML cannot yet replace conservative safety factors or exhaustive experimental validation for mission-critical components operating near material limits.

• **Comparative notes on ML potentials and generative crystal models**

ML potentials: Local-only models (PET-MAD[1]) prioritize computational speed and excel on bonded systems within their cutoff (~5 Å), but fail to capture long-range electrostatics in ionic systems. In contrast, physics-augmented potentials (PQEq model[2]) explicitly model charge equilibration and polarization, enabling accurate simulations of electrolytes and ferroelectrics where local models break down, albeit at a higher computational cost due to self-consistent charge solving. Compared to ensemble or evidential uncertainty methods, Bayesian CNNs (PM-TR-BCNN[9]) offer robust uncertainty quantification for small datasets but are less efficient for large-scale MD than graph-based potentials.

Generative crystal models: Symmetry-biased search algorithms (MAGUS[5]) efficiently navigate potential energy surfaces by respecting crystallographic constraints, making them ideal for ground-state structure prediction but prone to missing novel metastable phases. Conversely, diffusion-based models (MatterGen[4], PXRDnet[10]) generate diverse, chemically plausible structures without symmetry bias, excelling at *de novo* design and solving nanostructures from ambiguous data. However, MatterGen[4] often overproduces

low-symmetry structures and ignores synthesizability, while PXRDnet[10] requires known chemical formulas and struggles with experimental noise. Both diffusion models[4,10] demand significantly more computational resources than traditional CSP methods but offer a viable path to solving previously intractable "nanostructure problems" where symmetry-based approaches fail.

• Data foundations for AI-driven materials modeling

The efficacy of the discussed AI paradigms is inextricably linked to the heterogeneity, scale, and physical grounding of their training corpora, which span composition, structure, morphology, and functional properties:

Composition and atomic structure: Foundation models (PET-MAD[1]) predominantly rely on DFT-derived trajectories (e.g., MPtrj) covering the periodic table up to Pu. While broad, these datasets often suffer from inconsistencies in exchange-correlation functionals and basis sets, leading to biases in force and energy predictions. Generative models like MatterGen[4] and PXRDnet[10] are trained on experimentally validated, stable inorganic crystals (e.g., Materials Project), limiting their exposure to disordered, amorphous, metastable phases, or defect systems. A critical gap remains in data for structures with > 20 atoms per unit cell and complex solid solutions, restricting model generalizability.

Morphology and microstructure: Data bridging atomic structure to mesoscale morphology (e.g., grain boundaries, precipitates, porosity) are sparse and expensive to acquire via TEM/SEM/XRD. The creep-steel study[9] exemplifies the challenge: while thermodynamic databases (TCFE9) inform precipitate coarsening kinetics, they lack dynamic data on Laves phase evolution under service conditions. Similarly, PXRDnet[10] is constrained by simulated, idealized nanocrystal patterns, struggling with experimental artifacts like preferred orientation, strain broadening, and amorphous backgrounds that obscure diffraction peaks.

Functionality and properties: Datasets for functional properties (e.g., ionic conductivity, ferroelectric switching, catalytic activity) are often narrow and task-specific. CRESt's autonomous discovery[7] is bounded by the quality of prior catalytic performance data, while closed-loop workflows[8] depend on microkinetic models parameterized from limited experimental studies. Time-dependent degradation data (e.g., long-term creep, cyclic fatigue, SEI growth) are particularly scarce, forcing models like PM-TR-BCNN[9] to extrapolate from short-term tests - a process prone to uncertainty collapse.

Limitations and the path forward: Current data ecosystems suffer from a "rich-get-richer" problem - abundant data for stable, easy-to-synthesize compounds, but sparse data for reactive intermediates, defects, and extreme-condition performance. The field is shifting toward active learning loops (Chemma, CRESt) that strategically query experiments to fill data gaps, multimodal fusion (combining XRD, SEM, and spectroscopy), and physics-informed data augmentation to enhance model robustness beyond the confines of static, legacy datasets.

Compared to the broader scope and the key insights summarized for the DCTMD workshop in 2024[11], mainstream big atomic models - functioning as machine learning potentials - continue to advance in 2025 along two trajectories: toward more diverse yet compact datasets[1], and toward stricter physics-informed, data-driven methodologies that incorporate long-range interactions[2], polarizability, and magnetism. AI-driven autonomous laboratories are integrating increasingly automated hardware designs with more sophisticated workflows as interactive agents that combine computational tasks[7,8] and LLMs[6]. Generative AI, particularly when coupled with inverse design via global optimization, is being adopted more widely, with diffusion models playing a central role[4,5,10]. The most notable progress expected in the near future will likely be the transition of LLM-driven materials discovery toward greater scientific rigor and reliability[6].

Multimodal modeling is also expanding to encompass microstructural imaging, acoustic signals, diffraction[10], or absorption spectra.

The "AI for Materials Science" field is progressing toward a layered stack: transferable atomistic foundation models, generative design engines, control/optimization loops, benchmarking infrastructure, and LLM-driven agents. Persistent challenges include data quality and diversity and benchmarking; generalization to dynamics, excited states, and complex kinetics; multimodal/scale integration; small-data modeling; and industrial applications. The selected works collectively signal a discipline shifting from proof-of-concept toward layered, autonomous, and trustworthy materials discovery engines.

DECLARATIONS

Acknowledgement

Thank Prof. Baisheng Sa at Fuzhou University for the recommendation on the selected papers.

Authors' contributions

The author contributed solely to the article.

Availability of data and materials

Not applicable.

AI and AI-assisted tools statement

During the preparation of this manuscript, the AI tool DeepSeek (version V3, released 2024-12-26) was used solely for language editing. The tool did not influence the study design, data collection, analysis, interpretation, or the scientific content of the work. All authors take full responsibility for the accuracy, integrity, and final content of the manuscript.

Financial support and sponsorship

This work was supported by the National Natural Science Foundation of China (No. 52373227).

Conflicts of interest

Liu, Y. is an Editorial Board Member of the journal *Journal of Materials Informatics*, but was not involved in any steps of editorial processing, notably reviewer selection, manuscript handling, and decision-making.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

REFERENCES

1. Mazitov, A.; Bigi, F.; Kellner, M.; et al. PET-MAD as a lightweight universal interatomic potential for advanced materials modeling. *Nat. Commun.* **2025**, *16*, 10653. DOI PubMed PMC
2. Gao, R.; Yam, C.; Mao, J.; Chen, S.; Chen, G.; Hu, Z. A foundation machine learning potential with polarizable long-range interactions for materials modelling. *Nat. Commun.* **2025**, *16*, 10484. DOI PubMed PMC
3. Yang, C.; Wu, C.; Xie, W.; Xie, D.; Hu, P. General reactive element-based machine learning potentials for heterogeneous catalysis. *Nat. Catal.* **2025**, *8*, 891-904. DOI
4. Zeni, C.; Pinsler, R.; Zügner, D.; et al. A generative model for inorganic materials design. *Nature* **2025**, *639*, 624-32. DOI PubMed PMC
5. Han, Y.; Ding, C.; Wang, J.; et al. Efficient crystal structure prediction based on the symmetry principle. *Nat. Comput. Sci.* **2025**, *5*, 255-67. DOI PubMed

6. Zhang, Y.; Han, Y.; Chen, S.; et al. Large language models to accelerate organic chemistry synthesis. *Nat. Mach. Intell.* **2025**, *7*, 1010-22. DOI
7. Zhang, Z.; Ren, Z.; Hsu, C. W.; et al. A multimodal robotic platform for multi-element electrocatalyst discovery. *Nature* **2025**, *647*, 390-6. DOI PubMed
8. Jia, X.; Zhou, Z.; Liu, F.; et al. Closed-loop framework for discovering stable and low-cost bifunctional metal oxide catalysts for efficient electrocatalytic water splitting in acid. *J. Am. Chem. Soc.* **2025**, *147*, 22642-54. DOI PubMed PMC
9. Wang, C.; Wei, X.; van der Zwaag, S.; Wang, Q.; Xu, W. From creep-life prediction to ultra-creep-resistant steel design: an uncertainty-informed machine learning approach. *Acta. Mater.* **2025**, *292*, 121073. DOI
10. Guo, G.; Saidi, T. L.; Terban, M. W.; Valsecchi, M.; Billinge, S. J. L.; Lipson, H. *Ab initio* structure solutions from nanocrystalline powder diffraction data via diffusion models. *Nat. Mater.* **2025**, *24*, 1726-34. DOI PubMed
11. Kobayashi, R.; Amos, R. D.; Crawford, T. D.; et al. Unlocking the future of materials science: key insights from the DCTMD workshop. *J. Mater. Inf.* **2025**, *5*, 50. DOI