Commentary | Open Access

Chemical Synthesis

Veiga *et al. Chem. Synth.* **2026**, *6*, 66 **DOI:**10.20517/cs.2026.35

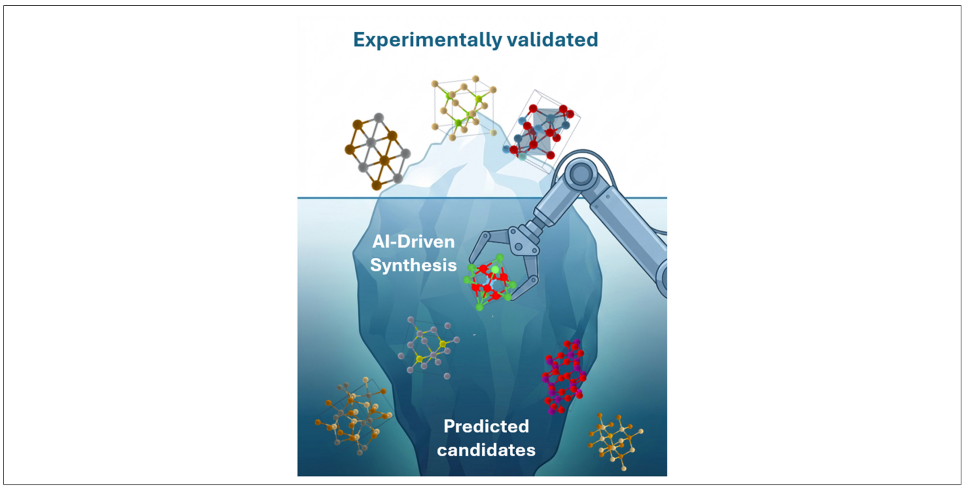
From prediction to realization: large language models and AI agents for inorganic materials discovery

Lionel Sebastian Veiga[1], Di Zhang[1,2], Tingyu Lu[1], Yiming Lu[1], Hao Li[1,*]

Citation: Veiga, L. S.; Zhang, D.; Lu, T.; Lu, Y.; Li, H. From prediction to realization: large language models and AI agents for inorganic materials discovery. *Chem. Synth.* **2026**, *6*, 66. https://dx.doi.org/10.20517/cs.2026.35

Received: 19 Jun 2026 **First Decision:** 16 Jul 2026 **Revised:** 19 Jul 2026 **Accepted:** 29 Jul 2026 **Published:** 13 Aug 2026

Academic Editor: Xiaoxin Zou **Copy Editor:** Pei-Yun Wang **Production Editor:** Pei-Yun Wang



FROM PREDICTION TO REALIZATION

The discovery of new inorganic materials is undergoing a structural transformation. Artificial intelligence (AI) has shifted the research paradigm from intuition-driven trial-and-error toward data-centric, algorithmically guided workflows encompassing hypothesis generation, experimental planning, characterization, and knowledge extraction[1,2]. At the frontier of this transformation, AI agents are beginning to engage directly with the synthesis problem that prediction alone cannot solve, converting computationally identified candidates into experimentally realized materials[3]. Here, we assess where this engagement produces genuine results and where fundamental challenges remain for inorganic materials synthesis.

The growing gap between prediction and synthesis

AI has made candidate generation and thermodynamic stability screening scalable, but it has not solved the more difficult problem of determining which predicted materials can be synthesized. Deep-learning models trained on millions of density functional theory (DFT) calculations now screen and generate millions of candidate

[1]Advanced Institute for Materials Research (WPI-AIMR), Tohoku University, Sendai 980-8577, Japan.
[2]The Frontier Research Institute for Interdisciplinary Sciences (FRIS), Tohoku University, Sendai 980-8577, Japan.

*Correspondence to: Prof. Hao Li, Advanced Institute for Materials Research (WPI-AIMR), Tohoku University, Sendai 980-8577, Japan. E-mail: li.hao.b8@tohoku.ac.jp

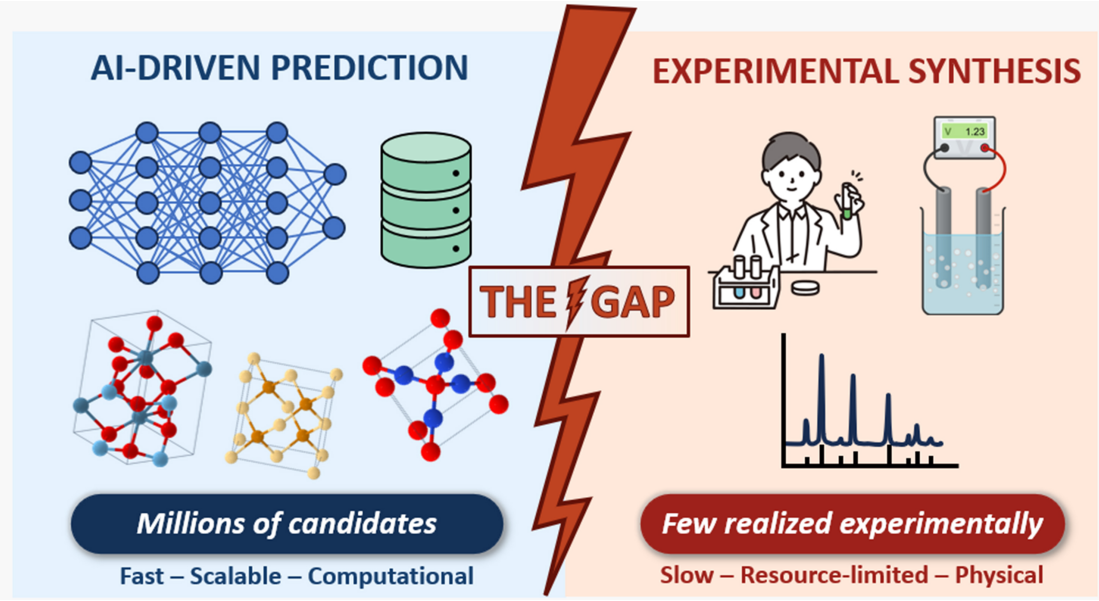


Figure 1. The prediction–synthesis gap in inorganic materials. AI has identified millions of thermodynamically stable structures, yet only several hundred had been experimentally realized at publication, highlighting the central challenge for AI in inorganic materials discovery. Crystal structures are visualized using the Materials Project[5], Copyright AIP Publishing LLC, licensed under a CC BY 3.0, and diagram elements created with Chemix (https://chemix.org). AI: Artificial intelligence.

structures predicted to be thermodynamically stable. As reported by Merchant *et al.*, among the approximately 2.2 million structures predicted to be stable, only 736 were matched to independently reported experimental structures at the time of publication. The scale of this gap is illustrated in Figure 1[4]. Predicting whether a hypothetical stable structure can be synthesized remains an open scientific problem: kinetic barriers, precursor incompatibility, and inaccessible reaction conditions mean that stability on paper does not guarantee a material can be made in the laboratory[6]. AI has widened the prediction-to-synthesis gap, and closing it is precisely the challenge that AI agents must now address.

Agents enter the inorganic synthesis laboratory

The closed-loop agent architecture, in which large language models (LLMs) orchestrate literature mining, experimental design, execution, characterization, and iterative refinement, has been in part practiced in organic synthesis[7-9]. A similar logic is now operational for inorganic materials: the A-Lab realized 36 of 57 targeted inorganic compounds through 17 days of continuous, machine learning (ML)-guided robotic synthesis[10]. The Digital Catalysis Platform (DigCat) implements a multi-step autonomous workflow integrating curated databases and literature with ML models and pH-dependent microkinetic modeling, deployed as a globally accessible cloud agent that self-improves through community feedback[11]. Applied to electrocatalyst discovery, this framework identified $RbSbWO_6$ as a new bifunctional non-noble metal oxide for acidic water splitting, synthesized, characterized, and validated experimentally to outperform many engineered catalysts reported in the literature[12]. At the bench scale, LLM-controlled robotic platforms have demonstrated synthesis of 13 structurally diverse inorganic compounds across coordination complexes, metal-organic frameworks (MOFs), nanoparticles, and polyoxometalates (including discovery of a previously unreported Mn-W polyoxometalate family through AI-guided exploration of synthesis space) while domain-trained LLMs have guided the synthesis of novel copper-based hydrogen storage MOFs in as few as three experimental iterations[13,14]. These are working systems producing new inorganic materials [Figure 2], though understanding where agents add the most value requires moving beyond headline results.

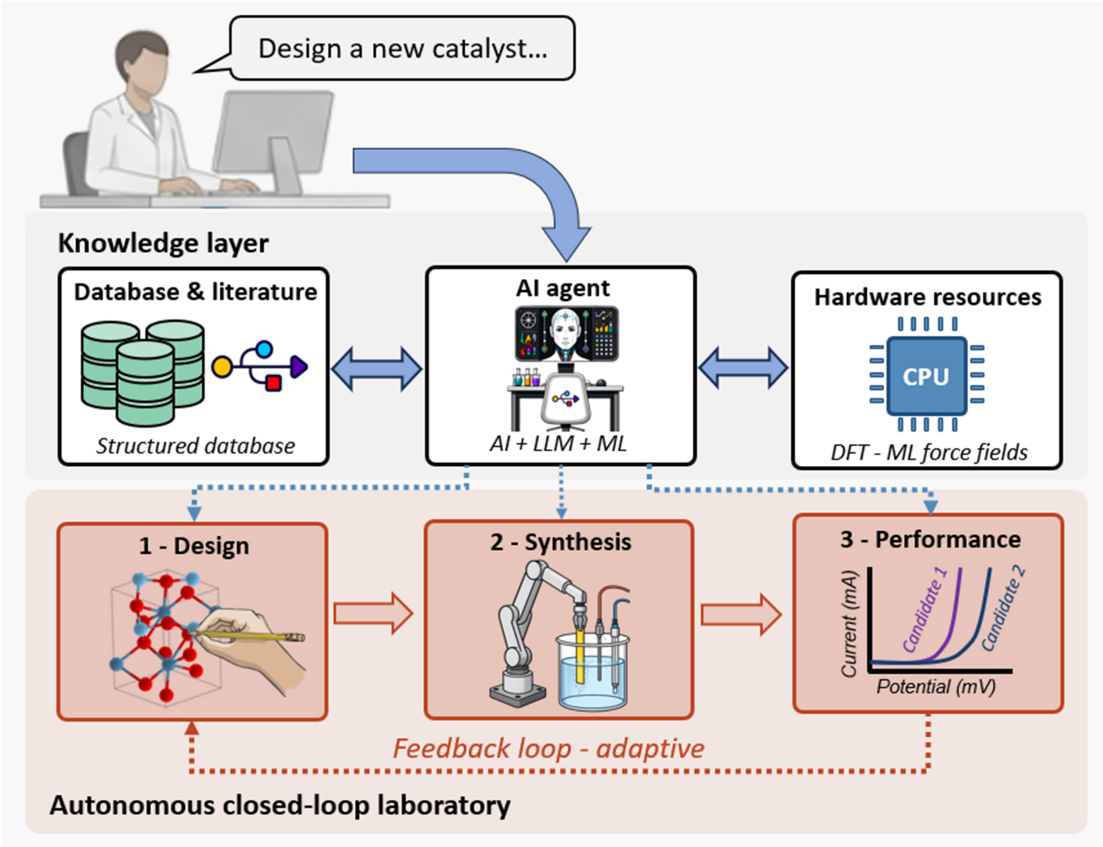


Figure 2. Closed-loop AI agent architecture for inorganic materials synthesis. An AI orchestrator integrating LLMs, ML models, and physics-based tools connects experimental databases and computational resources to an autonomous laboratory cycle comprising candidate design, robotic synthesis, and characterization, with experimental feedback refining each subsequent cycle. Diagram elements were created with Chemix (https://chemix.org); robot arm icon by Ehtisham Abid (Flaticon), adapted with ChatGPT. AI: Artificial intelligence; LLMs: large language models; ML: machine learning; DFT: density functional theory.

How AI agents drive autonomous experimentation

The architecture underlying these demonstrations shares a common logic: a data infrastructure layer that aggregates and standardizes experimental knowledge, an agent layer that reasons over that knowledge to propose candidates and synthesis conditions, and a hardware layer that executes, characterizes, and feeds results back into the loop. Synthesizability-prediction modules belong within the agent layer, acting as a triage step that screens thermodynamically stable candidates for kinetic and precursor feasibility before any are passed to the hardware layer for synthesis, reducing wasted experimental cycles on candidates unlikely to be realizable regardless of predicted stability[6]. At the data infrastructure level, a synthesis-ready database requires more than aggregated performance values: standardized synthesis conditions, reaction-specific metadata, and direct traceability to the primary literature, continuously updated through community contributions and AI-assisted curation. Platforms such as DigCat 4.0, spanning electrocatalysis, thermocatalysis, and photocatalysis, and literature-mined resources such as the solution-based synthesis dataset compiled by Wang *et al.*, illustrate this principle in practice[15,16]. Without this foundation, even the most capable agent cannot propose experiments that are experimentally actionable. Even where such platforms exist, they largely remain siloed: without shared data formats and ontologies, a database built for one chemistry cannot be readily queried or combined with another, limiting how far curated knowledge can travel across the field[17].

At the agent and hardware level, platforms such as CREST integrate a multimodal vision-language model with Bayesian optimization and robotic actuators to coordinate synthesis, characterization, and electrochemical testing end-to-end, exploring hundreds of compositions and thousands of tests to deliver an

eight-element electrocatalyst while diagnosing anomalies without human intervention[18,19]. Analogous workflows have used LLM-guided element selection to identify (RuNiFeMoCr)O_{3-4}, a stable acidic oxygen evolution catalyst validated at 1 A·cm^{-2} for over 150 h in a proton exchange membrane electrolyzer[20]. Across these systems, a consistent principle emerges: the closed loop generates its own training data with every cycle, and the quality of that data determines how rapidly the agent improves.

What agents genuinely contribute

The most defensible value proposition for AI agents in inorganic synthesis lies in knowledge aggregation and hypothesis generation, where the scale of the published literature exceeds human cognitive bandwidth. LLMs fine-tuned for materials science now extract structured synthesis knowledge from unstructured text to populate predictive databases at scale: Dagdelen *et al.* showed that entity-relation extraction models fine-tuned on as few as several hundred annotated abstracts can recover doping, MOF, and general materials data with high fidelity[21], but figure-centric data (the dominant format for electrochemical and thermodynamic performance) has remained systematically inaccessible to text-only approaches. The DIVE multi-agent workflow is designed to address this directly, achieving extraction accuracy gains exceeding 30% over open-source models, and enabling construction of a ~30,000-entry hydrogen storage database from ~4,000 publications[22]. Analogous workflows have accelerated materials discovery across chemistries and design tasks: photocatalysts synthesized within 5% of machine-learning-predicted values, inorganic crystal structures generated directly via autoregressive language modeling and validated against density-functional theory calculations, and MOF structures inverse-designed to experimentally validated materials in several synthesis iterations[14,23,24]. Conversational agent systems such as ChatMat interpret unstructured prompts through a manager agent that coordinates specialist sub-agents for property retrieval, computational design, and simulation, automating workflows from structure generation to potential-energy-surface construction[25]. Beyond extraction, agents contribute cross-dataset mechanistic reasoning invisible to single-study analysis: the MOFsyn agent, a retrieval-augmented generation framework, and other agents have integrated a non-intuitive stepwise reduction strategy that doubled active site density in Ni-loaded MOF catalysts, and have integrated specialist expertise fragmented across materials chemistry, electrochemistry, and cell engineering into solid-state battery development[26-28]. These examples share a common thread: agents excel at spotting patterns across more literature and data than anyone could read, but are not yet reliable at making the kind of consequential bench decision, where one wrong choice wastes materials or time. That is why agents currently add the most value before an experiment starts, not during it.

Where current systems fail

Three fundamental challenges constrain the current generation of inorganic synthesis agents. The first is a physics problem: agents reason about static structures, but synthesis transforms them. Data mining predicted Sb_2WO_6 would be unstable under alkaline oxygen reduction reaction (ORR) conditions, yet experimentally it undergoes electrochemical passivation to form a stable active surface[29]. This reflects a fundamental limitation of bulk thermodynamic screening that cannot model reaction-induced surface reconstruction, a limitation that only compounds as system complexity grows and design spaces increasingly exceed available training data. The second is a data structure problem: the synthesis literature is biased toward successes, and encodes critical quantitative data in figures rather than text. State-of-the-art multimodal models still fail systematically at spatial reasoning and quantitative interpretation of characterization outputs[30], and LLM-based extraction can introduce erroneous or unsupported values for quantities such as temperatures, stoichiometries, and synthesis conditions[17,31]. The third is a synthesizability problem: thermodynamically stable AI-generated structures are frequently unmakeable for reasons (e.g., kinetic traps, precursor incompatibility, phase changes) that energy calculations cannot capture[6].

Building the infrastructure for reliable human-agent collaboration

Progress in these systems depends critically on the quality of the underlying knowledge sources, but the deeper advantage of closed-loop design is that every experimental cycle (successful or not) generates data that refine the next recommendation[7,8,11,12,15,21]. The field currently loses most of this information because negative results go unreported and synthesis conditions are inconsistently documented. Recent text-mining efforts illustrate what becomes possible when this information is captured: a dataset of over 80,000 solid-state synthesis reactions, including nearly 19,000 documented impurity-phase outcomes, reveals reproducible synthesis routes and previously unexplored regions of the synthesis space[32]. Establishing community standards for reporting agent-assisted outcomes, including failures and condition-dependent variability, is what makes the loop self-improving.

The convergence of machine learning force fields, multimodal LLMs, and autonomous laboratory platforms points toward Digital Materials Ecosystems capable of accelerating functional materials discovery across domains[3]. The appropriate near-term model is not autonomous replacement of chemists but structured collaboration: agents handling knowledge aggregation and hypothesis generation at scales exceeding human bandwidth, while expert judgment remains essential at the bench, where errors are costly, and current AI systems remain unreliable. The materials that will define the next generation of clean energy, computing, and medicine may already exist as predictions in a database somewhere; the challenge, and the opportunity, is building the agent-driven infrastructure that can find, make, and explain them.

DECLARATIONS

Authors' contributions

Made substantial contributions to conception and design of the commentary and led the manuscript writing: Veiga, L. S.; Li, H.

Contributed to writing, scientific discussion, and manuscript revision: Zhang, D.; Lu, T.; Lu, Y.

Provided scientific guidance, supervision, and final approval of the manuscript: Li, H.

Availability of data and materials

Not applicable.

AI and AI-assisted tools statement

During the preparation of this work, the authors used Claude (Sonnet 5, Anthropic, released 2026-06-30) to assist with language editing and clarity improvements during manuscript revision. ChatGPT (GPT-5.5 Instant, OpenAI, released 2026-04-23) was used to adapt a stock icon used in Figure 2 and the graphic abstract. Neither tool influenced the study design, data collection, analysis, interpretation, or scientific content of the work. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the accuracy, integrity, and final content of the manuscript.

Financial support and sponsorship

We acknowledge the support from JSPS KAKENHI (Nos. JP25H01508 and JP25K01737, JP25K17991).

Conflicts of interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

REFERENCES

1. Cheng, M.; Fu, C. L.; Okabe, R.; et al. Artificial intelligence-driven approaches for materials design and discovery. *Nat. Mater.* **2026**, *25*, 174-90. DOI PubMed
2. Li, Y.; Wang, S.; Wang, J.; Qian, L.; Zhang, J. Artificial intelligence for Materials Science: transforming research paradigms. *Chem. Rev.* **2026**, *126*, 6776-803. DOI PubMed
3. Zhang, D.; Chen, Y.; Liu, C.; et al. Accelerating catalyst materials discovery with large artificial intelligence models. *Angew. Chem. Int. Ed. Engl.* **2026**, *65*, e26150. DOI PubMed PMC
4. Merchant, A.; Batzner, S.; Schoenholz, S. S.; Aykol, M.; Cheon, G.; Cubuk, E. D. Scaling deep learning for materials discovery. *Nature* **2023**, *624*, 80-5. DOI PubMed PMC
5. Jain, A.; Ong, S. P.; Hautier, G.; et al. Commentary: The Materials Project: a materials genome approach to accelerating materials innovation. *APL. Mater.* **2013**, *1*, 011002. DOI
6. Kim, S.; Schrier, J.; Jung, Y. Explainable synthesizability prediction of inorganic crystal polymorphs using large language models. *Angew. Chem. Int. Ed. Engl.* **2025**, *64*, e202423950. DOI PubMed PMC
7. Boiko, D. A.; MacKnight, R.; Kline, B.; Gomes, G. Autonomous chemical research with large language models. *Nature* **2023**, *624*, 570-8. DOI PubMed PMC
8. Ruan, Y.; Lu, C.; Xu, N.; et al. An automatic end-to-end chemical synthesis development platform powered by large language models. *Nat. Commun.* **2024**, *15*, 10160. DOI PubMed PMC
9. Bran, A. M.; Cox, S.; Schilter, O.; Baldassari, C.; White, A. D.; Schwaller, P. Augmenting large language models with chemistry tools. *Nat. Mach. Intell.* **2024**, *6*, 525-35. DOI PubMed PMC
10. Szymanski, N. J.; Rendy, B.; Fei, Y.; et al. An autonomous laboratory for the accelerated synthesis of novel materials. *Nature* **2023**, *624*, 86-91. DOI PubMed PMC
11. Zhang, D.; Jia, X.; Liu, H.; et al. Cloud synthesis: a global closed-loop feedback powered by autonomous AI-driven catalyst design agent. *AI. Agent.* **2025**, *1*, 2. DOI
12. 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
13. Huang, L.; Zhang, C.; Fu, Y.; et al. Natural-language-interfaced robotic synthesis for AI-copilot-assisted exploration of inorganic materials. *J. Am. Chem. Soc.* **2025**, *147*, 23014-25. DOI PubMed
14. Liu, Z.; Su, Y.; Wang, H.; et al. Domain-trained language model for inverse design and synthesis of high-performance hydrogen storage MOFs. *Angew. Chem. Int. Ed. Engl.* **2026**, *65*, e13366. DOI PubMed
15. Zhang, D.; Bao, Z.; Chu, Y.; et al. Digital catalysis platform as a gateway to big data and AI-powered innovations in catalysis. *Chem. Catal.* **2026**, *6*, 101775. DOI
16. Wang, Z.; Kononova, O.; Cruse, K.; et al. Dataset of solution-based inorganic materials synthesis procedures extracted from the scientific literature. *Sci. Data.* **2022**, *9*, 231. DOI PubMed PMC
17. Schilling-Wilhelmi, M.; Ríos-García, M.; Shabih, S.; et al. From text to insight: large language models for chemical data extraction. *Chem. Soc. Rev.* **2025**, *54*, 1125-50. DOI PubMed
18. 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
19. Peng, J.; Liu, C.; Luo, Y.; Dandapat, K. Accelerating multimetallic catalyst discovery with robotics and agentic AI. *AI. Agent.* **2025**, *1*, 5. DOI
20. Tu, Y.; Shan, X.; Chen, L.; et al. Accelerated discovery of highly stable ruthenium-based high-entropy oxides for acidic oxygen evolution. *Sci. Adv.* **2026**, *12*, eaed8479. DOI PubMed PMC
21. Dagdelen, J.; Dunn, A.; Lee, S.; et al. Structured information extraction from scientific text with large language models. *Nat. Commun.* **2024**, *15*, 1418. DOI PubMed PMC
22. Zhang, D.; Jia, X.; Tran, H. B.; et al. "DIVE" into hydrogen storage materials discovery with AI agents. *Chem. Sci.* **2026**, *17*, 3031-42. DOI PubMed PMC
23. Li, D.; Sun, X.; Sun, S.; et al. Prompt engineering accelerates the data-driven discovery of photocatalysts via an LLM-based model ensemble strategy. *Adv. Sci.* **2026**, *13*, e24215. DOI PubMed PMC
24. Antunes, L. M.; Butler, K. T.; Grau-Crespo, R. Crystal structure generation with autoregressive large language modeling. *Nat. Commun.* **2024**, *15*, 10570. DOI PubMed PMC
25. Lv, S.; Peng, L.; Jiao, S.; Yao, Y.; Wu, W.; Hu, W. ChatMat: a multi-agent chemist for autonomous material prediction and exploration. *Digital. Discovery.* **2026**, *5*, 2886-98. DOI
26. Lin, J.; Zhao, D.; Lu, S.; et al. Conversational large-language-model artificial intelligence agent for accelerated synthesis of metal-organic frameworks catalysts in olefin hydrogenation. *ACS. Nano.* **2025**, *19*, 23840-58. DOI PubMed

27. Zhong, Y.; Leng, Y.; Gu, Z.; et al. Breaking interdisciplinary barriers in solid-state battery research: BatteryAgent for multifaceted analysis. *J. Mater. Chem. A.* **2025**, *13*, 37031-43. DOI
28. Leng, Y.; Zhong, Y.; Gu, Z.; et al. Intelligent, Personalized Scientific Assistant Via Large Language Models For Solid-State Battery Research. *ACS. Mater. Lett.* **2025**, *7*, 1807-16. DOI
29. Jia, X.; Yu, Z.; Liu, F.; et al. Identifying stable electrocatalysts initialized by data mining: Sb_2WO_6 for oxygen reduction. *Adv. Sci.* **2024**, *11*, e2305630. DOI PubMed PMC
30. Alampara, N.; Schilling-Wilhelmi, M.; Ríos-García, M.; et al. Probing the limitations of multimodal language models for chemistry and materials research. *Nat. Comput. Sci.* **2025**, *5*, 952-61. DOI PubMed PMC
31. Bae, S.; Jeon, M.; Moon, H. R. Text mining in MOF research: from manual curation to large language model-based automation. *Chem. Commun.* **2025**, *61*, 11083-94. DOI PubMed
32. Lee, S.; Cruse, K.; Baibakova, V.; Ceder, G.; Jain, A. Text-mined dataset of solid-state syntheses with impurity phases using large language model. *Sci. Data.* **2025**, *12*, 1969. DOI PubMed PMC