Reprogramming synthesis

General-purpose artificial intelligence agents are at work in the materials chemistry laboratory

Zhiling Zheng

Experimental chemistry has long been defined by intuition, patience, and the painstaking execution of trial-and-error syntheses. Porous materials such as metal-organic frameworks (MOFs) show this clearly (1, 2). Their structures can, in principle, be drawn from enormous combinations of molecular building blocks to produce materials suited for clean water, air, and energy applications (3–5). Yet a substantial gap remains between the structures that can be imagined ("the conceptual space") and the materials that can reliably be made ("the realized space") with current synthetic methods (6). Without faster reasoning and experimentation tied together in one loop, much of this expansive design and synthesis landscape remains out of reach (7–9).

My research introduces a path toward predictive synthesis through a general-purpose artificial intelligence (AI) laboratory system built around three stages: read, design, and act (see

the figure). In this context, general-purpose AI refers to AI assistants or agents capable of interpreting chemical information, generating design hypotheses, and executing experimental decisions without being restricted to a single task or material system (10). These agents operate alongside human intuition within an interconnected loop that changes the way that researchers design and synthesize new materials to expand experimental reach, reduce failure rates, and unlock design-to-discovery timelines not feasible using traditional methods alone.

At the heart of this platform is the AI's ability to read chemistry like a chemist. The synthesis protocols and characterization outcomes for MOFs are often buried in prose, tables, and figures scattered across more than 50,000 academic papers containing 120,000 crystal structures amassed by three decades of prior MOF research (10). To extract these data at scale, we built natural language workflows that use large language models (LLMs) to parse synthesis procedures, identify reagents (e.g., metal precursors, linkers, solvents, and modulators), and extract detailed conditions (e.g., ratios, temperatures, and time) (11). Rather than rely on keyword heuristics, the literature data mining agents understand chemical name entities based on context and accept human natural language as input, which substantially lowers the barrier for synthetic chemists to use and apply AI in their workflows; our natural language workflow achieved more than 97% accuracy across diverse formats.

Experimental data often appear in figures; therefore, we extended the system to vision-language models (12) that read adsorption isotherms, diffraction patterns, and thermal profiles from figures. These outputs are digitized and merged with text-based data to ultimately yield a unified synthesis-structure-property dataset for MOFs. Manual curation of such a corpus with multimodality would have taken years; however, with our literature data mining agents, a workflow that is scalable, adaptable, and generalizable to many materials chemistry schemes was achieved within several hours.

With this structured knowledge base in hand, our hypothesis is that the way that chemists design organic linkers can be learned by AI. Through fine-tuning LLMs on a literature-based dataset of 3943 MOF linker transformations, we created generative design agents that propose new yet synthesizable building blocks for MOFs that respect topological and chemical constraints (13). As a result, these design agents proposed chemical structures that helped discover 10 new materials, referred to as the long-arm MOF family (LAMOF-1 through LAMOF-10), with computational simulations showing stronger water harvesting performances than those of state-of-the-art aluminum-based MOF adsorbents (4). These gains were not

achieved by traditional brute-force MOF linker screening but through targeted design guided by model suggestions learned from literature-based community knowledge (11).

The act of imagining a new material is only the beginning. The more difficult part is turning that design into a real crystal. The reaction space is large, and finding the right reaction parameters is often the limiting factor in MOF discovery. To tackle this, we built a multiagent system that acts as a group of experimental chemists (14), which plans reactions, communicates with laboratory automation, interprets results, and updates its ideas in real time. Agents' tasks include a planning stage informed by literature and learned priors, a control stage that converts natural language procedures into robotic commands, an analysis stage that evaluates powder x-ray diffraction patterns, and an optimization stage that proposes the next experiment using both in-context learning and Bayesian optimization. By learning from both successes and failures (15), the AI system improves its predictions as it moves through the synthesis landscape.

As a demonstration, we applied this closed-loop platform with predictive synthesis capability to the newly proposed materials LAMOF-1 and LAMOF-2 (13, 14), whose crystallization space spans more than 106 possible reaction conditions. Within 120 robotic experiments conducted within 2 weeks, the agents uncovered high-crystallinity conditions that previously would have required months of manual screening effort. The platform assigned confidence to its own suggestions, handled unexpected results, and updated its choices as the campaign progressed. What sets this platform apart is not a single model or algorithm but the behavior that emerges when a few agents work together to reason, much like an experienced human researcher would (15) but at a scale and pace that human intuition alone cannot reach.

Looking ahead, it is possible to envision a future where general-purpose agents become routine in materials chemistry laboratories. Crucially, their role is not to replace human creativity but to amplify it. By linking literature, design, and experimentation into a single system that learns from every step, AI agents change how we think about synthesis. They reprogram synthesis not only in the procedural sense but in the epistemological one: what materials we ask for, how we ask for them, and how quickly we can answer. In so doing, AI helps fulfill the promise of a true laboratory partner for predictive synthesis to greatly accelerate chemical and materials discovery in real time.

REFERENCES AND NOTES

1. H. Furukawa, K. E. Cordova, M. O'Keeffe, O. M. Yaghi, *Science* **341**, 1230444 (2013).
2. A. J. Howarth *et al*., *Chem. Mater.* **29**, 26 (2017).
3. Z. Chen *et al*., *Science* **368**, 297 (2020).
4. Z. Zheng *et al*., *Nat. Protoc.* **18**, 136 (2023).
5. J.-B. Lin *et al*., *Science* **374**, 1464 (2021).
6. H. Lyu, Z. Ji, S. Wuttke, O. M. Yaghi, *Chem* **6**, 2219 (2020).
7. B. A. Koscher *et al*., *Science* **382**, eadi1407 (2023).
8. D. A. Boiko, R. MacKnight, B. Kline, G. Gomes, *Nature* **624**, 570 (2023).
9. A. M. Bran *et al*., *Nat. Mach. Intell.* **6**, 525 (2024).
10. Z. Zheng *et al*., *Nat. Rev. Mater.* **10**, 369 (2025).
11. Z. Zheng, O. Zhang, C. Borgs, J. T. Chayes, O. M. Yaghi, *J. Am. Chem. Soc.* **145**, 18048 (2023).
12. Z. Zheng *et al*., *Digit. Discov.* **3**, 491 (2024).
13. Z. Zheng *et al*., *J. Am. Chem. Soc.* **145**, 28284 (2023).
14. Z. Zheng *et al*., *ACS Cent. Sci.* **9**, 2161 (2023).
15. Z. Zheng *et al*., *Angew. Chem. Int. Ed.* **62**, e202311983 (2023).