OPEN ACCESS

Materials Chemistry COMMENTARY

From automation to autonomy: Embracing multimodal intelligence in materials discovery

Haojie Liu[1] and Yonggang Yao[1,*]

[1]State Key Laboratory of Materials Processing and Die & Mould Technology, School of Materials Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China

*Correspondence: yaoyg@hust.edu.cn

Received: October 10, 2025; Accepted: December 5, 2025; Published Online: December 5, 2025; https://doi.org/10.59717/j.xinn-mater.2026.100178

Citation: Liu H. and Yao Y. (2026). From automation to autonomy: Embracing multimodal intelligence in materials discovery. *The Innovation Materials* **4**:100178.

The discovery of efficient and durable electrocatalysts underpins a broad spectrum of sustainable technologies, from renewable energy conversion and storage to carbon management and green chemical manufacturing. Yet progress remains constrained by the near-infinite complexity of chemical space: even small variations in composition, processing, or nanostructure can cause dramatic changes in catalytic performance.[1] Conventional trial-and-error or intuition-driven searches are slow and biased, while most optimization strategies remain unimodal, adjusting limited parameters against a single metric and failing to capture the nonlinear, multi-parameter interactions that govern catalytic behavior (Figure 1A).

Artificial intelligence (AI) and robotics have emerged as promising solutions. Several autonomous laboratories now couple machine learning with robotic synthesis and testing, accelerating the design–make–test–learn cycle and reducing human labor.[2,3] Yet most remain efficient "tools" rather than true "scientists." This efficiency defines automation: the fast execution of pre-programmed, repetitive tasks, typically relying on a single, simplified data stream (e.g., recipes and outputs). However, such systems remain blind to the broader scientific context—literature knowledge, microstructural cues from microscopy, or experimental subtleties that influence outcomes. Bridging this gap requires autonomy: the capacity for sophisticated, multimodal knowledge integration, reasoning, and most critically, self-diagnosis and self-correction of experimental anomalies to ensure reproducibility.

Now, writing in Nature, Li and colleagues recently reported a multimodal robotic platform, the Copilot for Real-world Experimental Scientists (CRESt), which represents a major step toward autonomous electrocatalysis.[4] Unlike earlier autonomous laboratories such as the A-Lab[2] for inorganic synthesis, Coley's flow-synthesis platform[3], or Boiko's language-model-driven system,[5] which remain confined to unimodal optimization and predefined search spaces, CRESt actively integrates multimodal knowledge—literature, structure, and composition—into a unified discovery loop. By embedding chemical priors, morphological cues, and experimental descriptors into an active learning framework, it demonstrates how AI and robotics can act as true copilots of scientific discovery rather than mere accelerators of repetitive tasks (Figure 1B).

At the algorithmic core of CRESt lies Knowledge-Assisted Bayesian Optimization (KABO), a refinement of Gaussian-process optimization designed to navigate dauntingly high-dimensional chemical spaces. KABO's architecture distinguishes it from other knowledge-embedded strategies, such as Transfer Learning Bayesian Optimization (TLBO), which typically transfers knowledge from a similar, pre-existing unimodal dataset, or Physics-Informed Gaussian Processes (PIGPs), which explicitly encode physical laws. KABO's novelty lies in its ability to seamlessly integrate multiple, heterogeneous data streams—namely, text embeddings (literature priors), image-derived features (microstructural cues), and compositional inputs—into a single, reduced-

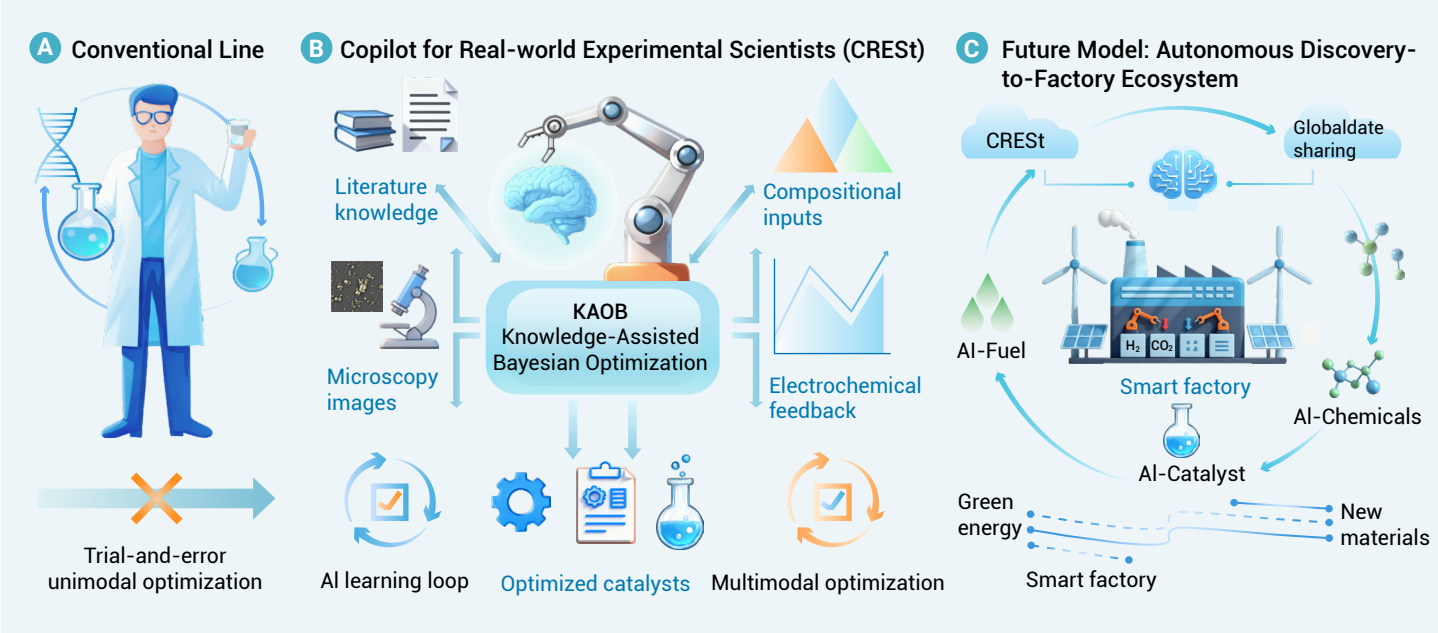

Figure 1. Paradigm shift of catalyst discovery models (A) Conventional line: traditional catalyst discovery relies on trial-and-error and unimodal optimization, leading to slow and biased progress. (B) Copilot for Real-world Experimental Scientists (CRESt): a multimodal AI–robotics framework that integrates literature knowledge, microscopy images, compositional inputs, and electrochemical feedback into Knowledge-Assisted Bayesian Optimization (KABO), enabling closed-loop learning and catalyst optimization. (C) Future model: vision of a fully automated smart factory powered by green energy.

dimensional latent space. Instead of operating blindly on elemental ratios, KABO leverages large multimodal models (LMMs) to compress heterogeneous data into a latent space enriched with prior knowledge. Language models extract literature-based descriptors of metallic elements, while automated microscopy supplies morphological features such as particle coverage and size distribution. When combined with compositional variables, these features create a fingerprint of candidate catalysts that is far richer than stoichiometry alone.

Equally important, CRESt's learning loop is inseparable from robotic execution. The platform is underpinned by a comprehensive infrastructure: liquid-handling robots prepare and mix precursors; carbothermal shock and electrodeposition methods generate nanostructured alloys; automated electro-

chemical workstations conduct rapid screening; and X-ray diffraction and microscopy provide structural feedback. These diverse modules, orchestrated by Python-based control, allow CRESt to function as a seamless closed-loop laboratory, continuously cycling from hypothesis to synthesis, measurement, and re-iteration without human interruption.

Perhaps the most striking advance is CRESt's use of vision–language models (VLMs) to overcome one of the least discussed but most critical obstacles in autonomous experimentation: reproducibility. In practice, real laboratories are rife with subtle errors—a misaligned pipette tip, a warped substrate, or a slight deviation in electrode geometry—that can silently compromise entire datasets. CRESt's VLMs combine computer vision with natural language reasoning to identify these anomalies in real time and propose corrective procedures, from detecting micrometer-scale shifts in carbon substrates to flagging irregular charring of laser-cut stages. This capacity to self-diagnose and self-correct transforms reproducibility from a chronic bottleneck into a manageable variable, marking a qualitative leap from automation to autonomy.

The capabilities of CRESt were validated in the electrochemical oxidation of formate. In just three months, the platform explored more than 900 unique chemistries and executed over 3500 electrochemical tests—an experimental throughput inconceivable for manual efforts. From the astronomical octonary Pd–Pt–Cu–Au–Ir–Ce–Nb–Cr space (on the order of 10^{17} possible compositions), CRESt identified catalysts with unprecedented performance. One optimized alloy delivered a 9.3-fold enhancement in cost-specific power density (defined as power density normalized by catalyst cost, mW USD^{-1} cm^{-2}, primarily driven by precious-metal content) compared with pure Pd, while simultaneously reducing precious-metal dependence—an outcome that balances performance with economic sustainability, a long-standing challenge in applied catalysis.

Beyond discovery, CRESt bridges exploration with mechanistic understanding, uniting rapid optimization with fundamental insight. It exemplifies a new paradigm where algorithms, robotics, and scientific reasoning converge to generate both materials and knowledge. In situ X-ray absorption spectroscopy revealed that Pd and Pt maintained metallic states under operating conditions, resisting deactivation through oxide formation. Complementary density functional theory calculations showed that alloying tuned the Pd d-band center downward, weakening adsorption of surface hydrogen and suppressing CO adsorption. These two canonical degradation pathways—hydrogen poisoning and CO poisoning—are well-known to block active sites and hinder formate oxidation kinetics to restrict turnover frequency and cause a rapid drop in current density. This dual evidence not only explained the superior activity but also highlighted how high-entropy alloying can engineer electronic structure for resilience.

The emergence of CRESt marks not just a technological advance but a paradigm shift in experimental science. By integrating multimodal data within a robotic framework, it moves beyond automation toward true autonomy. Unlike earlier platforms focused mainly on speed, CRESt delivers intelligence—embedding contextual knowledge that mirrors human reasoning and navigates chemical complexity with both breadth and depth. In electrocatalysis, this enables exploration of vast composition spaces while uncovering mechanistic insights that explain performance. The same multimodal strategy could accelerate discovery in oxygen evolution, CO_2 reduction, and ammonia oxidation, as well as broader materials research from solid electrolytes to photonics. As studies expand into high-entropy alloys and compounds with immense configurational diversity, platforms like CRESt may prove essential for turning combinatorial chaos into actionable knowledge.

Despite its promise, CRESt also underscores the formidable challenges that must be addressed before science becomes truly autonomous. The integration of heterogeneous data—ranging from literature embeddings to microscopy descriptors and electrochemical signals—carries the danger of bias propagation and misalignment, which can undermine the reliability of model reasoning. Bias propagation stems from the inherited statistical and conceptual biases within pre-trained multimodal models: for example, literature embeddings may overemphasize historically popular elements or neglect chemistries underexplored in the scientific record. Misalignment arises when different modalities contribute unequally to the shared latent space, such as when imaging artefacts or inconsistent data quality distort the balance between visual and numeric features during dimensionality reduction. While CRESt partially alleviates these issues, these mechanisms remain reactive rather than preventive. The deeper challenge lies in establishing bias-aware learning frameworks and cross-modal calibration standards capable of harmonizing heterogeneous datasets before model training. Addressing these frontiers will be essential to ensure not only faster but also fairer and more interpretable materials discovery. Besides, robotic execution, while already adept at liquid handling and electrochemical screening, still falls short when confronted with gas-phase synthesis, high-temperature processing, or the intricate dynamics of heterogeneous reactors.

The opportunities ahead, however, are unprecedented. The integration of advanced in situ probes—Raman spectroscopy, synchrotron scattering, neutron imaging—into closed-loop experimentation could transform raw measurements into mechanistic insight in real time, collapsing the gap between discovery and understanding. Coupling such systems with generative AI would move beyond incremental optimization toward the proactive invention of chemistries that lie outside human intuition. At the same time, the establishment of open-source databases and interoperable protocols could democratize this capability, enabling geographically distributed laboratories to collaborate through a federated network of AI–robotics platforms. In this model, the collective intelligence of machines and humans would no longer be bounded by local trial-and-error but would operate across a shared landscape of knowledge, accelerating progress while retaining scientific depth. Achieving this requires establishing open and interoperable data standards (e.g., machine-readable schemas that capture not only final results but also all multimodal inputs, process metadata, and VLM-derived experimental annotations) to ensure seamless knowledge transfer. Furthermore, the human absorption of machine-generated knowledge must be enhanced through interpretability layers; AI systems must propose recipes and present the scientific rationale (e.g., "This composition was selected because the text-embedding indicates element *X* tunes the d-band center, and the image-embedding predicts the optimal particle size distribution"), allowing human experts to efficiently verify the mechanistic hypothesis via complementary experiments (e.g., DFT or in situ spectroscopy).

Looking further, one can envision how platforms like CRESt could expand from autonomous laboratories into fully integrated intelligent factories (Figure 1C). Multimodal AI would not only design catalysts but also manage entire process chains—scaling synthesis, adjusting reactor conditions in real time, and linking discovery directly to renewable-powered production. Such factories could autonomously prototype electrocatalysts for water electrolysis, optimize CO_2-to-fuel conversion, and pilot ammonia or green chemical synthesis while continuously sharing data through global networks. In this vision, electrocatalysis becomes part of a planetary-scale discovery engine—a connected ecosystem of adaptive laboratories and smart factories that collectively transform materials science and redefine how knowledge is created.

REFERENCES

1. Yao Y., Dong Q., Brozena A., et al. High-entropy nanoparticles: Synthesis-structure-property relationships and data-driven discovery. *Science* **376**:eabn3103. DOI:10.1126/science.abn3103
2. Szymanski N. J., Rendy B., Fei Y., et al. (2023). An autonomous laboratory for the accelerated synthesis of novel materials. *Nature* **624**:86–91. DOI:10.1038/s41586-023-06734-w
3. Coley C. W., Thomas D. A., Lummiss J. A. M., et al. (2019). A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science* **365**:eaax1566. DOI:10.1126/science.aax1566
4. Zhang Z., Ren Z., Hsu C.-W., et al. (2025). A multimodal robotic platform for multi-element electrocatalyst discovery. *Nature* **647**:390–396. DOI:10.1038/s41586-025-09640-5
5. Boiko D. A., MacKnight R., Kline B., et al. (2023). Autonomous chemical research with large language models. *Nature* **624**:570–578. DOI:10.1038/s41586-023-06792-0

FUNDING AND ACKNOWLEDGMENTS

Y. Yao was supported by National Key R & D Program of China (2021YFA1202300), National Natural Science Foundation of China (No. 52371223). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.