

A multimodal robotic platform for multi-element electrocatalyst discovery

<https://doi.org/10.1038/s41586-025-09640-5>

Received: 3 January 2024

Accepted: 16 September 2025

Published online: 23 September 2025

 Check for updates

Zhen Zhang^{1,9}, Zhichu Ren^{1,9}, Chia-Wei Hsu^{1,9}, Weibin Chen^{2,9}, Zhang-Wei Hong³, Chi-Feng Lee⁴, Aubrey Penn⁵, Hongbin Xu¹, Daniel J. Zheng¹, Shuhan Miao⁶, Yimeng Huang², Yifan Gao², Weiyin Chen², Hugh Smith¹, Yaoshen Niu¹, Yunsheng Tian³, Ying-Rui Lu⁷, Yu-Cheng Shao⁷, Sipei Li², Hsiao-Tsu Wang⁴, Iwnetim I. Abate¹, Pulkit Agrawal¹, Yang Shao-Horn^{1,8} & Ju Li^{1,2}✉

One of the goals of ‘AI for Science’ is to discover customized materials through real-world experiments. Pioneering advances have been made in computational predictions and the automation of materials synthesis^{1–7}. Yet most materials experimentation remains constrained to using unimodal active learning approaches, relying on a single data stream. The potential of artificial intelligence to interpret experimental complexity remains largely untapped^{8,9}. Here we present Copilot for Real-world Experimental Scientists (CRESt), a platform that integrates large multimodal models (incorporating chemical compositions, text embeddings and microstructural images) with knowledge-assisted Bayesian optimization and robotic automation. CRESt uses knowledge-embedding-based search space reduction and adaptive exploration–exploitation strategy to accelerate materials design, high-throughput synthesis and characterization, and electrochemical performance optimization. CRESt enables monitoring with cameras and the generation of vision-language-model-driven hypotheses to diagnose and correct experimental anomalies. Applied to electrochemical formate oxidation, CRESt explored more than 900 catalyst chemistries and 3,500 electrochemical tests within 3 months, identifying a state-of-the-art catalyst in the octonary chemical space (Pd–Pt–Cu–Au–Ir–Ce–Nb–Cr) that exhibits a 9.3-fold improvement in cost-specific performance.

Design of experiments (DOE) for targeted materials discovery remains a fundamental challenge. Even subtle variations in chemistry or processing could yield markedly different properties, and experimental optimization is often constrained by human and machine errors, cost and limited throughput¹⁰. Active learning (AL), in the form of Gaussian process-based Bayesian optimization (GP-BO), provides a DOE framework by balancing exploration of uncertain parametric regions and exploitation of already-promising regions¹¹. However, conventional unimodal AL operates on a single stream of data of fixed dimensions—for example, mapping the ratios of a particular set of quinary elements directly to alloy properties—ignoring the diverse knowledge sources routinely used by human experts, such as literature, previous experience, microstructural features, intuition and hypothesis making.

Here we introduce Copilot for Real-world Experimental Scientists (CRESt), a multimodal platform that integrates large multimodal models (LMMs) with robotic automation¹². Chemical compositions are precisely controlled by robotic preparation, whereas large language models (LLMs) embed literature knowledge to guide design and explorations in reduced-dimensional spaces. Beyond chemical descriptors,

CRESt incorporates microstructural features from high-throughput scanning electron microscopy (SEM) imaging and computer vision analysis, enriching the surrogate model and accelerating optimization. Dimensionality reduction using principal component analysis (PCA) preserves variance before AL is performed in the latent space, with candidate solutions mapped back to compositions by sequential least squares programming (SLSQP).

Real-world experiments also face reproducibility challenges arising from mechanical, electrical, thermal, magnetic and even organizational and cognitive factors that remain invisible to standard algorithms. To address this, CRESt enables voice and text interaction, as well as camera-based monitoring of experiments. Leveraging vision language models (VLMs), CRESt analyses irreproducibility issues and proposes corrective procedures, offering a breadth of expertise that surpasses any individual researcher.

We demonstrate CRESt for the electrochemical formate oxidation reaction. Within 3 months, the platform explored more than 900 chemistries and 3,500 electrochemical tests, leading to the discovery of an octonary multi-element catalyst (Pd–Pt–Cu–Au–Ir–Ce–Nb–Cr). This catalyst achieved a 9.3-fold improvement in cost-specific performance

¹Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. ²Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. ³Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA. ⁴Department of Physics, Tamkang University, New Taipei City, Taiwan. ⁵MIT.nano, Massachusetts Institute of Technology, Cambridge, MA, USA. ⁶Department of Architecture, Massachusetts Institute of Technology, Cambridge, MA, USA. ⁷National Synchrotron Radiation Research Center, Hsinchu, Taiwan. ⁸Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. ⁹These authors contributed equally: Zhen Zhang, Zhichu Ren, Chia-Wei Hsu, Weibin Chen. ✉e-mail: liju@mit.edu

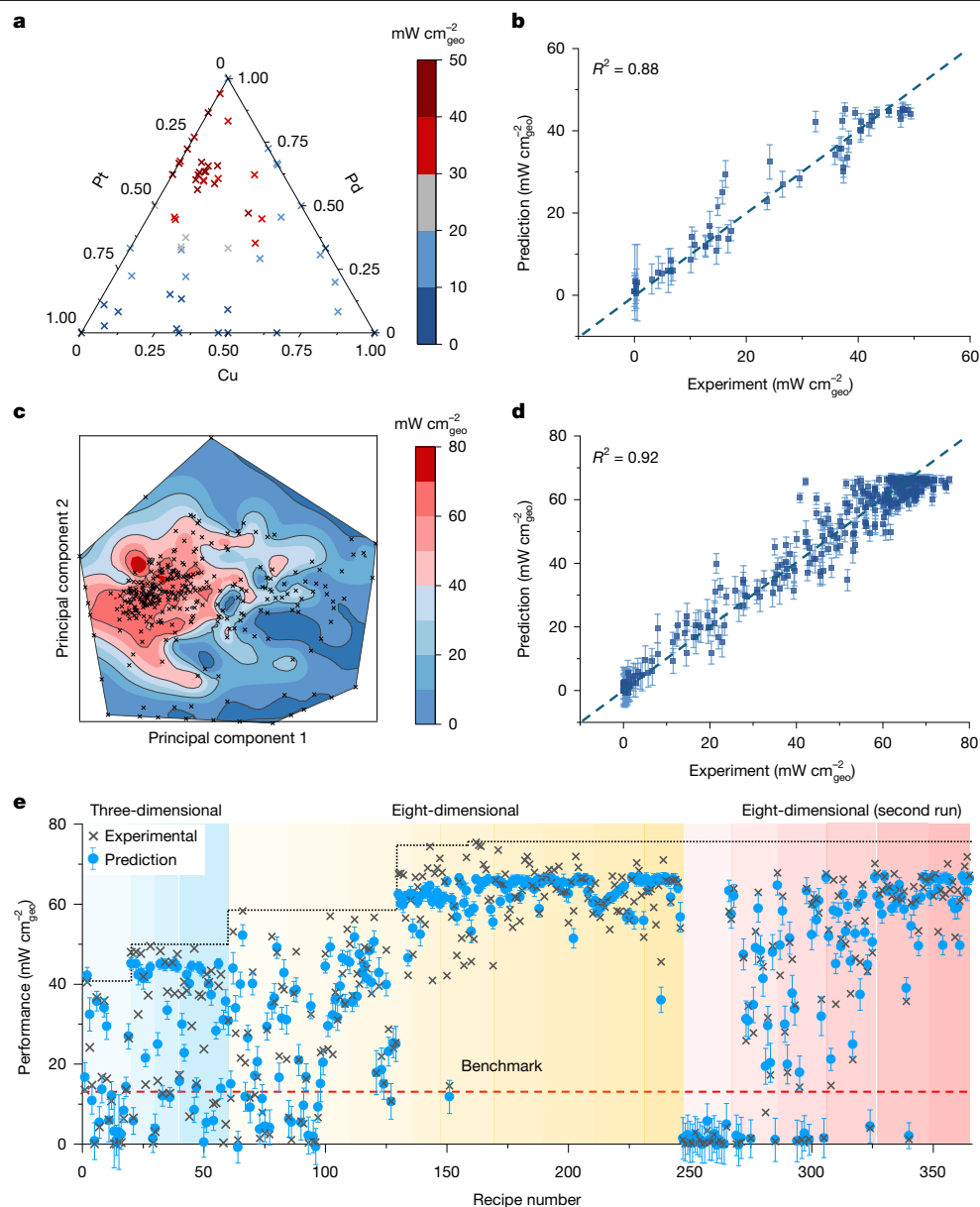


Fig. 1 | AL discovery of multi-element electrocatalysts for formate oxidation reactions. **a**, The trajectory plot of the AL optimization results for ternary electrocatalysts (Pd–Pt–Cu). The triangular axes indicate the composition ratios of the three metallic elements. **b**, Cross-validation plot for the ternary AL, with prediction versus experiments. **c**, PCA plot with contour lines for octonary alloys (Pd–Pt–Cu–Au–Ir–Ce–Nb–Cr). **d**, Cross-validation plot for the octonary AL,

with prediction versus experiments. **e**, The complete AL campaign plot with the baseline AL model, with the performance value compared with the recipe number. Blue shading indicates the ternary optimization region. The yellow and red shadings indicate the octonary optimization. The colour bars in **a** and **c** indicate the maximum power density ($\text{mW cm}^{-2}_{\text{geo}}$) of a catalyst. Error bars in **b**, **d**, **e** represent the predictive uncertainty estimated by the GP-BO model.

(power density normalized by catalyst cost) over the pure Pd benchmark (Fig. 1 and Supplementary Table 1). Mechanistic investigations showed enhanced tolerance to CO and hydrogen poisoning, consistent with electronic tuning in rational catalyst design. Finally, in a direct formate fuel cell, the optimized recipe delivered record power density at just one-quarter the precious-metal loading of previous devices. Together, these results demonstrate the potential of LMM- and robotics-driven materials experimentation.

Copilot for Real-world Experimental Scientists

We introduce our LMM-assisted robotic platform named CRES¹² (Extended Data Figs. 1 and 2 and Supplementary Videos 1 and 2). CRES consists of three main components: a user interface; the LMM-powered back end; and a range of actuators. The user interface supports both

voice and text interactions, connecting the LMM back end and allowing researchers to control the robotic platform without computer programming experience. The actuators encompass various robotic equipment, including sample preparation tools, such as the liquid-handling robot and the carbothermal shock system¹³; sample testing tools, such as an automated electrochemical workstation; characterization equipment, including X-ray diffraction (XRD) and software-driven SEM; and auxiliary devices, such as automated peristaltic pumps and gas valves. To facilitate coordinated electrochemical research, we customized these actuators extensively through coding, three-dimensional (3D) printing, laser cutting, electrical modifications, and so on. All actuators operate remotely through our customized Python codes. Apart from implementing instructions, CRES actively converses with human scientists in natural languages, making observations and hypotheses¹⁴ along the way.

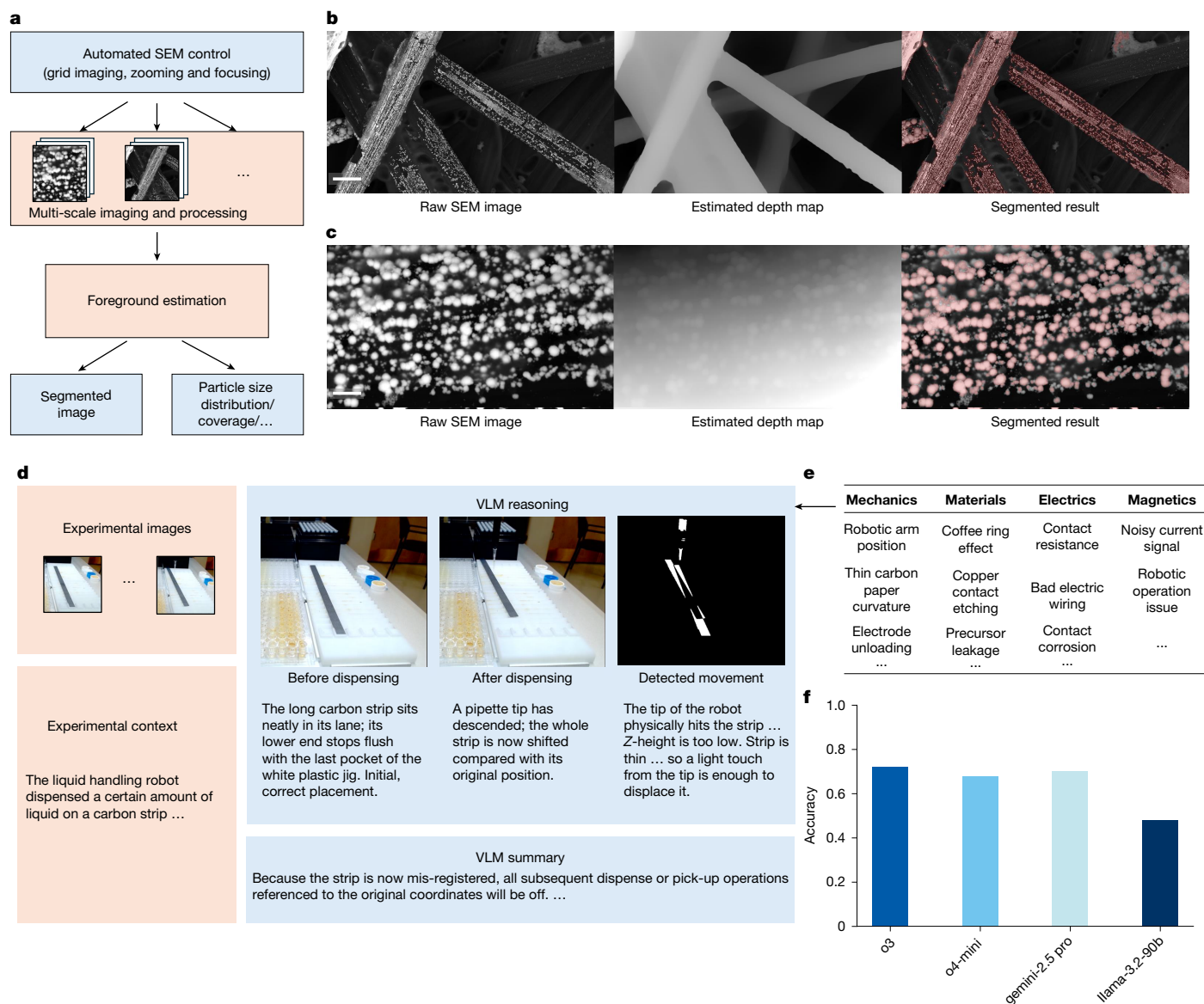


Fig. 2 | Vision analysis modules of the CREST platform. a–c, Computer vision analysis of microstructures by the automated SEM control. Pipeline of the multi-scale image feature acquisition and analysis (**a**) and visualization of microstructure analysis (**b,c**). The first column shows the original SEM images. The second column shows the predicted depth maps, highlighting the separation of foreground and background. The third column shows particles successfully identified by our algorithm, marked in red. The recipe shown is for $\text{Pd}_{0.232}\text{Pt}_{0.133}\text{Ir}_{0.330}\text{Nb}_{0.176}\text{Cr}_{0.128}$. Scale bar, 8 μm (**b**); 1 μm (**c**). **d–f,** Experimental

diagnosis by the VLM. An illustrative example of leveraging a VLM is shown in **d** to diagnose issues during the automated liquid pipetting process. Provided with a minimal experimental context by the human operator, the VLM performed reasoning, invoked tools such as movement detection to identify shifts in the carbon paper, and subsequently proposed relevant solutions. Diverse sources of experimental failure in the CREST pipeline are shown in **e**. Overall success rate of debugging experimental issues with different VLMs are shown in **f**.

Knowledge-assisted AL

Conventional unimodal AL relies on a single data stream of small, fixed dimensions, overlooking the richness and complexity of real-world materials and experiments. Embedding methods have proven effective across a range of downstream tasks in integrating diverse information sources^{15,16}, such as 3D molecular and crystal structures¹⁷, property databases¹⁸, lab logs and discussion notes, into unified vector representations. Here, modifying the standard Bayesian optimization method (Supplementary Note 1), we introduce a knowledge-assisted Bayesian optimization (KABO) method that leverages LMMs to bridge this gap (Supplementary Note 2). For the text-embedding-based dimensionality reduction, CREST first searches the literature for comprehensive descriptions of metallic elements relevant to formate oxidation

reactions and then embeds these responses into alloy-specific vectors using word embedding models¹⁹. For the image embedding, we developed a new workflow to perform high-throughput SEM imaging (Fig. 2a–c, Extended Data Fig. 2b and Supplementary Fig. 1) for our samples with different imaging resolutions (Extended Data Figs. 2b and 3). Taking advantage of the near-spherical morphology of our catalyst particles in most cases, we elected to extract four morphological features (slope of the particle distribution function, bias of the particle distribution function, particle coverage ratio and the particle count per unit area) for the AL training (Supplementary Note 4). Next, KABO applies PCA to update the latent space, reducing dimensionality while explaining at least 80% of the variability in the data, thus effectively representing both literature-derived knowledge and known experimental features. Using the knowledge gradient acquisition function²⁰, KABO

optimizes in this latent space, mapping candidate solutions back to practical materials compositions by SLSQP.

We also address the challenge of adaptively balancing exploration and exploitation in Bayesian optimization, which conventionally depends on a manually tuned weight parameter (κ) in the acquisition function. Here, we incorporate a dynamic tuning mechanism inspired by the Bayesian optimization with policy improvement constraints (BOPIC) framework in reinforcement learning²¹, introducing a Lagrange multiplier (λ) that adaptively adjusts the trade-off between exploration and exploitation without manual intervention (Supplementary Note 3).

We used CRES to discover catalysts for the electrochemical oxidation of formate. We elected to optimize the maximum power density of a fuel cell device using the synthesized catalyst for formate oxidation, with oxygen reduction reaction at the counter electrode. To simplify, we assumed that the current and potential data from the three-electrode linear sweep voltammetry reflect the performance of the formate oxidation half-cell and a constant overpotential for oxygen reduction reaction²² (Supplementary Note 1). These choices provided a relatively consistent ranking of the catalyst performance when compared with fuel cell experiments (Supplementary Figs. 2 and 3). We independently evaluated the performance of KABO, BOPIC and standard Bayesian optimization through experimentation (Extended Data Fig. 4). Notably, both KABO and BOPIC achieved continuous improvements by identifying higher-performing recipes batch after batch, whereas standard Bayesian optimization required several batches before substantial gains were obtained. This delay could be detrimental in high-throughput settings, in which limited optimization may cause premature termination of experiments. By pre-screening the design space, KABO provided substantial early-stage benefits and reached convergence first. BOPIC proved most effective in sampling high-performing recipes, thereby generating a richer candidate pool for downstream selection. We further benchmarked the convergence efficiency of the three algorithms using a trained gradient boosting model to simulate the optimization process (Supplementary Fig. 4). It is shown that, on average, there is a 36% (KABO) and 27% (BOPIC) performance improvement compared with Bayesian optimization under the same number of experimental trials. In the quest for high-performance samples (here, more than five times the performance of the benchmark), it only required 25% (KABO) and 47% (BOPIC) of the recipes. Notably, KABO algorithm shows equally good performance when using open-source LLMs under various conditions, demonstrating the generalizability of our method (Supplementary Note 2 and Supplementary Fig. 5). For the image-embedding method, to further validate the use of the selected morphological features for AL, we designed various experimental settings and ablation studies. We found that microstructural features, when optimized jointly with elemental ratios, led to faster convergence (Supplementary Fig. 6).

Owing to the complexity of real-world experiments with Bayesian optimization, we first attempted an optimization process in a ternary chemical space (Pd–Pt–Cu), a subset of the octonary chemical space (Pd–Pt–Cu–Au–Ir–Ce–Nb–Cr). As a result, in the ternary space, we identified an optimal catalyst composition with less than 60 experimental recipes, from around 5×10^5 potential recipes (Fig. 1a and Supplementary Note 5), with cross-validation confirming model accuracy against experiments (Fig. 1b). The optimized catalyst, $\text{Pd}_{0.635}\text{Pt}_{0.258}\text{Cu}_{0.107}$, delivered around 3.5 times the power density of pure Pd with the same molarity of metallic loading, which is commonly used as the benchmark catalyst²³. We then progressed to the octonary space, which introduces exponentially larger search space (around 2×10^{17} potential recipes). We achieved optimized performance, with PCA indicating a primary peak with smaller satellites (Fig. 1c; for further analysis of the physical importance of the two primary axes from PCA, see Supplementary Fig. 7), and cross-validation showing a good alignment with experimental data (Fig. 1d). To validate global optimization, we started a second trial of AL, but favoring exploration. The optimization again converged to a similar value, potentially confirming global optimization (Fig. 1e).

The optimized catalyst, $\text{Pd}_{0.487}\text{Pt}_{0.185}\text{Cu}_{0.018}\text{Ir}_{0.037}\text{Ce}_{0.106}\text{Nb}_{0.168}$, showed 5.6 times the power density, whereas the catalyst $\text{Pd}_{0.381}\text{Pt}_{0.080}\text{Cu}_{0.009}\text{Au}_{0.004}\text{Ir}_{0.02}\text{Ce}_{0.086}\text{Nb}_{0.338}\text{Cr}_{0.082}$ showed 9.3 times the power density normalized by catalyst cost, compared with the pure Pd catalyst benchmark (Supplementary Table 1).

Experimental error diagnosis by VLM

At the early stages of our electrochemical experiments, poor reproducibility emerged as the main obstacle and time sink. Although large datasets were generated, inconsistencies across trials rendered them unreliable for AL training (Supplementary Fig. 8a and Supplementary Data 1 and 2). Careful inspection showed numerous hidden errors—mechanical, thermal, electrical, magnetic and even organizational and cognitive—that subtly altered conditions despite seemingly identical steps, initially requiring extensive human oversight to debug (Fig. 2d,e). The advent of VLMs provided a new approach: by coupling computer vision with broad domain knowledge, AI proposed sources of irreproducibility and plausible corrective procedures (Supplementary Video 1). For example, the VLM can invoke tools such as coordinate alignment and movement detection to diagnose issues in which the pipette displaces the carbon paper, even from micrometre-scale height changes (Fig. 2d). In another case, a 1-mm deviation in a 1-cm sample geometry introduces an error of approximately 10%, substantially degrading AL performance. Although laser-cut wooden stages were designed to ensure uniformity, the VLM identified charring artefacts that caused dimensional variation, attributing them to surface non-uniformity in laser absorption (Supplementary Fig. 9). This insight prompted a switch to stainless steel stages, yielding improved consistency, with minimal human guidance during the debugging process. To systematically evaluate this ability, we curated a troubleshooting log ('Criminals in the AI City') and constructed a question and answer set from recorded failures (Appendix in the Supplementary Information). Benchmarking several VLMs demonstrated promising accuracies—72% (OpenAI o3), 70% (Gemini-2.5 Pro), 68% (OpenAI o4-mini) and 48% (Llama-3.2-90B)—highlighting their potential as practical experimental assistants (Fig. 2f and Extended Data Fig. 5). After hypothesis making¹⁴, checking and correcting the root causes, reproducibility was markedly improved (Supplementary Fig. 8b).

Performance testing

We leveraged in situ electrochemical deposition for the purpose of high-throughput testing. The synthesized nanoparticles showed uniform distribution across the entire electrode surface and on single carbon fibres (Fig. 3a). The optimized catalyst demonstrated a homogeneous mix of elements (Fig. 3b) and a single face-centred cubic (FCC) phase in the XRD analysis (Fig. 4a and Supplementary Fig. 10a,b). In the cyclic voltammetry analysis for formate oxidation, a pronounced peak appears during the forward scan, followed by a marked decrease in the current density as oxidation continues (Fig. 3c). This decrease is due to the formation of metal oxides, such as PdO_x , which are inactive for formate oxidation^{24,25}. A pronounced peak is observed in the reverse scan, which corresponds to a surge in formate oxidation activity when metal oxides are reduced back to their metallic form²⁶. In three-electrode testing, the optimized catalysts demonstrated a much higher current density, despite having an equivalent molar metallic catalyst loading, compared with the benchmark catalyst. The enhanced activity allowed for a reduction in the usage of precious metals. In the fuel cell tests, our catalyst (with 2.0 mg cm^{-2} of entire metal loading, and thus around 1.2 mg cm^{-2} of precious-metal loading) demonstrated a peak power density of 325 mW cm^{-2} , higher than the benchmark Pd catalyst with various loadings ($0.5\text{--}4.0 \text{ mg cm}^{-2}$), either tested by us (Fig. 3d and Supplementary Fig. 11) or reported in the literature^{23,27–35} (Fig. 3e). Furthermore, we found that the difference of catalyst activity could

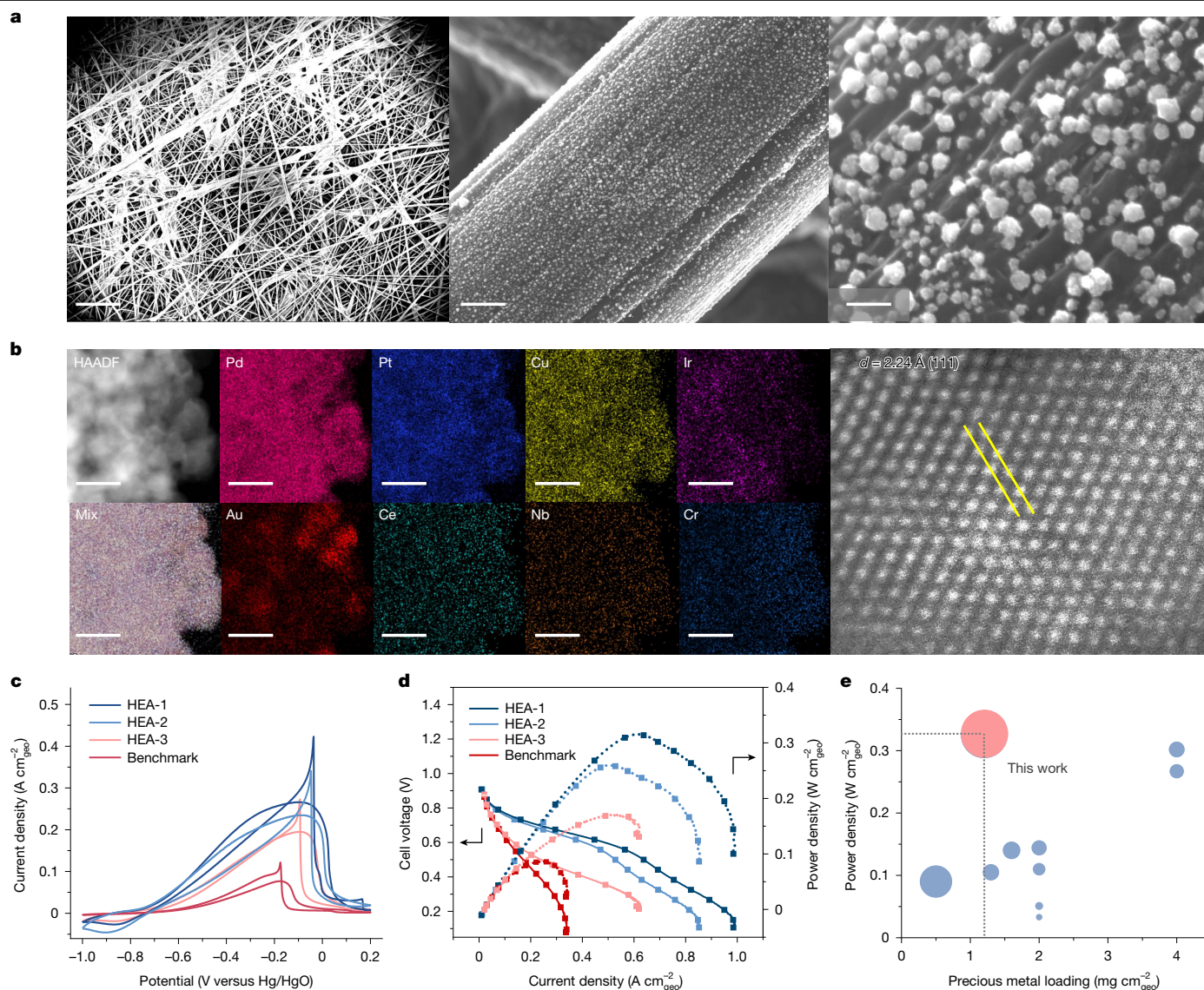


Fig. 3 | Morphology characterization and electrochemical testing of the optimized catalysts for the formate oxidation reaction. **a**, SEM images of the electrochemically prepared catalysts on the carbon substrate. Uniform catalyst distribution was observed on a single carbon fibre. Scale bar, 200 μm , 1 μm and 40 nm (left to right). **b**, High-angle annular dark-field scanning transmission electron microscopy and EDS of the optimized catalysts. Scale bar, 10 nm. **c**, The cyclic voltammetry plot of selected chemistries in a three-electrode setup under nitrogen sparging at ambient conditions. A low catalyst precursor solution loading of $10 \mu\text{l cm}^{-2}$ (corresponding to about 0.05 mg cm^{-2} for pure Pd) was used for fast catalyst screening. The scan rate was 50 mV s^{-1} .

at least partly come from the distinct chemical nature of these catalysts, as indicated by a consistent trend of voltage/current slope in the activation region in both the three-electrode and fuel cell testing (Supplementary Fig. 12).

Mechanistic analysis

For the octonary catalyst (HEA-8D), Pd and Pt serve as primary active elements for formate oxidation reactions²³. Small amounts of Au and Ir, included in the optimized composition, have also been previously used in direct formate fuel cells²⁶. The remaining elements (Cu, Ce, Nb and Cr) could alloy with the primary elements to provide tailored coordination environments and reduce catalyst cost^{13,36,37}. The optimized catalyst,

d, The polarization curve of selected chemistries tested in the direct formate fuel cells at 60°C in an electrolyte of 2.0 M KOH and 1.0 M HCOOK. To enable comparison with literature, all samples had a total catalyst loading of 2.0 mg cm^{-2} . Details are in the Methods. **e**, Comparison plot of our optimized catalysts with results in the literature under similar testing conditions (typically at 60°C in an electrolyte of 2.0 M KOH and 1.0 M HCOOK)^{23,27–35}. The catalyst HEA-1, HEA-2, HEA-3 and Benchmark represent the composition of $\text{Pd}_{0.487}\text{Pt}_{0.185}\text{Cu}_{0.018}\text{Ir}_{0.037}\text{Ce}_{0.106}\text{Nb}_{0.168}$, $\text{Pd}_{0.381}\text{Pt}_{0.080}\text{Cu}_{0.009}\text{Au}_{0.004}\text{Ir}_{0.02}\text{Ce}_{0.086}\text{Nb}_{0.338}\text{Cr}_{0.082}$, $\text{Pd}_{0.635}\text{Pt}_{0.258}\text{Cu}_{0.107}$ and Pd, respectively. HEA, high-entropy alloy.

$\text{Pd}_{0.381}\text{Pt}_{0.080}\text{Cu}_{0.009}\text{Au}_{0.004}\text{Ir}_{0.02}\text{Ce}_{0.086}\text{Nb}_{0.338}\text{Cr}_{0.082}$, demonstrated a single FCC phase in the XRD analysis with Rietveld refinement (Fig. 4a). The lattice parameters for Pd and HEA-8D were determined as 3.896 Å and 3.899 Å, respectively, indicating that here alloying does not lead to huge lattice deformation, potentially ensuring structural integrity (Supplementary Table 2).

To probe the oxidation states and local coordination environments of the optimized catalyst, in situ XAS was used. Owing to the complexity of the composition, we elected to investigate the primary catalytic elements Pd and Pt. X-ray absorption near-edge structure analysis showed that both Pd and Pt retained their metallic states during formate oxidation (Supplementary Fig. 13), an important factor given that both PdO and PtO₂ exhibit negligible catalytic activity²⁶. A slight shift

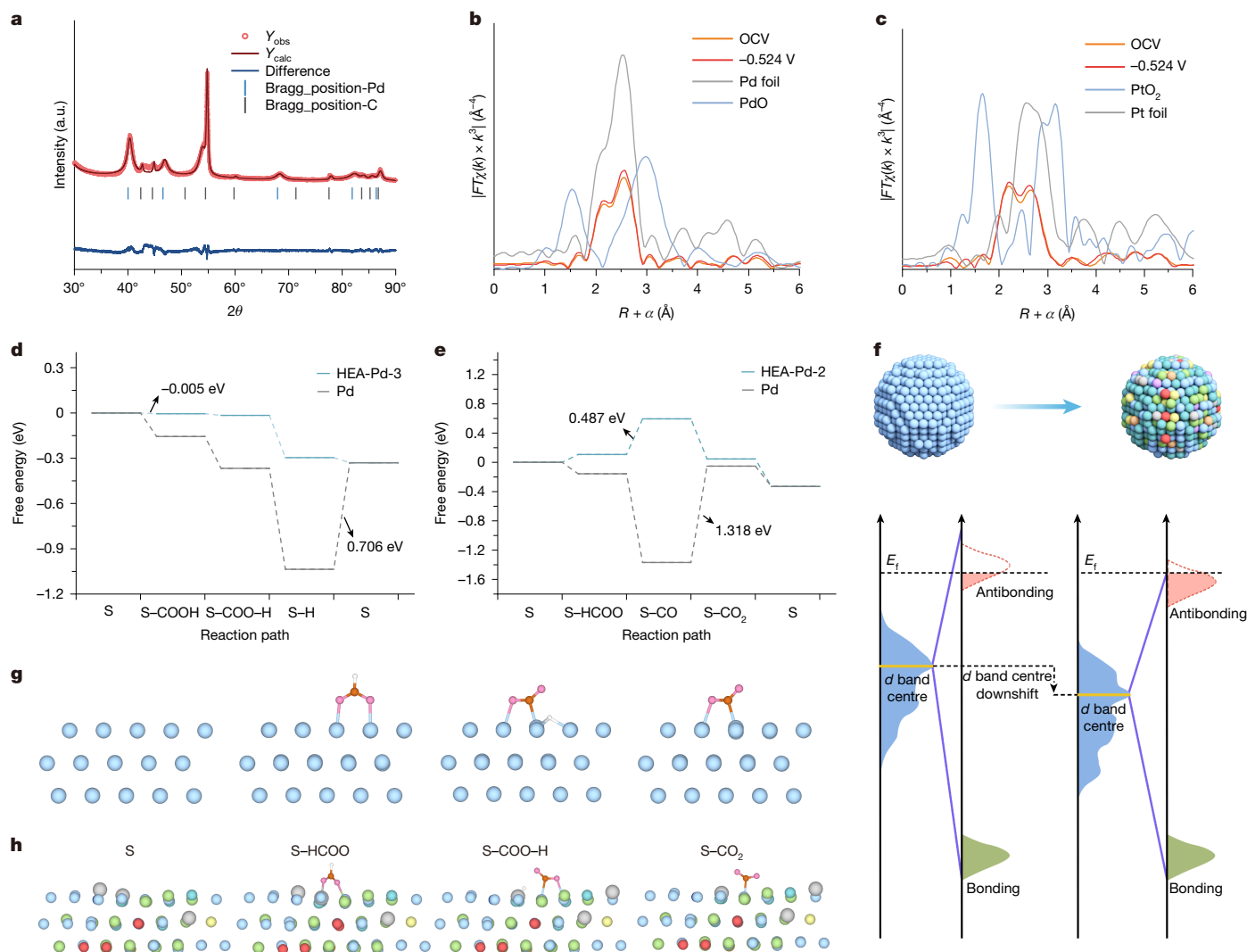


Fig. 4 | Mechanistic analysis of the optimized multi-element electrocatalysts.

a, XRD patterns with Rietveld refinement for the octonary catalyst, showing a singular FCC solid solution phase. **b,c**, In situ Fourier transform EXAFS spectra of the octonary catalyst collected at the Pd edge: the open circuit voltage spectrum, the potential-applied spectrum (at -0.524 V compared with Hg/HgO in 1.0 M KOH), as compared with reference samples, including metallic Pd and PdO (**b**) and spectra collected at the Pt edge: the open circuit voltage spectrum, the potential-applied spectrum (at -0.524 V compared with Hg/HgO in 1.0 M KOH), as compared with reference samples, including metallic Pt and PtO₂ (**c**). **d,e**, Reaction pathways computed by the DFT calculations. Benchmark Pd catalysts show a

reaction barrier of 0.706 eV in the direct pathway (**d**) and 1.318 eV in the indirect pathway (**e**), whereas the high-entropy catalysts showed a reduced activation barrier of -0.005 eV in the direct pathway (**d**) and 0.487 eV in the indirect pathway (**e**). **f**, Schematic of the standard Pd particle and the octonary alloy particle, and the position of the d -band centre. **g,h**, Schematic of the model and key intermediates and adsorbates on the Pd (**g**) and the optimized octonary catalyst (**h**). From top to bottom: the pure Pd reaction model and the HEA-Pd-3 reaction model. From left to right: the clean model surface, the model with HCOO adsorbed, the model with COO-H adsorbed and the model with CO₂ adsorbed. OCV, open circuit voltage.

in the Pd absorption edge relative to Pd foil indicated local structural perturbations arising from alloying. Furthermore, extended X-ray absorption fine structure (EXAFS) spectra showed distinct Pd-Pd and Pd-Pt coordination peaks, indicative of strong electronic interactions associated with tuned catalytic activity (Fig. 4b,c). Quantitative fitting of the EXAFS data showed minor variations in bond lengths and coordination numbers, suggesting a robust atomic structure in HEA-8D during the operation (Supplementary Note 6, Supplementary Figs. 14 and 15 and Supplementary Table 3).

A structural model was constructed based on refined XRD results, and high-throughput calculations were performed to screen and identify stable structures for subsequent simulations (Supplementary Fig. 16 and Supplementary Note 7). Reaction pathways and free energy profiles for both direct and indirect pathways were calculated for multiple possible adsorption sites (Supplementary Figs. 17–27 and Supplementary Notes 8 and 9) and benchmarked against a conventional Pd catalyst.

The octonary catalyst demonstrates strong resistance to hydrogen poisoning (direct pathway, Fig. 4d,g,h) and CO poisoning (indirect pathway, Fig. 4e) relative to pure Pd. Notably, the Pd site in the HEA exhibits a substantially lower potential-determining step barrier for the indirect pathway than that of pure Pd. To further explain the mechanism, projected density of states analyses were performed for the Pd sites in HEA-8D and pure Pd in DFT calculations (Supplementary Fig. 28). The d -band centres of Pd-1, Pd-2 and Pd-3 are all much lower than that of pure Pd, indicating weakened hydrogen binding due to reduced orbital overlap (Fig. 4f), which promotes desorption. Moreover, isotope study (with electrolytes containing HCOONa or DCOONa) and CO stripping experiments were performed to experimentally validate the tolerance of the catalyst to surface H_{ads} and CO_{ads}, respectively (Supplementary Figs. 29 and 30). These findings demonstrate that tailored alloying and atomic-level structural tuning effectively regulate surface electronic properties and reaction energetics.

Conclusion

In summary, we developed CReSt, an LMM- and robotic-driven platform for experimental materials discovery. Our method incorporates previous literature and database knowledge, human and machine hypothesis making, composition tuning and process design, and microstructural features into the materials experimentation framework. We further leverage VLMs to analyse experimental processes to find and correct the root causes of anomalies. Mechanistic studies confirm that the optimized multi-element catalyst exhibits enhanced tolerance to hydrogen and CO poisoning, attributed to tailored atomic and electronic structures. This demonstrates that LMM-based approach could more effectively explore the rich and complex real-world materials and experiments, uniting automation with intelligence.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-025-09640-5>.

1. Szymanski, N. J. et al. An autonomous laboratory for the accelerated synthesis of novel materials. *Nature* **624**, 86–91 (2023).
2. Dai, T. et al. Autonomous mobile robots for exploratory synthetic chemistry. *Nature* **635**, 890–897 (2024).
3. Joung, J. F. et al. Electron flow matching for generative reaction mechanism prediction. *Nature* **645**, 115–123 (2025).
4. Koscher, B. A. et al. Autonomous, multiproperty-driven molecular discovery: From predictions to measurements and back. *Science* **382**, eadi1407 (2023).
5. Boiko, D. A., MacKnight, R., Kline, B. & Gomes, G. Autonomous chemical research with large language models. *Nature* **624**, 570–578 (2023).
6. Coley, C. W. et al. A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science* **365**, eaax1566 (2019).
7. Tom, G. et al. Self-driving laboratories for chemistry and materials science. *Chem. Rev.* **124**, 9633–9732 (2024).
8. Ren, Z., Ren, Z., Zhang, Z., Buonassisi, T. & Li, J. Autonomous experiments using active learning and AI. *Nat. Rev. Mater.* **8**, 563–564 (2023).
9. Noorden, R. V. & Perkel, J. M. AI and science: what 1,600 researchers think. *Nature* **621**, 672–675 (2023).
10. Attia, P. M. et al. Closed-loop optimization of fast-charging protocols for batteries with machine learning. *Nature* **578**, 397–402 (2020).
11. Balandat, M. et al. BoTorch: a framework for efficient Monte-Carlo Bayesian optimization. *NeurIPS* <https://doi.org/10.48550/arxiv.1910.06403> (2020).
12. Ren, Z., Zhang, Z., Tian, Y. & Li, J. CReSt – copilot for real-world experimental scientist. Preprint at <https://doi.org/10.26434/chemrxiv-2023-tnz1x-v4> (2023).
13. Yao, Y. et al. Carbothermal shock synthesis of high-entropy-alloy nanoparticles. *Science* **359**, 1489–1494 (2018).
14. Park, Y. J. et al. Can ChatGPT be used to generate scientific hypotheses? *J. Materiom.* **10**, 578–584 (2024).
15. Nussbaum, Z., Morris, J. X., Duderstadt, B. & Mulyar, A. nomic embed: training a reproducible long context text embedder. Preprint at <https://doi.org/10.48550/arxiv.2402.01613> (2024).
16. Ramos, M. C., Michtav, S. S., Porosoff, M. D. & White, A. D. Bayesian optimization of catalysis with in-context learning. Preprint at <https://doi.org/10.48550/arxiv.2304.05341> (2025).
17. Park, Y. J., Kumaran, M., Hsu, C.-W., Olivetti, E. & Li, J. Contrastive learning of English language and crystal graphs for multimodal representation of materials knowledge. Preprint at <https://doi.org/10.48550/arxiv.2502.16451> (2025).
18. Park, Y. J., Jerng, S. E., Yoon, S. & Li, J. 1.5 million materials narratives generated by chatbots. *Sci. Data* **11**, 1060 (2024).
19. Neelakantan, A. et al. Text and code embeddings by contrastive pre-training. Preprint at <https://doi.org/10.48550/arxiv.2201.10005> (2022).
20. Frazier, P. I., Powell, W. B. & Dayanik, S. A knowledge-gradient policy for sequential information collection. *SIAM J. Control Optim.* **47**, 2410–2439 (2008).
21. Chen, E., Hong, Z.-W., Pajarinen, J. & Agrawal, P. Redeeming intrinsic rewards via constrained optimization. *NeurIPS* <https://doi.org/10.48550/arxiv.2211.07627> (2022).
22. Xu, H. et al. An actor–critic algorithm to maximize the power delivered from direct methanol fuel cells. *Nat. Energy* **10**, 951–961 (2025).
23. An, L. & Chen, R. Direct formate fuel cells: a review. *J. Power Sources* **320**, 127–139 (2016).
24. Choun, M., Hong, S. & Lee, J. Adsorbed hydrogen as a site-occupying species in the electrocatalytic oxidation of formate on Pd/C in alkaline medium. *J. Electrochem. Soc.* **165**, J3266–J3270 (2018).
25. Wang, Q. et al. Transition from core-shell to janus segregation pattern in AgPd nanoalloy by Ni doping for the formate oxidation. *Appl. Catal. B Environ.* **270**, 118861 (2020).
26. Folkman, S. J., González-Cobos, J., Giancola, S., Sánchez-Molina, I. & Galán-Mascarós, J. R. Benchmarking catalysts for formic acid/formate electrooxidation. *Molecules* **26**, 4756 (2021).
27. Tran, K., Nguyen, T. Q., Bartrom, A. M., Sadiki, A. & Haan, J. L. A fuel-flexible alkaline direct liquid fuel cell. *Fuel Cells* **14**, 834–841 (2014).
28. Bartrom, A. M. & Haan, J. L. The direct formate fuel cell with an alkaline anion exchange membrane. *J. Power Sources* **214**, 68–74 (2012).
29. Bartrom, A. M. et al. Optimization of an anode fabrication method for the alkaline direct formate fuel cell. *J. Power Sources* **229**, 234–238 (2013).
30. Lan, L. et al. Membrane-less direct formate fuel cell using an Fe–N-doped bamboo internode as the binder-free and monolithic air-breathing cathode. *ACS Appl. Mater. Inter.* **12**, 27095–27103 (2020).
31. Bartrom, A. M., Ognibene, G., Ta, J., Tran, J. & Haan, J. L. Catalysts for alkaline direct ethanol and direct formate fuel cells. *Electrochem. Soc. Trans.* **50**, 1913–1918 (2013).
32. Zeng, L., Tang, Z. K. & Zhao, T. S. A high-performance alkaline exchange membrane direct formate fuel cell. *Appl. Energy* **115**, 405–410 (2014).
33. Li, Y., Feng, Y., Sun, X. & He, Y. A sodium-ion-conducting direct formate fuel cell: generating electricity and producing base. *Angew. Chem. Int. Ed.* **56**, 5734–5737 (2017).
34. Castello, C. et al. Direct formate anion exchange membrane fuel cells with a PdAu bimetallic nanoparticle anode electrocatalyst obtained by metal vapor synthesis. *Energy Adv.* **3**, 2520–2529 (2024).
35. Sun, X., Li, Y., An, L. & Lv, X. Comparative performance evaluation of self-basifying direct formate fuel cells. *J. Electrochem. Soc.* **166**, F768–F773 (2019).
36. Abdelhafiz, A., Wang, B., Harutyunyan, A. R. & Li, J. Carbothermal shock synthesis of high entropy oxide catalysts: dynamic structural and chemical reconstruction boosting the catalytic activity and stability toward oxygen evolution reaction. *Adv. Energy Mater.* **12**, 2200742 (2022).
37. Tajuddin, A. A. H. et al. Corrosion-resistant and high-entropic non-noble-metal electrodes for oxygen evolution in acidic media. *Adv. Mater.* **35**, e2207466 (2023).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2025

Methods

Implementation of the CRES platform

The system supports a unified workflow in which the user engages with our AI platform, CRES, a large language-model-driven layer, through natural language prompts (Extended Data Fig. 2a). CRES interprets these requests and orchestrates the execution of relevant routines, here collectively referred to as 'end-effectors'. Each end-effector is accessed through standardized HTTP requests or Python calls. Some routines query local and remote databases (for example, Materials Project) for data retrieval, whereas others directly manipulate physical laboratory hardware, such as a liquid-handling robot (Opentrons), multi-axis robotic arm (xArm) or specialized lab components for pumping, gas flow control or illumination.

In the present implementation, many devices rely on PyAutoGUI for emulating keyboard and mouse interactions, enabling a universal approach to instrument control irrespective of the availability of hardware-specific application programming interfaces. This is expected to evolve as laboratory equipment increasingly offers dedicated endpoints compatible with AI-ready protocols. Although not limited to these, commonly invoked routines involve automatic experimental preparation (for example, pipetting liquids with Opentrons), advanced sample handling (xArm) or environment manipulation (pump and gas valves). Thus, the integration of CRES with both digital data repositories and physical lab instrumentation consolidates a broad range of abilities within a single AI-governed function pool, reducing operator overhead while ensuring reproducibility and streamlined experimental workflows. Specific AI approaches are discussed in detail in Supplementary Notes 1–3.

The autonomous materials characterization framework integrates three GPT (generative pre-trained transformer)-based agents with a Phenom Pharos scanning electron microscope using Python (Extended Data Fig. 2b). At the top layer, CRES serves as the user interface, capturing natural language descriptions of desired imaging objectives. Rather than issuing low-level instructions, CRES converts these requests into structured goals and forwards them to the SEM actuation agent, which is also powered by GPT. The scanning electron microscope actuation agent then interprets the goals and translates them into direct scanning electron microscope commands (for example, zoom in, zoom out and stage navigation). Autofocus, brightness-contrast and image acquisition routines proceed automatically at each iteration. Following each round of data capture, the scanning electron microscope actuation agent sends the newly obtained micrographs and metadata to the vision agent, built on GPT-4V. The vision agent assesses whether the micrographs fulfil the predefined objective and, if necessary, recommends subsequent zoom or stage manoeuvres back to the scanning electron microscope actuation agent. A set-of-mark layer augments the raw images with reference markers before they are processed by GPT-4V, allowing explicit spatial referencing of specific features and ensuring more precise stage navigation. This cycle repeats until the vision agent confirms that the target electron micrograph has been successfully acquired, at which point the scanning electron microscope actuation agent returns the final image and summary report to CRES. Detailed methods for the computer vision analysis on catalyst particles are provided in Supplementary Note 4.

Materials and reagents

Palladium(II) chloride ($\geq 99.9\%$), chloroplatinic acid hexahydrate (ACS reagent, $\geq 37.50\%$ Pt basis), platinum(IV) chloride ($\geq 99.99\%$ trace metals basis), copper(II) nitrate trihydrate (puriss. p.a., 99–104%), gold(III) chloride ($\geq 99.99\%$ trace metals basis), gold(III) chloride trihydrate ($\geq 99.9\%$ trace metals basis), hydrogen hexachloroiridate(IV) hydrate (99.9% trace metals basis), cerium(III) nitrate hexahydrate (99% trace metals basis), niobium(V) chloride ($\geq 99.9\%$ trace metals basis), chromium(III) chloride (anhydrous, 99.99% trace metals

basis), bismuth(III) chloride (99.99% trace metals basis), tin(II) chloride ($\geq 99.99\%$ trace metals basis), yttrium(III) chloride (anhydrous, powder, 99.99% trace metals basis), iron(III) chloride (anhydrous for synthesis), zinc(II) chloride (reagent grade, $\geq 98\%$), indium(III) chloride (99.999% trace metals basis), praseodymium(III) chloride (anhydrous, powder, 99.99% trace metals basis), lanthanum(III) chloride heptahydrate (ACS reagent), ruthenium(III) chloride hydrate ($\geq 99.9\%$ trace metals basis), vanadium(III) chloride (99+), Tungsten(VI) chloride ($\geq 99.9\%$ trace metals basis), nickel(II) chloride hexahydrate (99.9% trace metals basis), cobalt(II) chloride hexahydrate (ACS reagent, 98%), ethyl alcohol (pure, 200 proof, anhydrous, $\geq 99.5\%$), hydrochloric acid (ACS reagent, 37%), Nafion 115 ionomer solution (5 wt%), potassium formate (ReagentPlus, 99%), potassium hydroxide (ACS reagent, $\geq 85\%$, pellets), oleylamine (technical grade, 70%), hexadecyltrimethylammonium chloride (CTAC, $\geq 98.0\%$), palladium(II) acetylacetonate ($\text{Pd}(\text{acac})_2$, Umicore, 99%), platinum(II) acetylacetonate ($\text{Pt}(\text{acac})_2$, $\geq 99.98\%$ trace metals basis), copper(II) acetylacetonate ($\text{Cu}(\text{acac})_2$, $\geq 99.9\%$ trace metals basis), iridium(III) acetylacetonate ($\text{Ir}(\text{acac})_3$, 97%), cerium(III) acetylacetonate hydrate ($\text{Ce}(\text{acac})_3 \cdot x\text{H}_2\text{O}$), chromium(III) acetylacetonate ($\text{Cr}(\text{acac})_3$, 97%), were purchased from Sigma Aldrich. Hexacarbonylmolybdenum ($\text{Mo}(\text{CO})_6$, 98%) was purchased from Thermo Scientific. Avcarb MGL 370 carbon paper was used as the loading substrate and gas diffusion layer in the fuel cell, and a commercial Pt cathode (2 mg cm^{-2}) was used as the counterelectrode, and both were directly purchased from Fuel Cell Store. Sustainion XA-9 anion exchange membrane and Sustainion XA-9 ionomer were purchased from Dioxide Materials. Palladium black (high surface area) was purchased from Fuel Cell Store.

Solutions, each containing a single metallic salt, were prepared as precursors for the robotic system. Most metallic salts were dissolved in ethanol to prepare 50 mM solutions. Salts that are insoluble or with low solubility in pure ethanol (palladium chloride, bismuth chloride and niobium chloride) were dissolved in aqueous hydrochloric acid solution (37%) to final concentrations of 20 vol% HCl(aq) and 50 mM metallic element concentration in ethanol. Both AuCl_3 and HAuCl_4 would be slowly reduced to metallic gold in the presence of ethanol, even when the solution is acidified with 20 vol% HCl(aq). Thus, we dissolve the HAuCl_4 in deionized water to prepare a 50 mM solution.

Liquid handling with Opentrons

The OT-2 liquid-handling robot was purchased from Opentrons. Two pipette channels were installed: P20 Single Channel Gen2 (20 μL tip) and P300 Single Channel Gen2 (300 μL tip). The liquid dispensing rate was optimized to be $0.378 \mu\text{L s}^{-1}$. A customized 3D-printed 36-slot sample stage was used to place carbon strips of size $1.0 \text{ cm} \times 1.3 \text{ cm}$. A 96-well plate (300 μL volume for each well) was used for mixing different solutions. The mixing protocol was executed by the 300 μL tip with a fast mixing rate of $7.56 \mu\text{L s}^{-1}$ to ensure a turbulent, homogeneous mix.

In situ electrodeposition synthesis

For AL, because the goal was to select the best recipe, only 10 μL of the precursor mixture was dropcast by Opentrons on each sample (Avcarb MGL 370 strip) of size $1.0 \text{ cm} \times 1.3 \text{ cm}$, which was pre-cut by laser to ensure high size consistency. After dropcasting, the samples were naturally dried in air for at least 2 h for the ethanol solvent to evaporate. During the later robotic electrochemical treatment, the 0.3 cm fringe was clamped by the sample holder and did not take part in reactions. Each sample was first immersed in the electrolyte for 1 min, and then activated with cyclic voltammetry for 20 scans (50 mV s^{-1}), and a linear voltammetry scan for once (10 mV s^{-1}). The potential range was from -1.0 V to 0.2 V compared with Hg/HgO . The electrolyte was a mixture of 1.0 M potassium hydroxide (KOH) and 1.0 M potassium formate (HCOOK). The treatment occurred at ambient conditions. To process samples with a higher loading density, more activation cycles would be required. After this process, nanostructures would be generated in situ on the carbon fibre for further electrochemical testing.

Nanoparticle synthesis

This method was adapted from a previous report³⁸. Specifically, CTAC (50 mg) and oleylamine (5 ml) were mixed and sonicated for about 15 min in a glass vial. After that, Pd(acac)₂, Pt(acac)₂, Cu(acac)₂, HAuCl₄, Ir(acac)₃, NbCl₅, Ce(acac)₃·xH₂O, Cr(acac)₃, glucose (60 mg) and Mo(CO)₆ (33 mg) were added into the vial with designated ratios. The total amount of the metal precursor added was controlled to be 0.125 mmol, and the exact mass of metal precursor added depends on the specific recipe. The mixture was then subjected to sonication for 1 h. The vial was then heated to 220 °C and kept at this temperature for 2 h under vigorous magnetic stirring. The black colloidal product was collected by centrifugation and washed at least twice with a mixture of ethanol and cyclohexane (1:1 in volume ratio). Then, the product was subjected to ultrasonication in 0.5 M acetic acid (in ethanol) for 2 h and then centrifuged to further remove organic impurities. The final powder product was collected after washing with ethanol solution twice.

Electrochemical testing with the 7-axis robot

The 7-axis xArm robotic arm with gripper was purchased from UFactory. A customized Cu–Au connection plate was fabricated to enable the connection of the electrode sample holder with the BioLogic Potentiostat (SP-150e). The electrolyte for the three-electrode setup test was a mixture of 150 ml 1.0 M KOH and 1.0 M HCOOK. Hg/HgO electrode (filled with 1.0 M KOH) was used as the reference, and the Pt foil electrode of size 1.0 cm × 1.0 cm was used as the counter. Pure N₂ gas (Airgas, Ultra High Purity) was continuously sparged into the electrolyte during tests. All the three-electrode tests occurred at ambient temperature. For each sample, a typical test time is around 20 min. The electrolyte would be changed, and the cell would be cleaned thoroughly every 10 samples. Alkaline solutions such as KOH may etch glass and introduce impurities into the electrolyte, but considering the relatively short amount of time for the testing, such an effect was considered minimal in this work.

Membrane electrode assembly fuel cell testing

The membrane electrode assembly flow electrolyser was purchased from the Fuel Cell Store. It has two compartments: an anolyte chamber with a titanium anode flow field and a catholyte chamber with 904-L stainless steel flow field. It also has a PID (proportional–integral–derivative) temperature controller and two customized heating pads attached to the two metallic blocks. Sustainion X37-50 was used as the anion exchange membrane for ion conduction across the inner circuit, and a commercial platinum black catalyst (2.0 mg cm⁻²) was used as the standard cathode for the oxygen reduction reactions. The commercial palladium black anode was fabricated by air-spraying 2.0 mg cm⁻² of palladium black (with 30 wt% of Sustainion XA-9 ionomer) on the carbon paper. The formate anode size was 1.0 cm × 1.0 cm. To better compare the performance of the formate oxidation, we oversized our oxygen reduction cathode to be 1.4 cm × 1.6 cm. All the current and power densities were reported against the formate anode size (1.0 cm²), which was the research topic of this work. O₂ gas (Airgas, Ultra High Purity) was constantly flowing in and out at 10 sccm through silicone tubes with a mass flow controller (Alicat) at ambient pressure. The anolyte (2.0 M KOH and 1.0 M HCOOK) was also flowing in and out at a constant flow rate of 20 ml min⁻¹ with a peristaltic pump. The membrane electrode assembly block was heated to 60 °C, and the electrolyte was heated to 70 °C on a hot plate with a thermometer immersed in the solution. All the parameters (including flow rate and temperature) are the optimized values with our setup for this work. For the polarization curve test, the linear sweep voltammetry method at a scan rate of 10 mV s⁻¹ was applied, similar to that in a previous literature report³⁹. CO₂ and carbonate are generated during the formate oxidation reaction, leading to a change in the local pH. But for the flow cell test, considering that a fresh solution was pumped in and out of the flow field constantly, this effect was insignificant.

We used both catalysts synthesized from the in situ electrochemical deposition method and the multi-element nanoparticle bulk synthesis method. The in situ electrochemical method provides a high-throughput way for sample preparation and performance optimization but suffers from low mass loading (typically below 0.2 mg cm⁻²). AL was conducted using catalysts synthesized by in situ electrochemical deposition for rapid screening. Promising or representative recipes were subsequently validated in fuel cell testing (Supplementary Fig. 3) using HEA powder prepared by conventional synthesis. For the device testing, we mixed the HEA nanoparticles with 30 wt% of Sustainion XA-9 ionomer and air-sprayed them onto the carbon paper, which improves the mass loading (controlled to be 2.0 mg cm⁻²) of our catalyst, and thus the device performance. However, such a workflow remains mostly manual for us. The key point of this paper is to leverage the robotic platform to quickly screen catalyst recipes using the three-electrode testing method, which is a more widely used electrochemical testing method in the literature.

Structural characterization

XRD was conducted on the X-ray diffractometer (Aeris Research edition) using a copper target at a voltage of 40 kV and a current of 15 mA. Deionized water was used to rinse the remaining KOH and HCOOK off the electrode surface before XRD testing. SEM was performed with a Zeiss Merlin High-resolution scanning electron microscope at the MIT Materials Research Laboratory. Characterization with transmission electron microscopy (TEM), energy-dispersive spectroscopy (EDS) and scanning transmission electron microscopy were performed with the Thermo Fisher Scientific Themis Z G3 aberration-corrected scanning transmission electron microscope with a resolution of <0.6 Å at MIT.nano. In situ XAS experiments were performed at the SPring-8 BL12B2 Taiwan Beamline, using a custom-designed electrochemical cell tailored for in situ XAS measurements. Fluorescence signals were acquired using a Lytle detector. A custom-made in situ XAS cell was used for this experiment. The electrochemical setup used a three-electrode system, consisting of a working electrode, a counterelectrode (Pt wire) and a reference electrode (Hg/HgO), immersed in an electrolyte solution composed of 1.0 M KOH and 1.0 M HCOOK.

DFT calculations

First-principles calculations were performed using spin-polarized DFT^{40,41} implemented in the Vienna ab initio simulation package⁴² with the Perdew, Burke and Ernzerhof⁴³ exchange-correlation potential within the generalized gradient approximation. The projector augmented wave pseudo-potential⁴⁴ was used to describe core electrons. For all optimization calculations, the cutoff energy was set at 450 eV and *k*-space was sampled by 3 × 3 × 1 for all models. The convergence criteria for energy and force were set at 10⁻⁴ eV and 0.02 eV Å⁻¹, respectively. The van der Waals interaction has been considered using the Grimme dispersion scheme⁴⁵.

For a certain reaction (A → B), the reaction free energy Δ*G* (including the DFT total energy, zero-point energy, vibrational enthalpy (thermal corrections) and vibrational entropy (at *T* = 298.15 K)) is defined as follows:

$$\Delta G \equiv G_B - G_A$$

Further details on the modelling approaches and model selection criteria can be found in Supplementary Note 6, whereas the energy calculations and treatment methods for formate fuel cells are described in Supplementary Note 7.

Data availability

The data that support the findings of this study are included in the main text, and the source files are available from the corresponding author upon request. Source data are provided with this paper.

Code availability

The code supporting the findings of this study is available at GitHub (<https://github.com/zhang21mit/CRESt>) and can be obtained from the corresponding author upon request.

38. Li, H. et al. Fast site-to-site electron transfer of high-entropy alloy nanocatalyst driving redox electrocatalysis. *Nat. Commun.* **11**, 5437 (2020).
39. Ni, W. et al. Synergistic interactions between PtRu catalyst and nitrogen-doped carbon support boost hydrogen oxidation. *Nat. Catal.* **6**, 773–783 (2023).
40. Hohenberg, P. & Kohn, W. Inhomogeneous electron gas. *Phys. Rev.* **136**, B864–B871 (1964).
41. Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140**, A1133–A1138 (1965).
42. Kresse, G. & Hafner, J. *Ab initio* molecular dynamics for liquid metals. *Phys. Rev. B* **47**, 558–561 (1993).
43. Perdew, J. P., Ernzerhof, M. & Burke, K. Rationale for mixing exact exchange with density functional approximations. *J. Chem. Phys.* **105**, 9982–9985 (1996).
44. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).
45. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *J. Chem. Phys.* **132**, 154104 (2010).

Acknowledgements We thank A. Quinn and F. Brushett from the Department of Chemical Engineering at MIT for discussions. The in situ XAS spectra obtained from beamline TPS 32A at

the National Synchrotron Radiation Research Center and Taiwan Beamline 12B2 at SPring-8 are appreciated. The TEM images obtained from MIT.nano are appreciated. This work was supported by the Defense Advanced Research Projects Agency under agreement no. HR00112490369 and the Defense Threat Reduction Agency (award no. HDTRA1-20-2-0002) Interaction of Ionizing Radiation with Matter University Research Alliance. H.-T.W. was supported by the National Science and Technology Council, Taiwan (grant no. 112-2112-M-Q32-017-MY2).

Author contributions Z.Z., Z.R. and J.L. conceptualized the study. Z.Z., Z.R., C.-W.H., Weibin C., Z.-W.H., C.-F.L., Y.-R.L., Y.-C.S., H.-T.W., I.I.A., P.A., Y.S.-H. and J.L. devised the methodology. Z.Z., Z.R., C.-W.H., Weibin C., Z.-W.H., C.-F.L., A.P., S.M., Y.S.-H., H.X., D.J.Z., Y.G., Weiyin C., H.S., Y.N., Y.T., Y.-R.L., Y.-C.S., S.L., H.-T.W., I.I.A., P.A., Y.S.-H. and J.L. investigated the study. Z.Z., Z.R., C.W.H., Weibin C., C.-F.L., S.M., Y.S.-H. and H.-T.W. visualized the projects. J.L. helped with funding acquisition and supervision. Z.Z., Z.R. and J.L. wrote the original draft. Z.Z., Z.R., C.W.H., Weibin C., Z.-W.H., C.-F.L., A.P., S.M., Y.S.-H., H.X., D.J.Z., Y.G., Weiyin C., H.S., Y.N., Y.T., Y.-R.L., Y.-C.S., S.L., H.-T.W., I.I.A., P.A., Y.S.-H. and J.L. reviewed and edited the paper.

Competing interests The authors declare no competing interests.

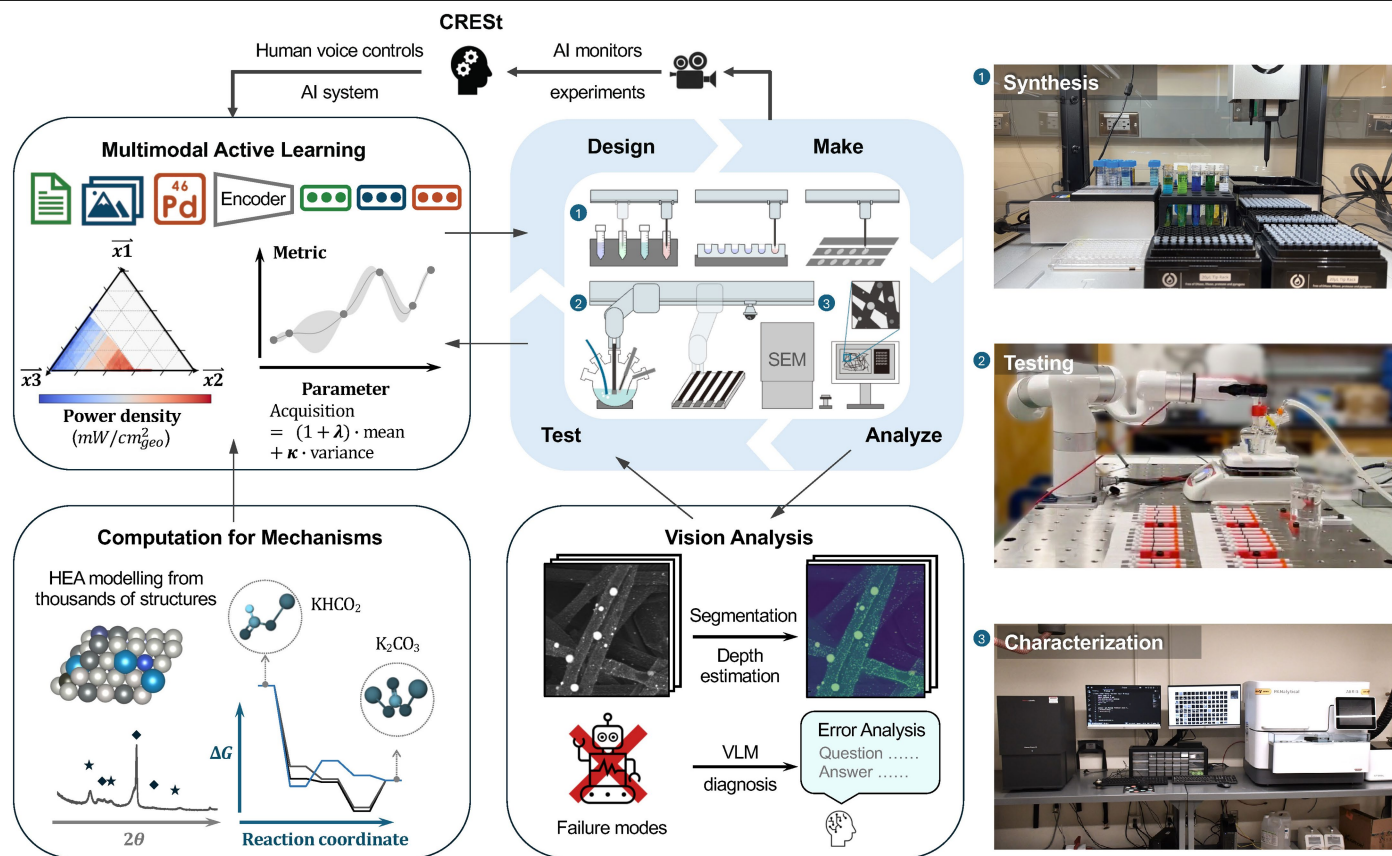
Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-025-09640-5>.

Correspondence and requests for materials should be addressed to Ju Li.

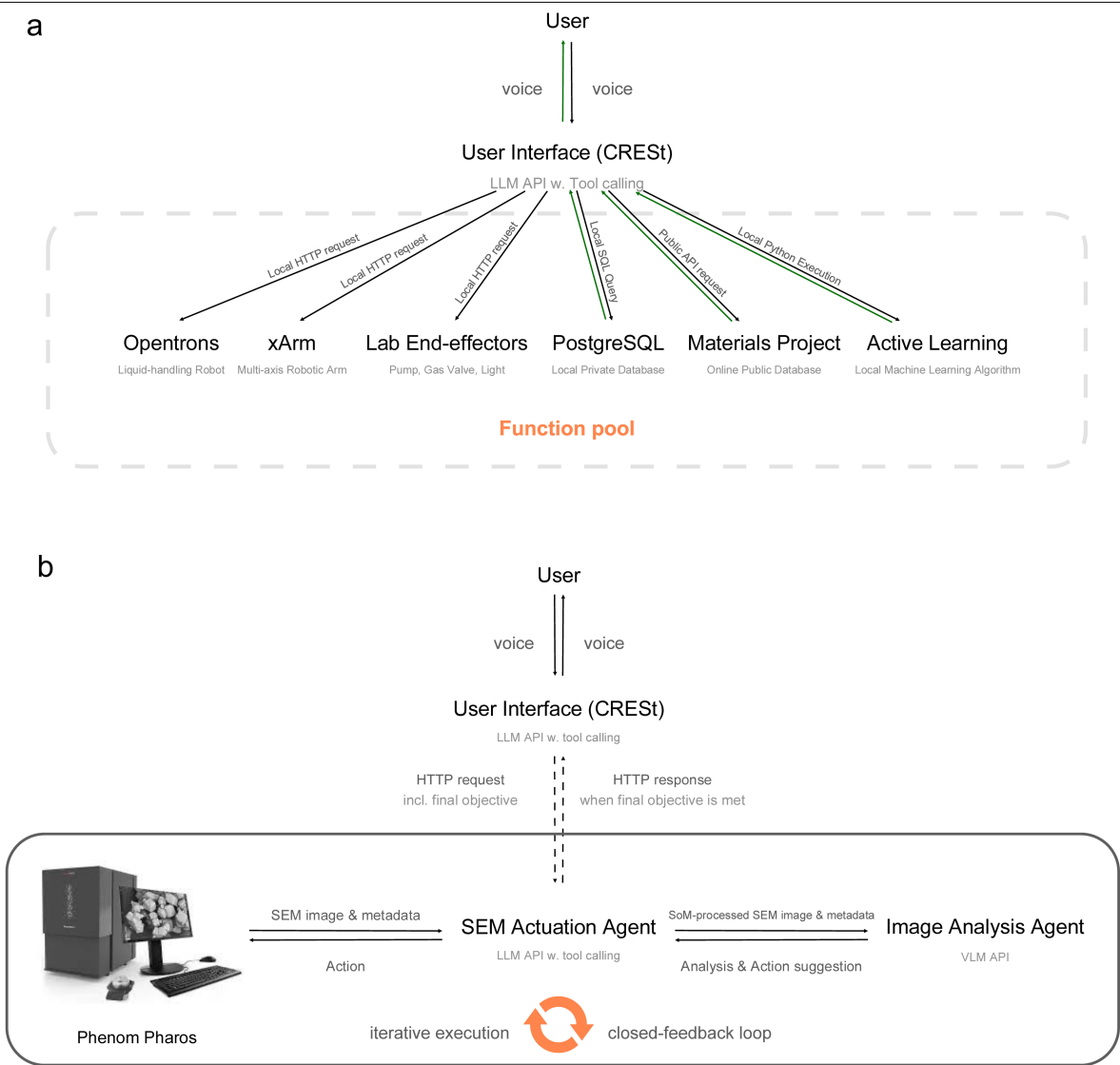
Peer review information *Nature* thanks Liang Zhang and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



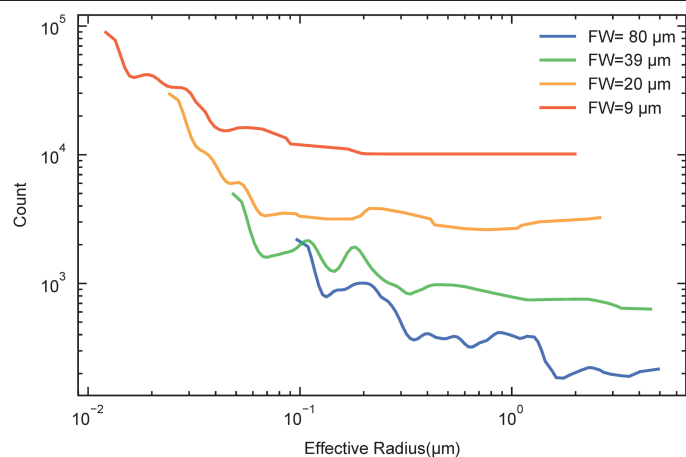
Extended Data Fig. 1 | The workflow of electrocatalyst discovery guided by CREST. A large multimodal model interfaces with researchers through text and voice, monitors experiments via visual analysis, and autonomously controls robotic instruments for electrochemical materials research. The multimodal active learning (MAL) module integrates literature and human knowledge, microstructural images, and composition and process tuning, operating in a compressed latent space to guide candidate selection. Robotic subsystems

carry out sample synthesis, testing, and characterization, while the vision module applies computer vision techniques such as segmentation and depth estimation to analyze microstructures. In parallel, a vision language model diagnoses experimental error modes to improve reproducibility. Experimental results are iteratively incorporated into the active learning loop, and optimized materials are validated through mechanistic studies and device-level testing.

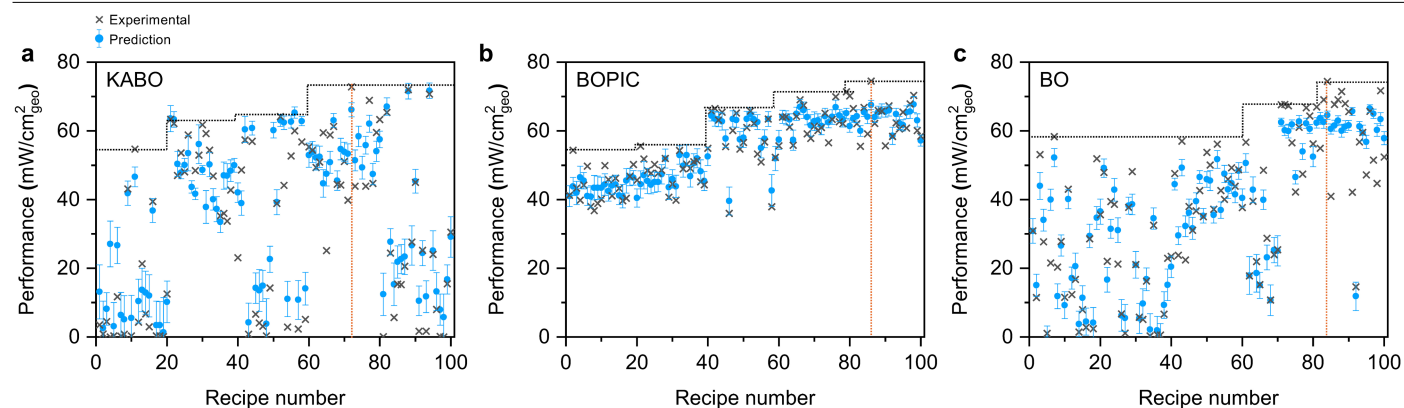


Extended Data Fig. 2 | The CREST coding architecture. (a) Architecture of the CREST main experimental platform. CREST acts as a laboratory orchestrator, dynamically invoking routines from a shared function library that includes instrument-control systems and Python workflows to execute real-world experiments. (b) Architecture of the autonomous scanning electron microscope (SEM) module. A three-agent loop coordinates (i) a dialogue/orchestration agent at the CREST layer that interfaces with users and sets goals, (ii) an SEM control

agent that performs navigation, focusing and zooming, and (iii) a vision agent that analyzes images and recommends the next action. The loop iterates until the imaging objective is met, after which the SEM agent returns the final micrograph and an execution report to CREST. A separate computer vision analysis module is also utilized to analyze SEM images for statistical features (Supplementary Note 4).

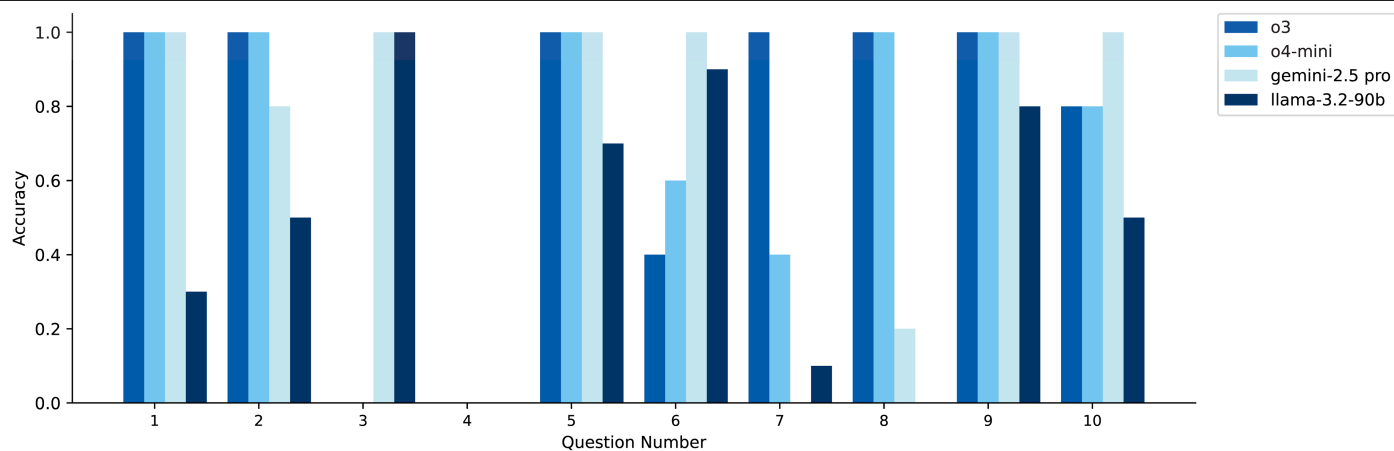


Extended Data Fig. 3 | Example of the particle number distribution function in the logarithm scale. The feature statistics were obtained from SEM images captured at varying field widths across our dataset. The effective radius is calculated by converting the measured particle area into an equivalent spherical radius. Our analysis shows that obtaining accurate feature statistics, especially those related to particle size, requires selecting a field of view that is appropriately scaled to the size of the particles being analyzed.



Extended Data Fig. 4 | Comparison plot of the active learning campaign for the three algorithms. (a) Knowledge-Assisted Bayesian Optimization (KABO). (b) Bayesian Optimization with Policy Improvement Constraints (BOPIC). (c) Standard Bayesian Optimization using the Upper Confidence Bound (UCB) acquisition function. Both KABO and BOPIC achieved continuous improvements batch after batch, whereas standard BO required ~70 samples before making progress—a critical limitation when each experiment is costly. KABO, likely due

to its integration of domain knowledge, showed the strongest alignment between predictions and experimental results, particularly in high-performance regions, and also identified the global optimum first (indicated by the orange vertical line). BOPIC proved especially effective in sampling high-performance candidates relative to the other two methods. Ultimately, all three algorithms converged to a similar high-performance value, consistent with the fixed ground truth in the chemistry space.



Extended Data Fig. 5 | Comparison of performance for the state-of-the-art vision language model (VLM) on a real-world materials science experimental question set. Considering the stochasticity of the vision language models,

each question was tested for five times for each model to compute the average accuracy. Examples could be found in the Appendix of the Supplementary file.