EES Catalysis

PERSPECTIVE

View Article Online
View Journal

Cite this: DOI: 10.1039/d6ey00079g

Finding the hidden catalytic knowledge from literature data

Yuhang Wang, Yong Wang and Hao Li *

Over decades of catalytic research, a vast amount of knowledge has been documented in the literature, yet its potential scientific value has not been fully explored. Catalytic performance is determined by multiple factors such as electronic structure and reaction conditions, exhibiting complex nonlinear structure–performance relationships. Simultaneously, differences in experimental conditions and inconsistent data dimensions among different studies make it difficult to directly use existing data for rule extraction and catalyst design. This perspective systematically summarizes three strategies for discovering new catalytic knowledge from literature data: first, leveraging "human intelligence" and statistical analysis to discover new catalytic knowledge through literature integration and mechanistic insights; second, constructing interpretable descriptors using symbolic regression and machine learning to achieve quantitative prediction of catalytic performance; and third, combining large language models and AI agents for multi-source data integration, knowledge extraction, and intelligent catalyst recommendation. Overall, these data-driven methods can transform scattered experience into computable design criteria, enabling a new paradigm for the rational design and efficient screening of catalytic materials, and accelerating the development of catalysis research towards a digital materials ecosystem and closed-loop research model that integrates AI, theoretical calculations, and autonomous experiments.

Received 16th April 2026,
Accepted 12th May 2026

DOI: 10.1039/d6ey00079g

rsc.li/eescatalysis

Broader context

Catalysis plays a central role in sustainable energy conversion, clean fuel production, carbon recycling, and environmental remediation. Over decades of research, enormous amounts of catalytic knowledge have been accumulated in the scientific literature, covering experimental performance, reaction conditions, material structures, and theoretical calculations. However, much of this knowledge remains scattered across individual studies and is difficult to reuse for rational catalyst design. This perspective discusses how hidden catalytic knowledge can be rediscovered from existing literature data through three complementary strategies: human-intelligence-guided data analysis, regression-model-based descriptor development, and AI-agent-assisted knowledge extraction. By summarizing representative examples across key electrocatalytic reactions, we show how historical data can be transformed into transferable design principles and computable structure–performance relationships. More broadly, this perspective highlights the importance of high-quality databases, expert validation, and AI–theory integration for building a digital materials ecosystem and accelerating the transition from empirical catalyst discovery to predictable, closed-loop catalyst design.

1. Introduction

With the rapid development of catalysis science and the continuous accumulation of numerous experimental data and first-principles calculations, the field of catalysis has entered a data-intensive research phase.[1] The rapid development of high-throughput computing platforms, automated experimental techniques, and open databases (*e.g.*, Materials Project,[2] AFLOW,[3] Digital Catalysis Platform (DigCat),[4,5] Catalysis-Hub,[6] and Open Catalyst[7]) has enabled the systematic storage and sharing of massive amounts of structural, energy, and reaction performance data. Compared with the traditional research model centered on a single system, current research focuses more on the common laws and transferable design principles across material systems.[8] Against this backdrop, how to extract reliable knowledge from existing literature and databases, and realize the transformation from data accumulation to knowledge discovery, is becoming an important research direction driving the rational design of catalysts. Systematic integration and reanalysis of existing data can reveal underlying structure–performance relationships, reduce

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

experimental trial-and-error costs, and accelerate the discovery of novel catalytic materials.

Despite the rapid accumulation of catalytic data, uncovering their full potential remains particularly challenging.[9] On the one hand, experimental conditions (electrolytes, pH, potential ranges, support environments, *etc.*) vary significantly across studies, limiting the comparability of data within small sample ranges, which makes it difficult to directly extract cross-system patterns. On the other hand, literature data often suffers from inconsistent dimensions, missing key parameters, and varying characterization standards, causing potential structure–performance correlations to be obscured by noise. Therefore, there is an urgent need to develop systematic methods to reorganize, reanalyze, and remodel historical data to extract physically meaningful catalytic patterns from complex and heterogeneous data, providing a reliable basis for predictable catalyst design.

In recent years, several key strategies have emerged for extracting new catalytic knowledge from existing data.[10,11] First, through knowledge induction driven by ''human intelligence'', combined with mechanistic understanding and cross-literature statistical analysis, potential commonalities and anomalous correlations can be identified, further proposing new physical insights or reaction mechanisms. Second, regression models and descriptor engineering methods based on large-scale data can construct physically meaningful structure–performance relationships from historical data, enabling quantitative prediction of catalytic performance and providing an interpretable theoretical basis for catalyst design. Furthermore, with the development of large language models (LLMs) and automated tools, AI agents are emerging as powerful tools for data mining. They can integrate literature, databases, and computational results to achieve high-throughput knowledge extraction and candidate material screening, thereby significantly improving the exploration efficiency of complex materials space.

Despite the broad applicability of data-driven approaches, such as their successful application in data-driven materials design,[12] this perspective does not attempt to cover all aspects of the field. Instead, we focus on a key but underdisc ussed question: what new catalytic insights can be gained by revisiting existing literature data? Based on the above background, this perspective systematically summarizes recent progress along three representative research paths: human intelligence-driven knowledge induction, regression model-based descriptor development, and AI agent-assisted knowledge discovery, illustrated through typical reactions such as carbon dioxide reduction reaction (CO_2RR),[13–18] oxygen reduction reaction (ORR),[19–24] oxygen evolution reaction (OER),[25–28] hydrogen evolution reaction (HER),[29–31] electrocatalytic ammonia synthesis,[32] and chlorine evolution reaction (CER).[33] By summarizing the applicability and advantages of different strategies, this perspective aims to construct an overall framework for data-driven catalysis research, clarify the fundamental role of high-quality databases in the digital catalysis ecosystem, and highlight emerging opportunities brought about by the deep integration of AI and theoretical computation, thereby providing guidance for achieving efficient and predictable rational catalyst design.

2. Strategies for discovering new catalytic knowledge based on old data

Owing to the rapid development of catalysis science in recent years, a vast amount of experimental and theoretical data has been accumulated in the field, providing an important foundation for a deeper understanding of structure–performance relationships. Notably, through the systematic organization and re-mining of existing catalysis data, researchers are constantly gaining new insights and design principles for catalysis, making the discovery of new knowledge from old data a crucial paradigm driving catalysis research. However, facing the rapid expansion of data scale and the increasing complexity of information, there is an urgent need to systematically summarize existing methods to clarify the advantages and applicability of different strategies in knowledge extraction and catalyst design. Currently, the three most representative approaches include: (1) knowledge induction and mechanism understanding based on human intelligence and direct statistical analysis; (2) regression models and descriptor construction based on big data to establish predictable structure–performance relationships; and (3) automated data mining and catalytic knowledge discovery using advanced AI agents. Moving from human-experience-driven mechanistic understanding to data-driven predictable modeling and then to AI-driven automated knowledge discovery, this progression reflects a paradigm shift in catalysis research from empirical induction toward intelligent design. In the following sections, we systematically discuss how these strategies facilitate the discovery of new catalytic knowledge.

2.1. Discovering new knowledge using human intelligence and statistical analysis

In this section, the examples are grouped into three categories: (i) extraction of common patterns, (ii) anomaly-driven discovery, and (iii) identification of research gaps and design directions.

Electrocatalytic CO_2RR has attracted widespread attention due to its potential applications in carbon resource recycling and renewable energy conversion,[15,34] resulting in a wealth of experimental and theoretical data. Formic acid (HCOOH) has become an important target product due to its high economic feasibility and ease of storage and transportation.[14] However, it remains challenging to summarize universal structure–performance rules from single experiments or a small number of independent studies. Notably, when examining CO_2RR-to-HCOOH catalysts from a higher-dimensional, systematic perspective, some potential commonalities gradually emerge. For example, Guo *et al.*,[35] through systematic mining of Sn-based catalyst data, discovered that various materials, including Sn-based metals, oxides, sulfides, and alloys, exhibit high intrinsic HCOOH selectivity, revealing the crucial role of the Sn center in

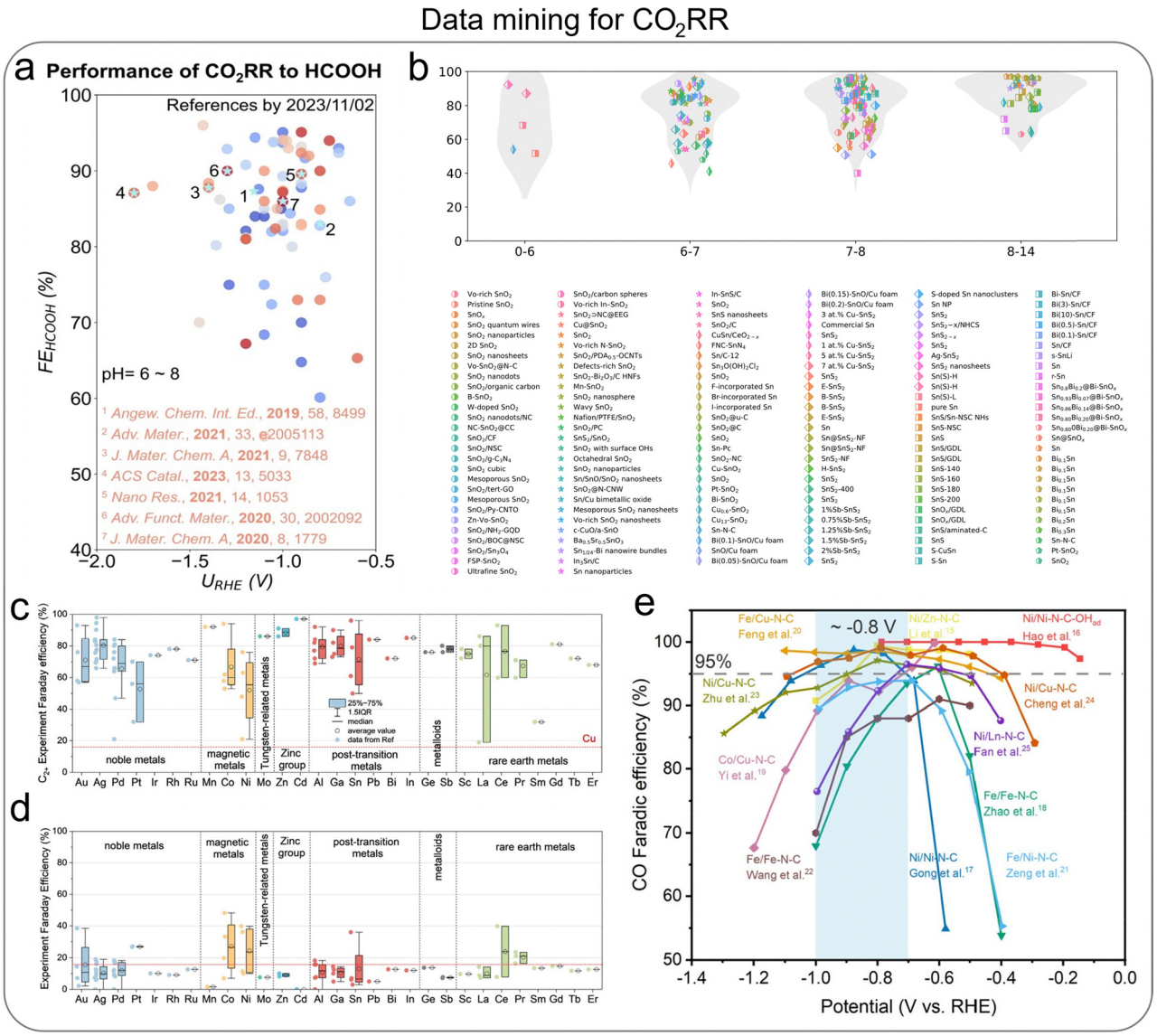

Fig. 1 (a) Systematic statistical results of the selectivity of the main product in CO_2RR by Sn-based catalysts. Reproduced with permission from ref. 35 copyright 2024, Wiley-VCH. (b) pH dependence of Faraday efficiency (FE) for Sn-based catalysts *via* large-scale data mining. Reproduced with permission from ref. 36 licensed under a Creative Commons License CC BY 4.0. Comparison of experimental Faraday efficiencies of Cu-based single-atom alloys (SAAs) in CO_2RR for C_{2+} product formation (c) and HER (d), compiled from the DigCat database. Reproduced with permission from ref. 37 licensed under a Creative Commons License CC BY 4.0. (e) Comparison of CO faradaic efficiency for various DACs reported in the past three years. Reproduced with permission from ref. 38 copyright 2023, American Chemical Society.

regulating the reaction pathway (Fig. 1a). This result indicates that systematic analysis of existing data reveals consistent HCOOH selectivity across Sn-based materials while capturing the general trend of reported literature, providing important guidance for the rational design of highly selective CO_2RR catalysts.

With the continuous development of the DigCat database, more and more scattered experimental data can be systematically integrated and quantitatively analyzed, thus providing new opportunities to reveal universal rules. Based on large-scale statistical data, Wang *et al.*[36] constructed a relationship between pH and the FE of HCOOH, finding that Sn-based catalysts exhibited a good correlation in reported experimental data (Fig. 1b): as the electrolyte pH increased, the FE for CO_2RR to HCOOH formation by Sn-based catalysts generally showed a significant upward trend. This finding further contributes to the establishment and development of theoretical understanding of pH-dependent volcano relations. Overall, this study demonstrates that systematic analysis based on experimental databases and the extraction of patterns using human intelligence can effectively promote the construction of new theoretical models, reflecting the important role of data-driven research in the development of catalytic mechanisms.

Although Cu-based single-atom alloys (SAAs) have been extensively studied in CO_2RR, their overall performance patterns and further optimization directions still lack systematic

understanding. To overcome this limitation, Wang *et al.*[37] conducted a systematic statistical analysis of approximately 50 articles and over 80 data entries in the DigCat database, extracting potential common patterns from scattered historical data. As shown in Fig. 1c, doping Cu-based SAA with 29 different elements significantly improved C_{2+} product selectivity (average FE exceeding 50%). More importantly, despite significant differences in C_{2+} selectivity among different doping systems, the FE of the competitive HER remained relatively stable (Fig. 1d). This counterintuitive phenomenon indicates that the main role of single-atom doping is not to suppress HER, but rather to directionally promote C–C coupling kinetics, *i.e.*, to improve C_{2+} selectivity by accelerating the target reaction pathway. Overall, this finding highlights that that human-driven re-mining of old data can extract catalytic laws across research scales, providing a more targeted theoretical basis for the rational design of Cu-based alloy catalysts.

To date, most dual-atom catalysts (DACs) have been employed in the electrocatalytic CO_2RR with the initial hope of generating substantial multicarbon products (*e.g.*, ethane and ethanol). Given that DACs offer two adjacent metal sites capable of enhancing surface *CO coverage, C–C coupling would appear feasible. By systematically reviewing the experimental CO_2RR data from 11 typical M–N–C DACs and combining it with DFT calculations and surface Pourbaix analysis, Yang *et al.*[38] found that although DACs possess adjacent bimetal sites, CO is the dominant product in nearly all experimental reports, and even Cu-containing DACs fail to effectively generate C–C coupling products (Fig. 1e). Theoretical simulations further reveal that under reaction conditions, the stable surface state of DACs is pre-covered by *CO at the metal–metal bridge sites, and the hydrogenation of *CO to *COH or *CHO is thermodynamically unfavorable, rendering C–C coupling a high-barrier process. In summary, both statistical analysis and computational results indicate that the strong *CO adsorption leading to site poisoning and the difficulty of *CO hydrogenation are the fundamental reasons why M–N–C DACs struggle to achieve C–C coupling. Overall, these CO_2RR examples collectively show that human-guided reanalysis of literature data can reveal both general selectivity trends and counterintuitive deviations that are difficult to recognize from individual studies.

Electrochemical nitrate reduction reaction (NO_3RR) offers a potential low-carbon ammonia synthesis route as an alternative to the traditional Haber–Bosch process, but the rational design of efficient catalysts still faces challenges such as unclear structure–performance relationships. As shown in Fig. 2a and b, Jiang *et al.*[39] systematically analyzed experimental data from over 60 reported M–N–C catalysts, extracting two key patterns. First, most studies focus on alkaline and neutral conditions, while acidic systems remain relatively underexplored due to strong competition from the hydrogen evolution reaction. Second, although experimental characterization of coordination environments is limited, statistical results indicate that pyrrole-coordinated M–N–C catalysts generally exhibit high faradaic efficiency under both alkaline and neutral conditions. Overall, this study demonstrates that even with scattered and incomplete historical data, systematic reanalysis can still reveal catalytic patterns across different studies, providing a crucial data-driven basis for the rational design of NO_3RR catalysts.

In the fields of electrocatalysis and thermocatalysis, the regulatory effects of external fields and unconventional promoters on reaction performance have attracted increasing attention.[40] However, their underlying mechanisms are often scattered across different studies, lacking a unified understanding. Through systematic mining of existing literature data, researchers have begun to reveal potential commonalities from a higher perspective. As shown in Fig. 2c, You *et al.*,[40] based on large-scale data statistics, found that research reports on the magnetic field (MF) enhancement effect in electrocatalysis show a continuous upward trend. Further DFT calculations show that in the constructed RuN_{4r} model, the spin state of single-atom Ru significantly affects its adsorption capacity for $*NH_2$ intermediates. Among them, the low-spin ground state of Ru (0.30 μB) exhibits the strongest adsorption, while the applied MF can induce higher spin states, thereby weakening the $*NH_2$ adsorption strength (Fig. 2d). On the other hand, Cao *et al.*[41] summarized experimental data on the promotion of thermocatalytic ammonia synthesis by various non-traditional promoters (*e.g.*, Ba, Ca, Li, electride, amide, and rare earth nitrides), and found that these promoters can still significantly enhance activity on weakly N-adsorbed metals (*e.g.*, Co and Ni) (Fig. 2e). Although individual studies usually explain this from the perspective of electronic or structural effects,[42] the comprehensive analysis shows that various promoters can lower the N_2 dissociation-related energy barrier, which is particularly evident in magnetic transition metal systems. Overall, these studies, through human-driven data reanalysis, propose a more unified physical picture: some promoters not only regulate electronic structure, but may also lower the spin-related energy barrier by adjusting the spin polarization of active sites, generating an additional spin promotion effect, providing a new theoretical perspective for overcoming traditional catalytic activity limitations.

The ORR is a crucial process in new energy devices such as hydrogen fuel cells and metal–air batteries.[20,43–45] However, its slow kinetics have long relied on expensive Pt-group metal catalysts, severely restricting the large-scale application of clean energy technologies. Therefore, a data-driven approach to deeply understand the structure–performance relationship of non-Pt catalysts and identify their intrinsic activity limitations has become an important direction in current catalysis research. Through systematic data mining, Zhang *et al.*[46] summarized experimental data from 1018 M–N–C catalysts. Although traditional volcano plots predict low activity for weakly adsorbed systems, weakly adsorbing catalysts such as Ni–N–C, Cu–N–C, and Zn–N–C still exhibit significant ORR activity and a clear pH dependence under alkaline conditions, revealing an anomalous correlation between adsorption strength and catalytic performance and highlighting limitations of conventional scaling relationships (Fig. 3a and b). Further theoretical studies show that there is a systematic shift

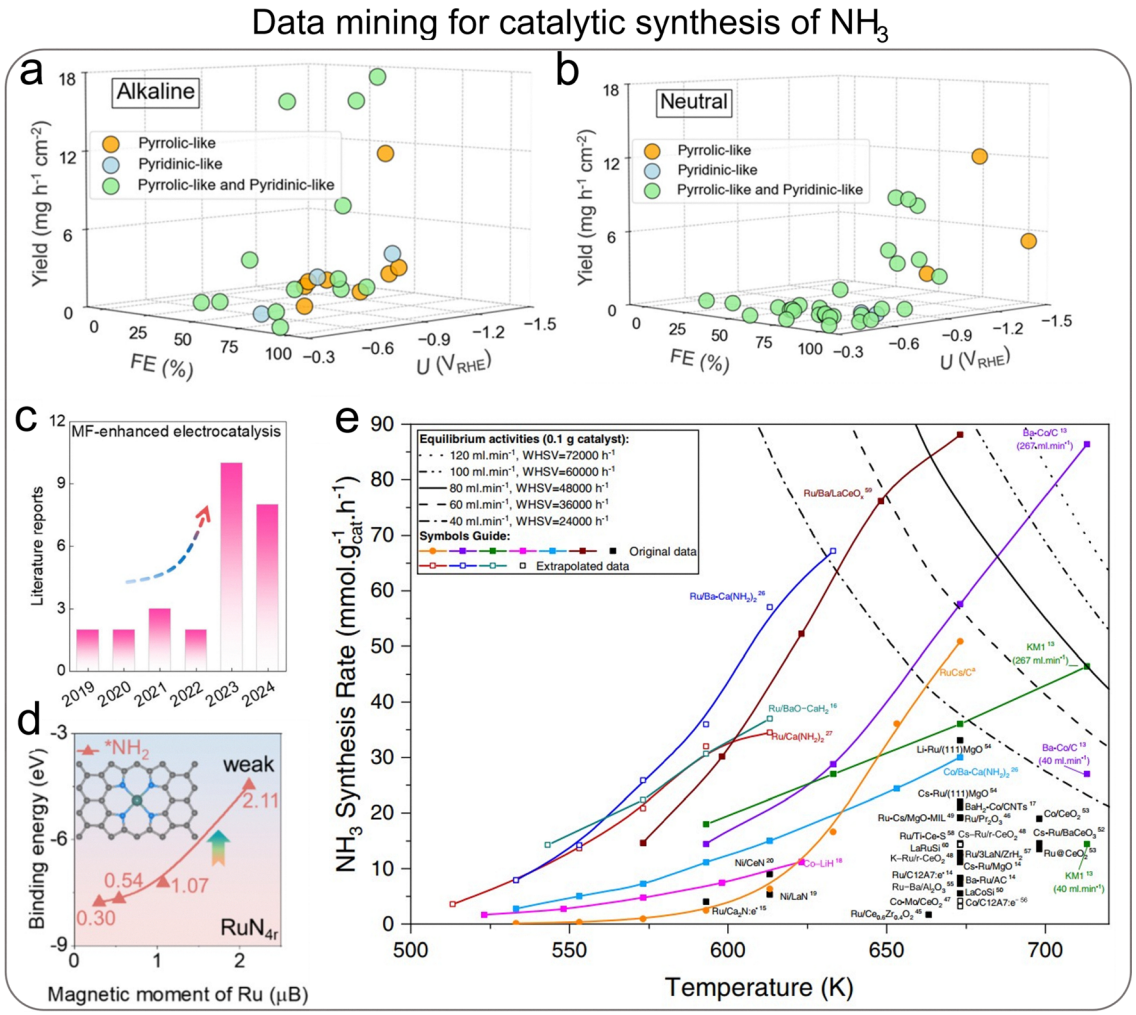

Fig. 2 Data mining analysis based on the experimental performance of 60 M–N–C catalysts in electrocatalytic NO_3RR: (a) alkaline conditions; (b) neutral conditions. Reproduced with permission from ref. 39 licensed under a Creative Commons License CC BY 4.0. (c) statistical analysis of literature data on the enhanced electrocatalytic effect of magnetic field; (d) changes in the adsorption energy of pyrrole-type N-coordinated single atom Ru (illustrated structure) under different magnetic moments for $*NH_2$. Reproduced with permission from ref. 40 licensed under a Creative Commons License CC BY 4.0. (e) Data-mined summary of previously reported experimental activities for ammonia synthesis. Reproduced with permission from ref. 41 licensed under a Creative Commons License CC BY 4.0.

in the scaling relationship between weakly adsorbed systems and moderately adsorbed systems, and their activity is significantly regulated by electric field response and solvation effects, thus exhibiting a reaction pathway different from the typical Fe/Co–N–C system. On the other hand, large-scale data analysis by Li *et al.*[47] on transition metal oxides (TMOs), a potential non-noble metal system, shows that even after optimization, the intrinsic ORR activities of Mn, Fe, and Ni-based TMOs are generally lower than those of Pt, exhibiting a clear performance upper limit (Fig. 3c). Mechanistic analysis indicates that the weak adsorption of oxygen intermediates makes O–O bond breaking the rate-determining step, while the interfacial electric field effect further increases the energy barrier, explaining the activity bottleneck of the TMO system at the atomic scale.

In addition, Liu *et al.* performed large-scale statistical mining of ORR experimental data for SACs and DACs using the DigCat database and discovered an anomaly in which the activity volcano plot of DACs exhibits a bimodal distribution inconsistent with classical theory (Fig. 3d). By further combining theoretical calculations with interpretable machine learning, they revealed that this anomaly originates from a dissociative mechanism and a dynamic switch in the rate-determining step, thereby proposing a new dual-Sabatier optima principle that revises the conventional design paradigm for single-site catalysts. This work vividly demonstrates how large-scale data mining can uncover deviations between theory and experiment, and how human intelligence (mechanistic analysis and advanced modeling) can transform such deviations into new scientific discoveries, thus providing a general paradigm for the design of multi-atom catalysts in multi-step catalytic reactions.

CER is a key reaction in the chlor-alkali industry and electrochemical oxidation processes.[33] However, the development of efficient catalysts still heavily relies on noble metal

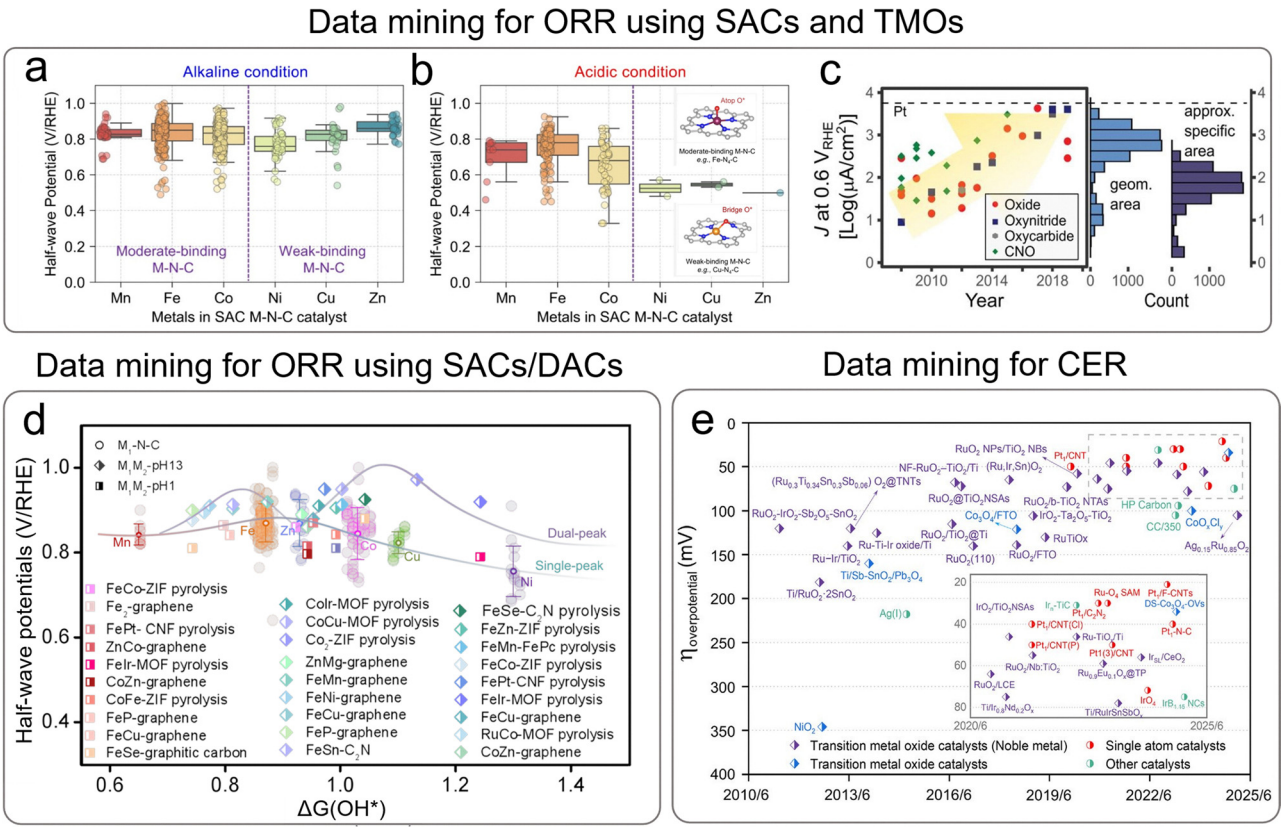

Fig. 3 (a and b) Half-wave potentials of M–N–C catalysts under alkaline and acidic conditions, respectively. Reproduced with permission from ref. 46 licensed under a Creative Commons License CC BY 4.0. (c) Current densities of transition metal oxides (TMOs) at 0.6 V/RHE by anion classification (blue: geometric; purple: specific from high-throughput data). Reproduced with permission from ref. 47 copyright 2021, Springer Nature. (d) Comparison of experimental ORR performance between SACs and DACs. Reproduced with permission from ref. 48 licensed under a Creative Commons License CC BY 4.0. (e) Overpotential summary at 10 mA cm^{-2} for various catalysts in CER. Reproduced with permission from ref. 49 licensed under a Creative Commons License CC BY 4.0.

systems, limiting their cost advantage and potential for large-scale application. Yang *et al.*[49] systematically analyzed nearly 15 years of CER reaction reports, as shown in Fig. 3e. Based on systematic data mining of CER catalyst literature since 2010 and combined with mechanistic analysis, it was found that current mainstream catalysts still focus on noble metal materials such as Ru and Ir. Most mixed metal oxide systems still require approximately 5–30% noble metal content to maintain high activity. In contrast, SACs developed in recent years can achieve lower overpotentials with extremely low noble metal loading, exhibiting higher atom utilization efficiency and catalytic potential. However, further summarizing existing data shows that non-noble metal single-atom catalysts are still almost non-existent in the CER field, especially carbon-supported non-noble metal systems, which lack systematic research. Overall, data mining reveals the continued dependence of CER catalysts on noble metals and potential research gaps, while human-powered intelligent analysis based on mechanistic understanding further points to feasible design directions for developing low-cost, high-efficiency CER catalysts.

In another fundamental electrocatalytic reaction (*i.e.*, HER), G_{H^*} (adsorption free energy of H*) is usually used as an effective activity descriptor.[50] However, Ye *et al.*[51] directly discovered a significant anomaly through data mining (Fig. 4a): G_{H^*} alone is insufficient to describe HER activity. For example, Ni–N–C still requires a large overpotential at higher G_{H^*} (~1.6 eV), while V–N–C exhibits poor activity even with a G_{H^*} close to 0 eV, demonstrating the limitations of weak adsorption descriptors in predicting actual catalytic performance. Fig. 4b further reveals that V single-atom catalysts tend to be poisoned by OH under acidic conditions with U_{RHE} = 0. Furthermore, a 2D volcano (Fig. 4c) as a function of G_{HO^*} (adsorption free energy of HO*) and G_{H^*} successfully corrects the shortcomings of G_{H^*} as the descriptor, compensating for the lack of catalytic knowledge discovery driven by data mining.

Achieving efficient water splitting relies heavily on discovering highly stable and active electrocatalysts capable of accelerating both OER and HER.[53–56] Although certain non-noble stoichiometric metal oxides (MOs) have demonstrated OER stability under alkaline conditions, their behavior in acidic environments remains largely unexplored. To address this challenge, Jia *et al.*[52] proposed a data-driven, human-assisted research strategy. The overall research workflow is illustrated in Fig. 4d–f, involving data mining from the reported literature, synthesis and electrochemical testing, and characterizations

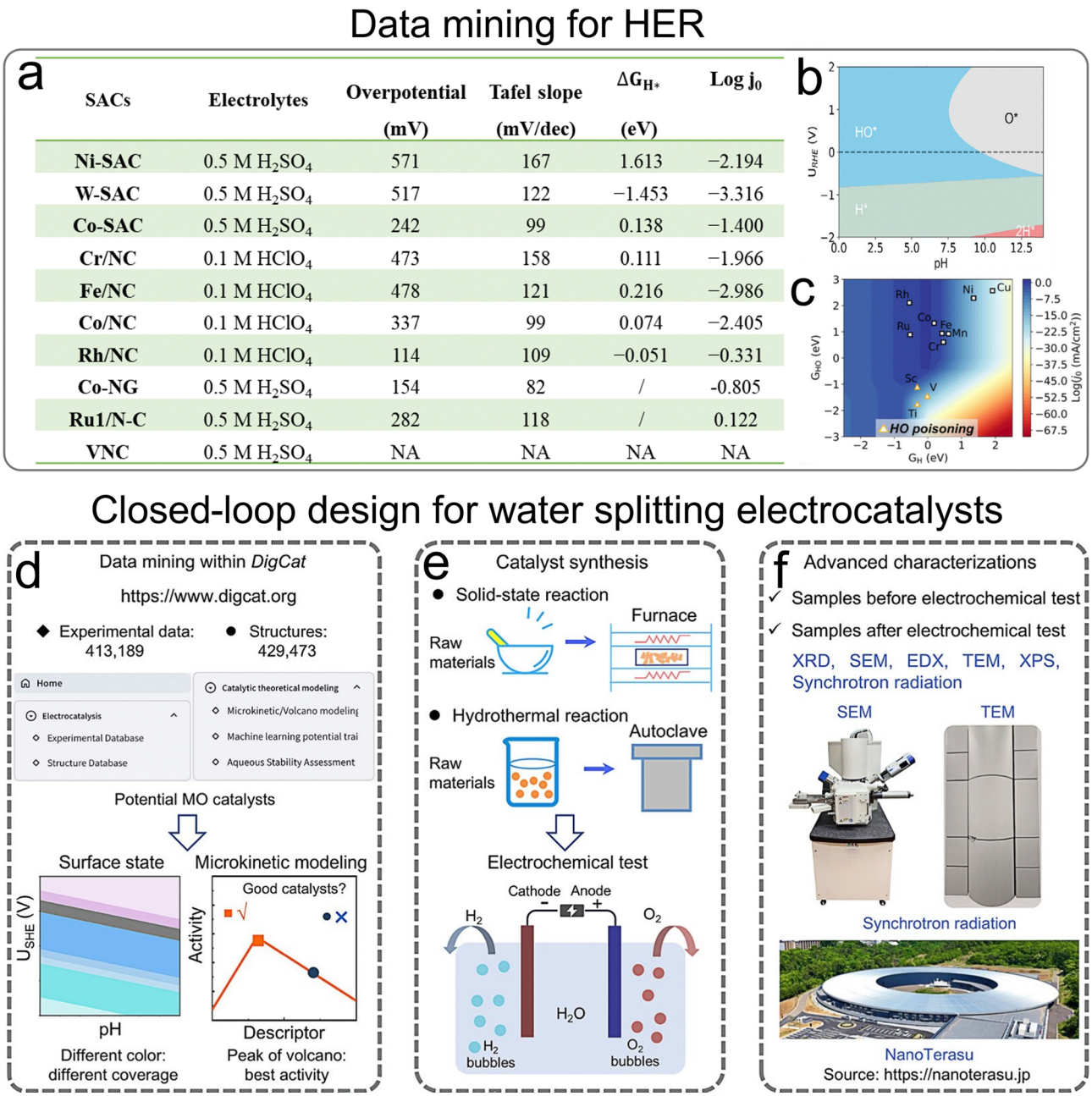

Fig. 4 (a) Comparison of HER activities of M–N–C SACs reported in the literature; (b) surface Pourbaix phase diagram of V-pyrrole-N_4 sites; (c) two-dimensional HER microkinetic volcano plot constructed with G_{H^*} and G_{OH^*} as variables under −0.5 V/RHE (pH = 0). Reproduced with permission from ref. 51 copyright 2025, Wiley-VCH. Closed-loop design for discovering non-noble MO electrocatalysts for water splitting, including data mining (d), catalyst synthesis (e), and advanced characterizations (f). Reproduced with permission from ref. 52 licensed under a Creative Commons License CC BY 4.0.

using advanced techniques. By leveraging large-scale data mining on the DigCat platform (which contains 413 189 experimental datasets and 429 473 computational data sets), combining data-driven methods such as surface state analysis and microkinetic modeling, and integrating human-intelligence-guided mechanistic understanding, the researchers successfully screened massive datasets and discovered for the first time a novel non-noble metal oxide bifunctional catalyst, $RbSbWO_6$. The selected catalyst exhibits excellent activity and stability for both OER and HER under acidic water splitting conditions. Notably, OER and HER occur on different facets of the stable surface, highlighting its unique dual-terminal catalytic feature. This achievement effectively integrates data-driven mining with human-intelligence-guided theoretical modeling, thereby effectively narrowing the catalyst search space and compensating for the limitations of conventional bulk stability analysis. Moreover, the closed-loop strategy of "data mining → theoretical screening → experimental validation → data

feedback'' provides a generalizable paradigm for the efficient and rational discovery of novel electrocatalysts, accelerating the transformation of catalysis research toward digitalization and intelligentization.

Overall, the above examples indicate that human-guided statistical analysis can contribute to catalytic knowledge discovery in different ways. Specifically, Sn-based CO_2RR catalysts and M–N–C catalysts for NO_3RR exemplify common pattern extraction, revealing general trends from dispersed experimental data. Cu-based SAAs, ORR DACs, and HER descriptor studies represent anomaly-driven discovery, where unexpected deviations from established descriptors or mechanisms lead to new theoretical insights. In comparison, the CER and acidic water-splitting studies mainly demonstrate research gap identification and design-direction mining, helping to locate underexplored catalyst systems and propose rational optimization strategies. Therefore, this section is not a simple accumulation of literature examples, but rather shows how existing data, when reanalyzed with human intelligence, can generate different levels of catalytic knowledge.

2.2. Developing new descriptors by building regression models using old data

Another important strategy for uncovering new catalytic knowledge from existing data is to construct novel descriptors based on regression models, thereby establishing quantitative structure–performance relationships. Taking symbolic regression based on genetic programming (GPSR) as an example (Fig. 5a), Weng *et al.*[57] started from existing experimental data, selecting electronic and structural parameters with clear physical meaning as input variables, such as the number of d electrons in transition metals (N_d), electronegativity at A/B sites (χ_A, χ_B), valence state (Q_A), ionic radius (R_A), and tolerance factor (t), and octahedral factor (μ) reflecting perovskite stability. The candidate mathematical expressions were iteratively optimized using a genetic algorithm, and the model performance was evaluated based on mean absolute error (MAE). From approximately 8640 candidate descriptors, the optimal relationship balancing predictive accuracy and physical interpretability was selected, enabling the systematic extraction of structure–performance relationships from existing data. On the other hand, Xin *et al.*[58] constructed an interpretable machine learning workflow (Fig. 5b). By organizing literature-reported ORR experimental data of perovskite oxides and extracting features such as electronegativity, ionic radius, Lewis acidity, and ionization energy, they analyzed key influencing factors using linear regression and neural network models. The results showed that the combination of lower A-site acidity and higher B-site acidity favors the formation of oxygen vacancies and reactive oxygen species, thereby promoting low-temperature ORR activity. Overall, the regression model can transform scattered experimental data into physically meaningful descriptors and design criteria, providing a key approach for systematically mining catalytic patterns from existing data and enabling data-driven catalyst design.

Electrochemical synthesis of H_2O_2 *via* $2e^-$ water oxidation ($2e^-$ WOR) is considered a low-cost, environmentally friendly, and sustainable pathway for converting electrical energy into chemical energy using water as the single source, showing significant application potential in renewable energy utilization and energy storage.[60,61] Against this backdrop, Liu *et al.*[59] systematically mined adsorption energy data from existing literature and databases, extracted the geometric and chemical characteristics of atomic centers, developed a weighted atom-centered symmetry function (wACSF) descriptor (Fig. 5c), and further established a high-precision XGBoost regression model, achieving rapid prediction of G_{HO^*} and G_{O^*}. Notably, this wACSF method can, for the first time, incorporate different types of catalysts (*e.g.*, metal alloys, metal oxides, SACs, and perovskites) into the same training set, which helps address the transferability challenge in AI for catalyst design. Based on this, combined with a microkinetic volcano model, they were able to efficiently screen catalysts for the $2e^-$ WOR, exhibiting both high activity and selectivity, thus achieving rapid transformation from old data to new material design. Overall, experimental verification using $LiScO_2$ as an example demonstrated $\sim$90% H_2O_2 Faradaic efficiency and excellent stability under mildly alkaline conditions, highlighting the reliability and practical value of regression models derived from literature data for efficient catalyst design.

In summary, regression models based on big data can extract key features from historical data and construct interpretable descriptors, establishing reliable structure–performance relationships. This strategy significantly improves catalyst screening efficiency, enabling rapid transformation from old data to new material design, and provides an important method for data-driven rational catalyst development. However, the reliability of such descriptors still strongly depends on data quality and feature selection, which may limit their transferability across different catalytic systems.

2.3. Discovering new knowledge using AI agents

With the continuous development of the digital materials ecosystem,[62] building data-driven AI agents for mining catalytic knowledge is gradually becoming an important approach to promoting the rational design of catalysts.[63] For example, Wang *et al.*,[37] used Cu-based SAAs as a model system, systematically mined large-scale experimental data from the DigCat database and combined it with a Catalysis AI Agent based on a LLM to systematically evaluate the influence of different doping elements on the selectivity of CO_2 electroreduction to generate multi-carbon products (C_{2+}). Fig. 6 outlines the overall research framework, including experimental data statistics, computational simulations, AI-assisted descriptor construction, and the catalyst design process. Based on the integration of experimental data and first-principles calculations, the researchers constructed a quantitative model of the relationship between energy descriptors and structure-selectivity, and further proposed a structure descriptor φ to establish a correlation between atomic-scale electronic structure characteristics, key intermediate adsorption behavior, and macroscopic product

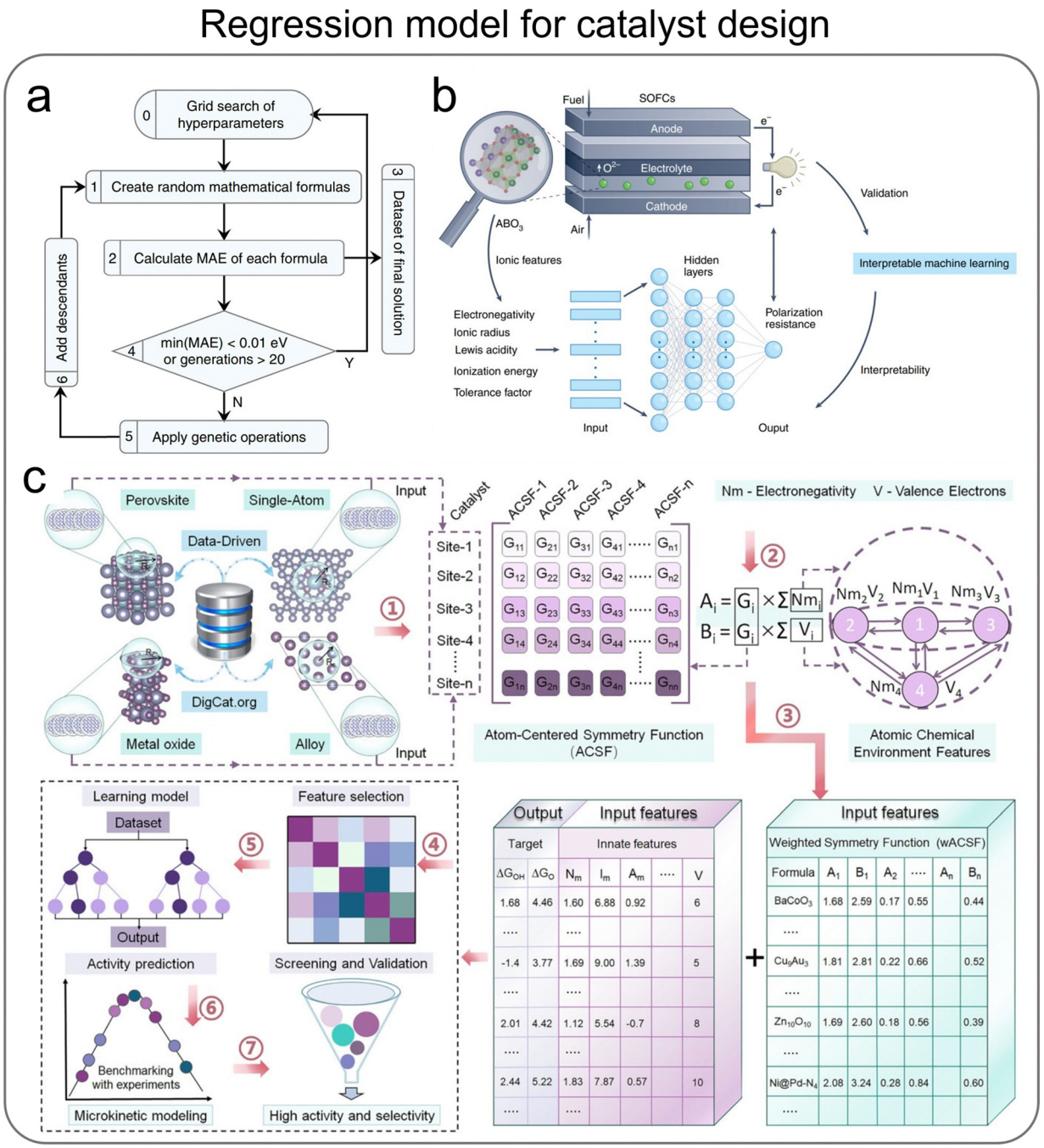

Fig. 5 (a) Schematic diagram of symbolic regression based on genetic programming. Reproduced with permission from ref. 57 licensed under a Creative Commons License CC BY 4.0. (b) Explainable machine learning framework for the discovery of catalyst materials for solid oxide fuel cells. Reproduced with permission from ref. 58 copyright 2022, Springer Nature. (c) Overview of the overall catalyst design process constructed for electrochemical hydrogen peroxide synthesis *via* $2e^-$ water oxidation ($2e^-$ WOR). Reproduced with permission from ref. 59 licensed under a Creative Commons License CC BY 4.0.

selectivity, achieving rapid prediction and efficient design of C_{2+} product selectivity. The research results not only reveal the periodic trends governing the effects of different dopants but also provide reliable guidance for the screening of novel Cu-based SAAs and dual single-atom alloys (DSAAs). Overall, this work demonstrates that combining knowledge mining of existing data with AI agents and descriptor modeling can significantly improve the efficiency and predictability of catalyst design, providing an important paradigm for data-driven catalysis research.

MOs have long been considered important electrocatalytic materials owing to their abundant resources, diverse structures, and tunable chemical properties. However, traditional experimental screening is inherently inefficient and struggles

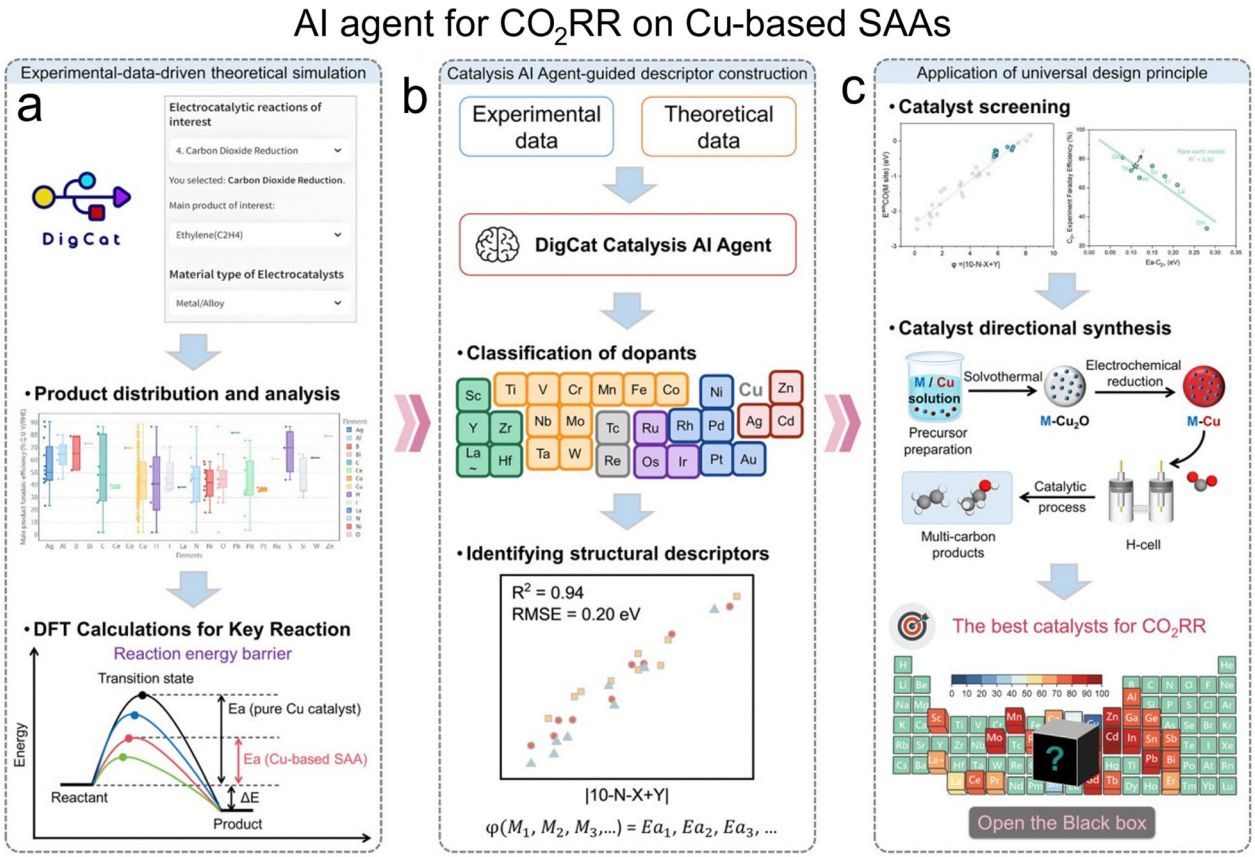

Fig. 6 A smart design process for Cu-based SAAs CO_2RR electrocatalysts for multi-carbon product generation: (a) experimental data-driven theoretical simulation; (b) descriptor construction assisted by Catalysis AI Agent; (c) application and verification of general design principles. Reproduced with permission from ref. 37 licensed under a Creative Commons License CC BY 4.0.

to support the rapid exploration of complex materials spaces. Databases such as the Materials Project and DigCat have accumulated extensive data on material structures, electrochemical stability, and experimental performance metrics. However, accessing and analyzing these datasets typically require expertise in programming and computational methods, which limits their accessibility to the broader research community. In recent years, AI agents combining LLMs and automated workflows have offered a new solution to this problem. For example, as shown in Fig. 7a, the StableOx-Cat AI agent automates the material screening process through natural language interaction, integrating functions such as thermodynamic and electrochemical stability analysis, elemental filtering, and custom evaluation, transforming complex calculation processes into an intuitive and easy-to-use research interface. Overall, the AI agent significantly lowers the barrier to data utilization, improves the efficiency of MOs stability assessment, and provides a reproducible, systematic data-driven workflow for mining potential novel catalysts based on existing data.

The design of multimetallic catalysts faces an explosive increase in the combinations of composition, structure, and process parameters, making traditional Bayesian optimization strategies, which rely on single data streams, ineffective in handling complex experimental spaces. The Copilot for Real-world Experimental Scientists (CRESt) system (Fig. 7b), developed by Zhang *et al.*,[67] integrates a multimodal large visual language model (LVLM), knowledge-assisted Bayesian optimization, and a robotic automated experimental platform to enable closed-loop high-throughput exploration from natural language interaction to catalyst synthesis, characterization, and electrochemical testing. In a direct formate fuel cell system, CRESt screened over 900 formulations and completed approximately 3500 tests within several months, successfully discovering an octagonal multimetallic catalyst with excellent performance and significantly reduced noble metal loading. This system achieves a dynamic balance between exploration and utilization by fusing literature, images, and experimental data embedding, thereby significantly improving material discovery efficiency and result reproducibility. Overall, multimodal and agent-driven self-powered laboratories provide an efficient and scalable technical path for mining new catalytic knowledge from existing data, offering a new research paradigm for the accelerated design of complex catalytic systems.

In catalysis research, the large volume and structural complexity of literature data make it difficult for traditional trial-and-error approaches and rule-based text mining methods to effectively uncover the intrinsic relationships among catalyst composition, structure, reaction conditions, and performance.

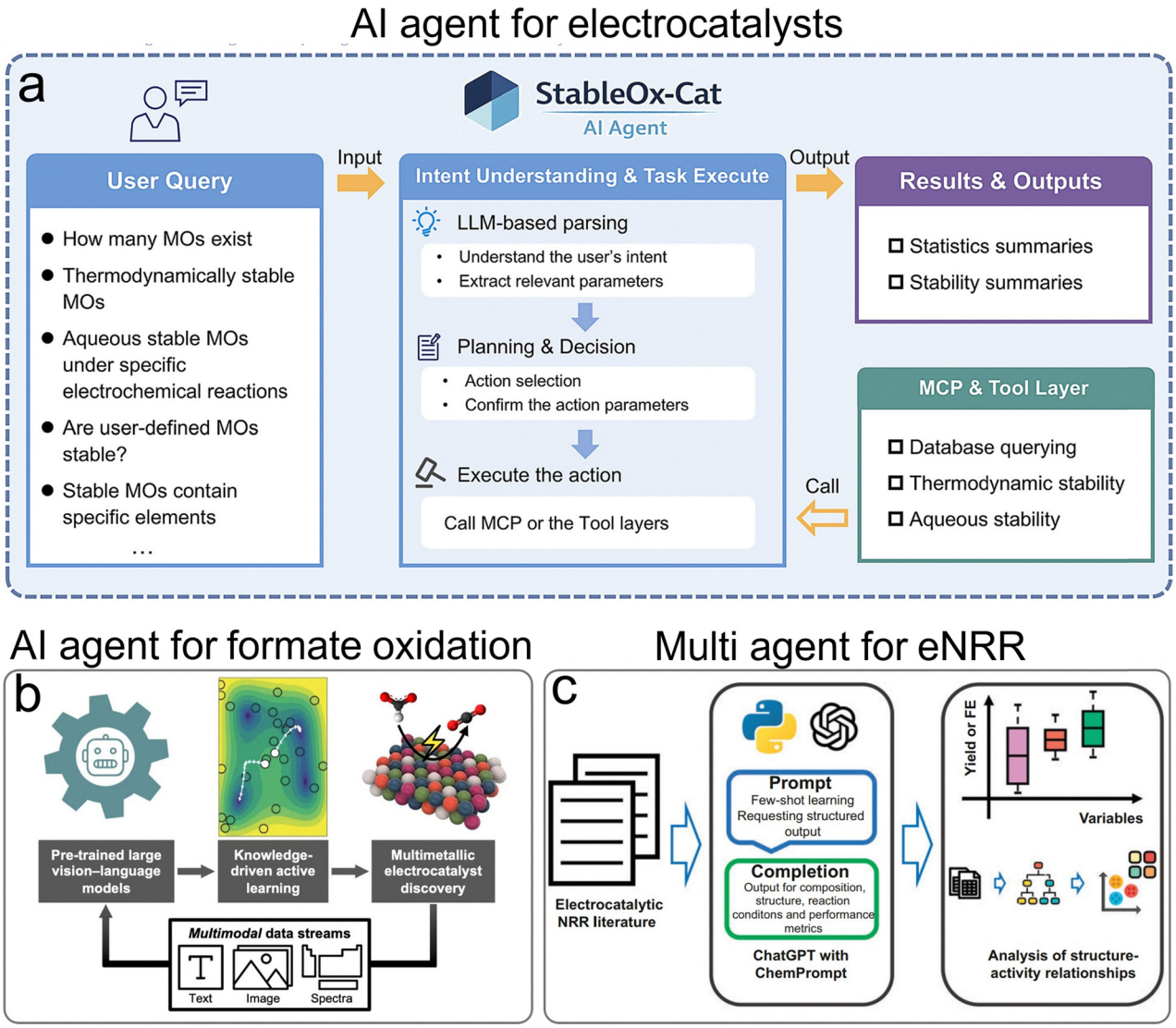

Fig. 7 (a) Schematic diagram of the overall architecture of the StableOx-Cat AI agent, where MOs represents metal oxides and MCP represents the Model Context Protocol. Reproduced with permission from ref. 64 licensed under a Creative Commons License CC BY 4.0. (b) Schematic diagram of the CRESt system, which is based on a large visual language model (LVLM)-driven agent framework. Reproduced with permission from ref. 65 licensed under a Creative Commons License CC BY 4.0. (c) Schematic diagram of the process of predicting and analyzing the structure–performance relationship of electrochemical nitrogen reduction reaction (eNRR) literature based on LLMs enhanced by ChemPrompt. Reproduced with permission from ref. 66 licensed under a Creative Commons License CC BY 4.0.

With the development of LLMs and AI agents, researchers can now achieve high-throughput data parsing and automated extraction of catalytic knowledge. For example, as shown in Fig. 7c, the eNRRCrew framework,[66] by integrating ChemPrompt-enhanced LLMs, knowledge graphs, and machine learning algorithms, extracts catalyst characteristics and performance information from 2321 literature abstracts and utilizes an agent to conduct structured analysis and performance prediction. Furthermore, such agents can not only parse existing data but also proactively recommend potentially efficient catalyst combinations based on learned structure–performance relationships, thereby improving catalyst screening efficiency. Overall, AI-driven data mining and agent analysis provide an efficient path for transforming scattered historical knowledge into actionable design rules, significantly accelerating the discovery and optimization of novel electrocatalytic materials.

In summary, AI agents can integrate literature, databases, and computational data to achieve automated catalytic-knowledge extraction and design optimization, significantly improving the efficiency of exploring complex materials spaces. This strategy promotes the transformation of catalysis research from an experience-driven to data-driven and self-driven modes, accelerating the discovery and design of novel electrocatalytic materials. Despite these advantages, the reliability of AI-agent-based discovery remains dependent on data quality and may be affected by model hallucination and bias.

3. Summary and outlook

Fig. 8 summarizes the main message of this perspective. The rapid development of digital and intelligent research paradigms is reshaping catalysis research by enabling the rediscovery of knowledge from accumulated literature, experimental records, and computational datasets. Reviewing current research paths for mining new catalysis knowledge from old data, three representative strategies have gradually emerged: First, knowledge induction based on human intelligence, establishing structure–performance relationships through theoretical understanding and empirical summarization;[4,68] second, statistical learning methods based on regression models and descriptor engineering to achieve quantitative prediction of catalytic activity and construction of design criteria;[59] and third, automated data mining and knowledge reasoning frameworks based on AI agents, integrating literature, databases, and computational results to achieve high-throughput knowledge extraction and candidate material recommendation.[37,69] Notably, although AI has made remarkable progress in energy materials research, literature-mining workflows based on AI tools remain nascent.[70] Zhang *et al.*[71,72] tested non-fine-tuned general LLMs (*e.g.*, Gemini 2.5 Flash, LLaMA-4) to extract data from solid-state hydrogen storage materials literature, and identified a high hallucination rate. By contrast, their Descriptive Interpretation of Visual Expression (DIVE) workflow, which uses multi-agent coordination and embedding-based alignment evaluation, automatically penalizes hallucinations and boosts extraction accuracy by 10–30%. This highlights that AI models with targeted fine-tuning or structured constraints can effectively safeguard the reliability of scientific data mining. Overall, these three strategies, at different levels, have promoted the structured expression of catalytic knowledge and significantly improved the efficiency of extracting design patterns from historical data.

However, the increasing use of literature-derived data also requires a more critical perspective. First, published catalytic data are not equivalent to the complete materials space because they are inevitably affected by publication bias. Catalysts with attractive activity, selectivity, or stability are more likely to be reported, whereas failed experiments, poorly performing materials, and negative results are rarely included in the literature. Therefore, statistical trends extracted from published studies should be interpreted as patterns within the reported knowledge space rather than universal laws of the entire chemical space. Second, catalytic performances collected from different publications are often measured under non-uniform conditions, including differences in electrolyte composition, pH, reactant concentration, catalyst loading, electrode configuration, potential calibration, product quantification, and stability testing protocols. Such heterogeneity may introduce hidden variables into data mining and machine learning models, making it necessary to combine statistical correlations with mechanistic analysis, uncertainty evaluation, and experimental validation.

Looking ahead, the deep integration of AI and theoretical computation will be an important direction for catalysis research. Several specific future directions deserve attention. First, universal machine learning interatomic potentials (MLIPs) should be systematically developed and benchmarked across realistic catalytic environments, such as solid–liquid interfaces, dynamic adsorbate coverages, surface defects, and catalyst reconstruction under operating conditions. Their combination with multi-scale simulations can help connect atomic-level mechanisms with macroscopic catalytic performance. Second, multimodal LLMs[73,74] may integrate

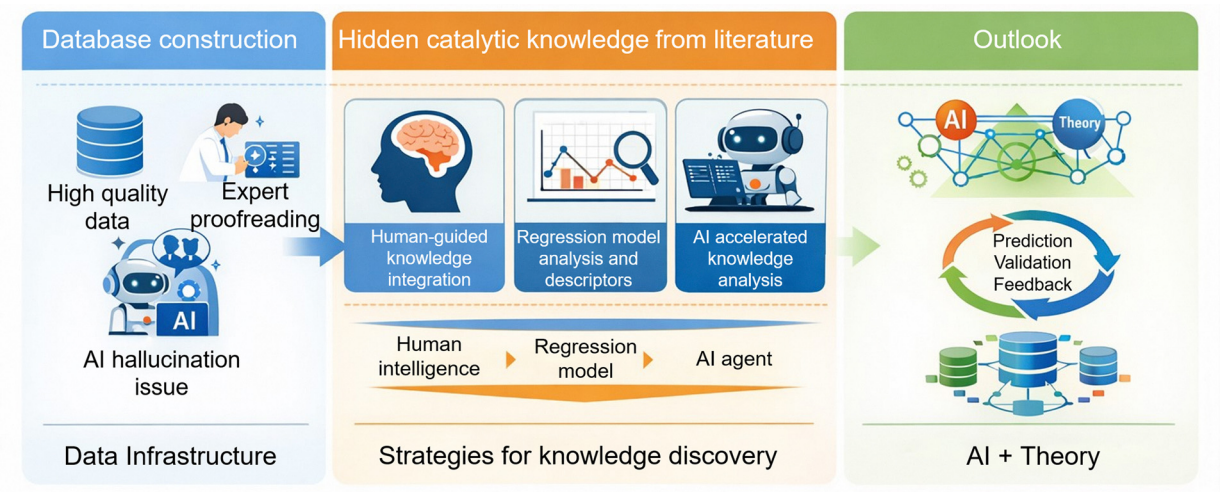

Fig. 8 A schematic diagram of the digital ecosystem for data-driven catalyst design. The left side emphasizes the importance of high-quality databases as digital infrastructure and points out the key role of expert verification in reducing the AI hallucination; the middle summarizes three strategies for mining catalytic knowledge based on old data: human intelligent, regression modeling, and AI agent knowledge mining; the right side shows the development direction of the integration of AI and theory, promoting the construction of the digital materials ecosystem through universal machine learning interatomic potentials (MLIPs), LLMs, and the "prediction-validation-feedback" closed-loop, and achieving efficient and predictable catalyst design.

heterogeneous catalytic information, including literature, spectra, microscopy images, computational data, and experimental protocols, thereby assisting automated data extraction, mechanism hypothesis generation, and cross-modal consistency checking. Third, closed-loop platforms combining digital databases, autonomous synthesis, high-throughput characterization, and catalytic testing should be established to realize a self-iterative ''prediction–verification–feedback'' workflow. Active learning can further guide experiment selection and continuously refine predictive models. Fourth, a cross-scale and cross-system digital materials ecosystem[70,75,76] should emphasize standardized metadata, shared benchmark datasets, and unified evaluation protocols, enabling knowledge transfer across different reactions and material families. Overall, these efforts will promote catalysis research toward a more reliable, interpretable, and closed-loop paradigm for efficient and predictable catalyst design.

Conflicts of interest

The authors declare no competing financial interest.

Data availability

No new data is generated for this article.

Acknowledgements

The authors acknowledge the support from JSPS KAKENHI (No. JP25H01508 and JP25K01737).

References

1 Y. Chen, M. Ross Kunz, X. He and R. Fushimi, *Curr. Opin. Chem. Eng.*, 2022, **37**, 100843.
2 A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner and G. Ceder, *APL Mater.*, 2013, **1**, 011002.
3 S. Curtarolo, W. Setyawan, G. L. Hart, M. Jahnatek, R. V. Chepulskii, R. H. Taylor, S. Wang, J. Xue, K. Yang and O. Levy, *Comput. Mater. Sci.*, 2012, **58**, 218–226.
4 D. Zhang, X. Jia, H. Liu, Y. Wang, S. Ye, Q. Jiang, Y. Wang, Z. Guo, L. Zhang, L. Wei, W. Yang, H. Liu, S. Zhao, H. Xu, D. Cheng, Y. Hashimoto, T. Tomai and H. Li, *AI Agent*, 2025, **1**, 1.
5 D. Zhang, Z. Bao, Y. Chu, Z. Guo, X. Jia, Q. Jiang, H. Liu, T. Liu, T. Lu, Y. Lu, D. Devang Shah, Y. Wang, Y. Wang, Y. Wang, S. Ye, S. Ying, Z. Yu, L. Zhang, S. Zhao and H. Li, *ChemRxiv*, 2026, preprint, DOI: 10.26434/chemrxiv-2024-9lpb9.
6 K. T. Winther, M. J. Hoffmann, J. R. Boes, O. Mamun, M. Bajdich and T. Bligaard, *Sci. Data*, 2019, **6**, 75.
7 S. J. Sahoo, D. S. Levine, Z. Ulissi, C. L. Zitnick, J. B. Varley, J. A. Gauthier, N. Govindarajan and M. Shuaibi, *arXiv*, 2025, preprint, arxiv:2509.17862, DOI: 10.48550/arxiv.2509.17862.
8 Z.-J. Zhao, S. Liu, S. Zha, D. Cheng, F. Studt, G. Henkelman and J. Gong, *Nat. Rev. Mater.*, 2019, **4**, 792–804.
9 Y. Zhuang, X. Yang, C. Zhang, X. Jia, D. Zhang, M. Li, T. Yao, J. Peng, Z. Gao, W. Yang and H. Li, *Precis. Chem.*, 2026, DOI: 10.1021/prechem.5c00449.
10 J. Benavides-Hernández and F. Dumeignil, *ACS Catal.*, 2024, **14**, 11749–11779.
11 C. Tong, Q. Liu, H. Wang and M. Yao, *Environ. Sci.: Nano*, 2026, **13**, 1830–1852.
12 Y. Wang, Q. Wang, S.-H. Jang, E. J. Cheng and H. Li, *Chem. Commun.*, 2026, DOI: 10.1039/D6CC01716A.
13 S. W. Ying, Y. Wang, P. Du, Q. Wang, C. Yue, D. Zhang, Z. C. Chen, J. W. Zheng, S. Y. Xie and H. Li, *Angew. Chem., Int. Ed.*, 2025, **64**, e202511924.
14 Y. Wang, Z. Wu, Y. Jiang, D. Zhang, Q. Wang, C. Wang, H. Li, X. Jia, J. Fan and H. Li, *Adv. Funct. Mater.*, 2025, e06314.
15 Y. Wang, N. Ma, Y. Zhang, B. Liang and J. Fan, *Appl. Surf. Sci.*, 2023, 157126.
16 Y. Chu, Y. Wang, D. Zhang, X. Song, C. Yu and H. Li, *J. Chem. Phys.*, 2025, **162**, 17.
17 S. Zhao, Y. Wang, H. Liu, X. Jia, Y. Zhang, L. Zhang, B. Da, Q. Wang, H. Zheng, H. Li and W. Li, *J. Catal.*, 2026, **453**, 116502.
18 R. Ren, Y. Wang, B. Li, H. Fan and H. Li, *J. Phys. Chem. Lett.*, 2026, **17**, 5583–5590.
19 D. Zhang, Z. Wang, F. Liu, P. Yi, L. Peng, Y. Chen, L. Wei and H. Li, *J. Am. Chem. Soc.*, 2024, **146**, 3210–3219.
20 N. Ma, C. Leung, Y. Wang, Y. Zhang, S. Luo, H. Liu, B. Liang, C. Huang, Z. Wei, Y. Ren and J. Fan, *Small Methods*, 2025, **9**, 2500310.
21 N. Ma, Y. Zhang, Y. Wang, C. Huang, J. Zhao, B. Liang and J. Fan, *Appl. Surf. Sci.*, 2023, **628**, 157225.
22 D. Qi, H. Zhang, K. Su, W. Li, Y. Yuan, Y. Xiao and J. Xu, *ChemSusChem*, 2025, **18**, e202401267.
23 Y. Zhao, H. Wu, Y. Wang, L. Liu, W. Qin, S. Liu, J. Liu, Y. Qin, D. Zhang, A. Chu, B. Jia, X. Qu and M. Qin, *Energy Storage Mater.*, 2022, **50**, 186–195.
24 Y. Zhao, Z. Zhang, L. Liu, Y. Wang, T. Wu, W. Qin, S. Liu, B. Jia, H. Wu, D. Zhang, X. Qu, G. Qi, E. P. Giannelis, M. Qin and S. Guo, *J. Am. Chem. Soc.*, 2022, **144**, 20571–20581.
25 Y. Wang, B. Jia, W. Qin, Y. Wang, S. Liu, Y. Qin, Y. Zhao, L. Liu, D. Zhang, H. Liu, H. Zhong, J. Liu, J. Tu, Y. Liu, H. Wu, D. Zhang, J. Fan, X. Qu, H. Li and M. Qin, *J. Am. Chem. Soc.*, 2025, **147**, 32249–32262.
26 Y. Wang, Y. Qin, S. Liu, Y. Zhao, L. Liu, D. Zhang, S. Zhao, J. Liu, J. Wang, Y. Liu, H. Wu, B. Jia, X. Qu, H. Li and M. Qin, *J. Am. Chem. Soc.*, 2025, **147**, 13345–13355.
27 Y. Wang, Y. Zhao, L. Liu, W. Qin, S. Liu, J. Tu, Y. Liu, Y. Qin, J. Liu, H. Wu, D. Zhang, A. Chu, B. Jia, X. Qu, M. Qin and J. Xue, *J. Am. Chem. Soc.*, 2023, **145**, 20261–20272.
28 Y. Wang, Y. Zhao, L. Liu, W. Qin, S. Liu, J. Tu, Y. Qin, J. Liu, H. Wu, D. Zhang, A. Chu, B. Jia, X. Qu and M. Qin, *Adv. Mater.*, 2022, **34**, 2200088.
29 J. Du, Y. Yan, X. Li, J. Chen, C. Guo, Y. Chen and H. Wang, *Chem. Sci.*, 2025, **16**, 9424–9435.

30 L. Liu, Y. Wang, Y. Zhao, Y. Wang, Z. Zhang, T. Wu, W. Qin, S. Liu, B. Jia, H. Wu, D. Zhang, X. Qu, M. Chhowalla and M. Qin, *Adv. Funct. Mater.*, 2022, **32**, 2112207.
31 S. Zhen, Y. Wang, Y. Wang, J. Wang, X. Niu, K. Li, D. Su, H. Duan, B. Jia, M. Qin and L. Zhang, *eScience*, 2026, **6**, 100484.
32 H. Jing, J. Long, H. Li, X. Fu and J. Xiao, *Chin. J. Catal.*, 2023, **48**, 205–213.
33 J. Yang, W.-H. Li, H.-T. Tang, Y.-M. Pan, D. Wang and Y. Li, *Nature*, 2023, **617**, 519–523.
34 Y. Wang, Y. Zhang, N. Ma, J. Zhao, Y. Xiong, S. Luo and J. Fan, *Surf. Interfaces*, 2024, **50**, 104498.
35 Z. Guo, Y. Yu, C. Li, E. Campos Dos Santos, T. Wang, H. Li, J. Xu, C. Liu and H. Li, *Angew. Chem., Int. Ed.*, 2024, e202319913.
36 Y. Wang, D. Zhang, B. Sun, X. Jia, L. Zhang, H. Cheng, J. Fan and H. Li, *Angew. Chem., Int. Ed.*, 2024, e202418228.
37 X. Wang, Z. Li, D. Zhang, H. Li, H. Xu and D. Cheng, *Angew. Chem., Int. Ed.*, 2026, e24612.
38 W. Yang, Z. Jia, B. Zhou, L. Chen, X. Ding, L. Jiao, H. Zheng, Z. Gao, Q. Wang and H. Li, *ACS Catal.*, 2023, **13**, 9695–9705.
39 Q. Jiang, M. Gu, S. Pei, T. Wang, F. Liu, X. Yang, D. Zhang, Z. Wu, Y. Wang, L. Wei and H. Li, *J. Am. Chem. Soc.*, 2025, **147**, 26029–26039.
40 X. You, Z. Guo, Q. Jiang, J. Xia, S. Wang, X. Yang, Z. Zhuang, Y. Li, H. Xiang, H. Li and B. Yu, *Nano Lett.*, 2025, **25**, 8704–8712.
41 A. Cao, V. J. Bukas, V. Shadravan, Z. Wang, H. Li, J. Kibsgaard, I. Chorkendorff and J. K. Norskov, *Nat. Commun.*, 2022, **13**, 2382.
42 D. Szmigiel, H. Bielawa, M. Kurtz, O. Hinrichsen, M. Muhler, W. Raróg, S. Jodzis, Z. Kowalczyk, L. Znak and J. Zieliński, *J. Catal.*, 2002, **205**, 205–212.
43 S. Luo, N. Ma, J. Zhao, Y. Wang, Y. Zhang, Y. Xiong and J. Fan, *J. Mater. Sci. Technol.*, 2024, **199**, 145–155.
44 N. Ma, H. Liu, L. Yu, Q. Yu, J. Cheng, Y. Ren, J. Fan and Z. Wei, *Small*, 2026, **22**, e13102.
45 N. Ma, Y. Zhang, Y. Wang, J. Zhao, B. Liang, Y. Xiong, S. Luo, C. Huang and J. Fan, *J. Colloid Interface Sci.*, 2024, **654**, 1458–1468.
46 D. Zhang, F. She, J. Chen, L. Wei and H. Li, *J. Am. Chem. Soc.*, 2025, **147**, 6076–6086.
47 H. Li, S. Kelly, D. Guevarra, Z. Wang, Y. Wang, J. A. Haber, M. Anand, G. T. K. K. Gunasooriya, C. S. Abraham, S. Vijay, J. M. Gregoire and J. K. Nørskov, *Nat. Catal.*, 2021, **4**, 463–468.
48 J. Liu, H. Li, H. Xu and D. Cheng, *Angew. Chem., Int. Ed.*, 2026, e8386838, DOI: 10.1002/anie.8386838.
49 K. Ma, L. Yang, X. Yang, C. Zhang, D. Zhang, L. Zhang, L. Wei, C. Ye, H. Li and W. Yang, *Adv. Funct. Mater.*, 2025, 36.
50 Y. Zhang, Y. Wang, N. Ma, B. Liang, Y. Xiong and J. Fan, *Small*, 2023, e2306840.
51 S. Ye, F. Liu, F. She, J. Chen, D. Zhang, A. Kumatani, H. Shiku, L. Wei and H. Li, *Angew. Chem., Int. Ed.*, 2025, **64**, e202425402.
52 X. Jia, Z. Zhou, F. Liu, T. Wang, Y. Wang, D. Zhang, H. Liu, Y. Wang, S. Ye, K. Amezawa, L. Wei and H. Li, *J. Am. Chem. Soc.*, 2025, **147**, 22642–22654.
53 Y. Wang, Z. Yuan, Z. Cen, S. Yu, C. Li, H. Tang, A. Cao, T. Wu, X. Ren and D. Ma, *Angew. Chem., Int. Ed.*, 2026, e2500154.
54 Y. Wang, S. Liu, Y. Qin, Y. Zhao, L. Liu, D. Zhang, J. Liu, Y. Liu, A. Chu, H. Wu, B. Jia, X. Qu, H. Li and M. Qin, *ACS Catal.*, 2024, **14**, 13759–13767.
55 X. Jiang, Y. Wang, B. Jia, X. Qu and M. Qin, *ACS Appl. Mater. Interfaces*, 2022, **14**, 41141–41148.
56 S. Liu, B. Jia, Y. Wang, Y. Zhao, L. Liu, F. Fan, Y. Qin, J. Liu, Y. Jiang, H. Liu, H. Zhao, H. Li, W. Zhou, H. Wu, D. Zhang, X. Qu and M. Qin, *Adv. Mater.*, 2024, **36**, 2409530.
57 B. Weng, Z. Song, R. Zhu, Q. Yan, Q. Sun, C. G. Grice, Y. Yan and W. J. Yin, *Nat. Commun.*, 2020, **11**, 3513.
58 H. Xin, *Nat. Energy*, 2022, **7**, 790–791.
59 Z. Liu, Y. Liu, Y. Zhang, Y. Deng, Z. Zheng, R. Knibbe, T. Gao, M. Li, Z. Wang, B. Zhang, X. Jia, D. Zhang, H. Liu, X. Shao, Z. Gao, L. Wei, H. Li and W. Yang, *Angew. Chem., Int. Ed.*, 2026, **65**, e18027.
60 X. Shi, S. Siahrostami, G. L. Li, Y. Zhang, P. Chakthranont, F. Studt, T. F. Jaramillo, X. Zheng and J. K. Norskov, *Nat. Commun.*, 2017, **8**, 701.
61 X. Shi, S. Back, T. M. Gill, S. Siahrostami and X. Zheng, *Chem*, 2021, **7**, 38–63.
62 D. Zhang, X. Jia, Y. Wang, H. Liu, Q. Wang, S.-H. Jang, D. Shah, S. Ye, H. B. Tran and H. Li, *Chem. Sci.*, 2026, **17**, 5782–5804.
63 P. Mopgar, *ChemRxiv*, 2026, preprint, DOI: 10.26434/chemrxiv.15000226/v2.
64 D. Z. Xue Jia, Y. Lu, Q. Wang and H. Li, *AI Agent*, 2026, **2**, 1.
65 J. Peng, C. Liu, Y. Luo and K. Dandapat, *AI Agent*, 2025, **1**, 1.
66 X. Hu, S. Chen, L. Chen, H. Wang, X. Zhang and Z. Zhou, *Natl. Sci. Rev.*, 2025, **12**, nwaf372.
67 Z. Zhang, Z. Ren, C. W. Hsu, W. Chen, Z. W. Hong, C. F. Lee, A. Penn, H. Xu, D. J. Zheng, S. Miao, Y. Huang, Y. Gao, W. Chen, H. Smith, Y. Niu, Y. Tian, Y. R. Lu, Y. C. Shao, S. Li, H. T. Wang, I. I. Abate, P. Agrawal, Y. Shao-Horn and J. Li, *Nature*, 2025, **647**, 390–396.
68 F. Yang, E. Campos dos Santos, X. Jia, R. Sato, K. Kisu, Y. Hashimoto, S.-I. Orimo and H. Li, *Nano Mater. Sci*, 2024, **6**, 256–262.
69 D. Zhang, Y. Chen, C. Liu, Y. Liu, H. Xin, J. Peng, P. Ou and H. Li, *Angew. Chem., Int. Ed.*, 2026, e26150.
70 X. Jiang, W. Wang, S. Tian, H. Wang, T. Lookman and Y. Su, *npj Comput. Mater.*, 2025, **11**, 79.
71 D. Zhang, X. Jia, H. B. Tran, S. H. Jang, L. Zhang, R. Sato, Y. Hashimoto, T. Sato, K. Konno, S. I. Orimo and H. Li, *Chem. Sci.*, 2026, **17**, 3031–3042.
72 Y. Hong and X. Mao, *AI Agent*, 2026, **2**, 2.
73 T. Yao, J. Huang, Y. Yan, Y. Yang, Z. Wang, X. Shao, Z. Gao and W. Yang, *AI Agent*, 2025, **1**, 9.
74 S. Yang, Z. Yang, Y. Liu and H. Wang, *AI Agent*, 2025, **1**, 4.
75 D. Zhang, X. Jia, H. B. Tran, F. Yang, Q. Wang, H. Liu, Y. Qi, E. J. Cheng and H. Li, *ChemRxiv*, 2024, preprint, DOI: 10.26434/chemrxiv-2024-p9fvc.
76 A. R. Mishra, J. Pascasio, J. Yang and W.-L. Li, *AI Agent*, 2026, **2**, 6.