

The artificial intelligence-catalyst pipeline: accelerating catalyst innovation from laboratory to industry

Aoming Li¹, Peng Cui², Xu Wang², Adrian Fisher³, Lanyu Li (✉)¹, Daojian Cheng (✉)¹

¹ State Key Laboratory of Organic-Inorganic Composites and Beijing Advanced Innovation Center for Soft Matter Science and Engineering, Beijing University of Chemical Technology, Beijing 100029, China

² Petrochemical Research Institute, PetroChina, Beijing 102206, China

³ Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge CB3 0AS, UK

© Higher Education Press 2025

Abstract The integration of high-throughput experimental technologies with artificial intelligence is transforming catalyst research and development. This study explores the synergistic convergence of artificial intelligence and high-throughput experimentation in chemical catalysis, highlighting both current and emerging experimental techniques. It examines how AI-driven methodologies enhance data analysis, automate complex decision-making processes, and optimize catalyst design for industrial applications. The future of research laboratories is envisioned as autonomous, self-driven environments that streamline and accelerate the transition from conceptualization to practical implementation. Key challenges, including data quality, model interpretability, and the scalability of industrial applications, are critically analyzed. Future research should focus on addressing these challenges through strategic methodologies, establishing a systematic framework to fully harness the potential of artificial intelligence and high-throughput experimentation. These advancements will enhance research efficiency and drive innovation in catalysis.

Keywords high-throughput experimentation, artificial intelligence, chemical catalysis, self-driving labs, industrial application

1 Introduction

The performance of catalysts is governed by the interplay of multiple factors, including process conditions, preparation methods, and multi-component formulations.

Received December 20, 2024; accepted March 29, 2025;
online May 16, 2025

E-mails: lilanyu@buct.edu.cn (Li L.),
chengdj@mail.buct.edu.cn (Cheng D.)

Identifying an optimal catalyst formulation typically requires extensive experimental validation, thereby increasing demand for high-throughput synthesis technologies. The concept of high-throughput screening originates from combinatorial chemistry and has been extensively applied in various fields, including pharmaceuticals [1], materials science [2], and catalysis [3,4]. In catalysis, combinatorial techniques involve the batch synthesis of catalyst libraries, followed by high-throughput screening to identify promising candidates. Through iterative experimentation, catalyst composition and synthesis parameters are systematically refined to achieve optimal performance for a specific reaction [5]. High-throughput experimental (HTE) is increasingly recognized as a powerful methodology for accelerating catalyst development.

Simultaneously, advancements in computational chemistry and artificial intelligence (AI) have introduced transformative innovations in chemical engineering, giving rise to intelligent methodologies that enhance experimental research. These developments are reshaping traditional experimental frameworks, significantly improving efficiency and accuracy. The rapid progress of AI is driving a paradigm shift in scientific discovery, particularly in the natural sciences. AI now plays a pivotal role in advancing scientific inquiry, enhancing, accelerating, and expanding our understanding of natural phenomena across diverse spatial and temporal scales. This has led to the emergence of a novel research domain known as AI for Science [6].

The key milestones in AI development and its application in chemical catalysis research are illustrated in Fig. 1. The origins of AI trace back to the 1950s, marked by the introduction of the Turing test, the inception of neural networks, and early advancements in natural language processing. In recent years, the emergence of data-driven methodologies, machine

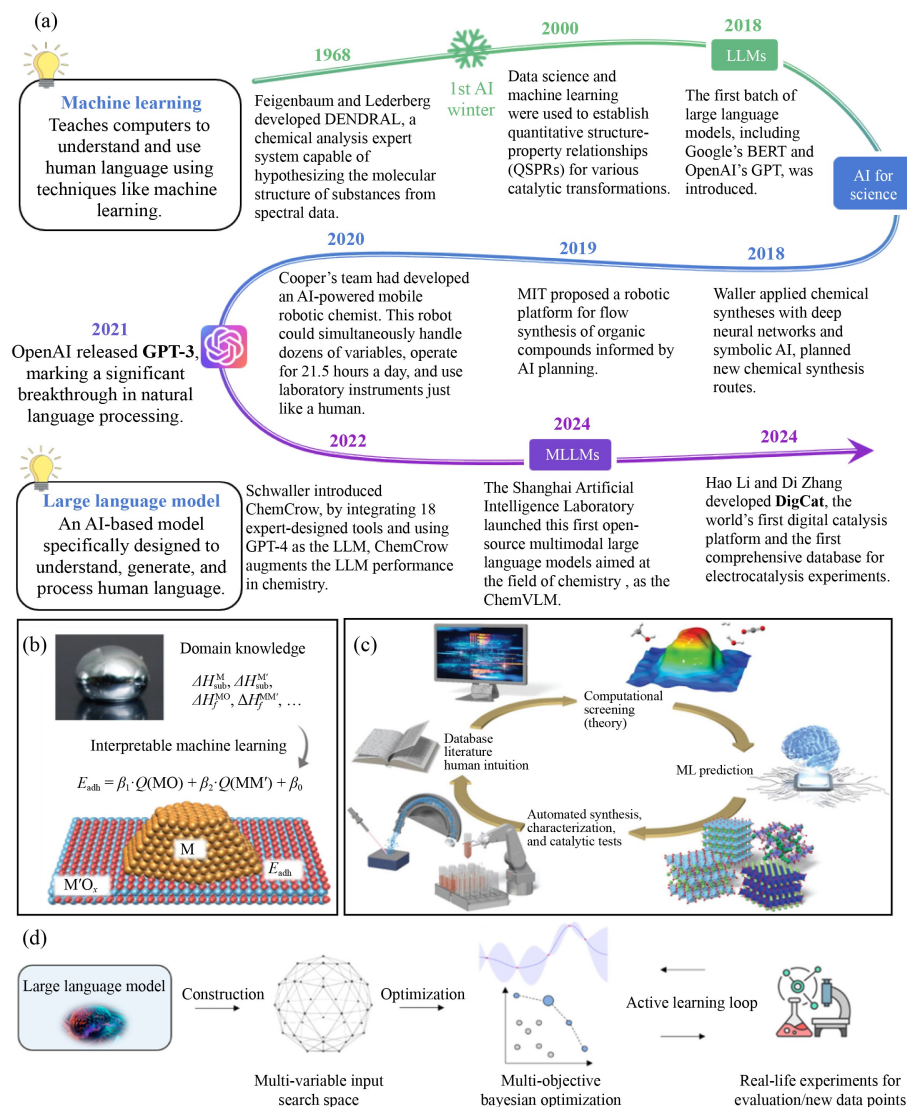


Fig. 1 (a) The key milestones of AI development and its application in chemical catalyst research. (b) Derivable formulas identified through interpretable machine learning. Reprinted with permission from Ref. [7], copyright 2024, American Association for the Advancement of Science. (c) Schematic representation of ML-aided catalysis research. Reprinted with permission from Ref. [8], copyright 2019, American Chemical Society. (d) LLM-integrated workflow for catalyst design and optimization. Reprinted with permission from Ref. [9], copyright 2023, American Chemical Society.

learning (ML), and large language models (LLMs) has significantly expanded AI applications in chemical research and development (R&D), including catalysis. Notable examples include the integration of ML with domain expertise and computational chemistry for mechanistic interpretation [7], the development of closed-loop intelligent catalyst synthesis [8], and the synergy of LLMs and active learning for catalyst discovery [9].

Traditional catalyst development relies on empirical trial-and-error approaches, necessitating extensive experimentation to screen materials and optimize reaction conditions. This process is often time-consuming, spanning several years, and incurs substantial costs. Additionally, data fragmentation and scale-up challenges represent major obstacles to the efficient development and successful commercialization of new catalysts.

Industrial catalytic processes involve the interplay of multiple variables, including temperature, pressure, and feedstock ratios. However, relevant data are typically dispersed across laboratories, production facilities, and published literature, making systematic utilization difficult. Furthermore, catalysts that demonstrate high performance under laboratory conditions frequently fail during scale-up due to discrepancies in mass transfer and thermodynamic behavior. The complexity of catalyst deactivation mechanisms and the necessity for long-term stability testing further constrain the advancement and practical implementation of catalytic systems.

This perspective underscores the transformative potential of integrating AI with HTE in catalysis, highlighting how this synergy can overcome traditional limitations and accelerate the development and

commercialization of advanced catalysts. By leveraging the rapid screening capabilities of HTE alongside the predictive power of AI, this integration not only expedites the discovery of novel materials but also optimizes industrial processes and establishes new paradigms in scientific research and industrial production. The following sections provide an in-depth analysis of the specific strategies and applications of this technological convergence, examining its role in shaping the future of catalysis and materials science.

2 High-throughput technology and equipment innovation

High-throughput methodologies facilitate the rapid generation of material libraries and parallel screening for specific chemical or physical properties of interest. This approach is particularly advantageous in scenarios requiring the efficient exploration of a vast parameter space, where numerous variables, conditions, and combinations must be systematically assessed to accelerate discovery, optimization, and screening of target compounds. By enabling a more comprehensive and data-driven investigation, high-throughput techniques enhance innovation rates, improve cost-effectiveness, strengthen intellectual property development, shorten time-to-market, and increase the probability of successful outcomes. These advantages make HTE especially valuable in catalyst research and development.

Chemical catalysis is generally classified into homogeneous and heterogeneous catalysis. The application of HTE in homogeneous catalysis necessitates the integration of multiple disciplines, including the synthesis of diverse ligand libraries with sufficient steric and electronic variability, the adaptation of catalyst synthesis and screening to microscale formats, microreactor design, and the implementation of rapid screening techniques. In heterogeneous catalysis, HTE involves the generation and evaluation of high-density libraries of inorganic materials. Similar to conventional catalyst R&D, HTE for both homogeneous and heterogeneous systems typically follows a two-stage process: primary and secondary screening. Primary screening focuses on catalyst discovery and often employs non-traditional reactor designs for qualitative or semiquantitative evaluations using small sample sizes. Secondary screening is used for validation and optimization of promising candidates, typically utilizing smaller-scale equipment for refined assessments. To meet the evolving demands of HTE, various instrumental modifications have been introduced, leading to advancements in reactor design and rapid screening methodologies. These innovations are critical for the discovery and optimization of both homogeneous and heterogeneous catalysts. High-throughput such as magnetron sputtering [10], multi-arc ion plating [11], and additive

manufacturing [12] have been employed to produce composition-spread alloy films, which play a pivotal role in the development of solid catalytic materials.

Figure 2 illustrates the overarching application framework of HTE, which integrates HTE with AI, computational chemistry, and human expertise in a synergistic manner. Within this system, computational chemistry—primarily based on density functional theory (DFT) and molecular dynamics simulations—provides quantum-level predictions, facilitates rational catalyst design, and collaborates with experimental validation to minimize the need for extensive trial-and-error experimentation. AI-driven optimization accelerates the screening of catalyst formulations through data-driven modeling, while HTE systems validate these predictions in real time, ensuring a seamless transition from theoretical insights to practical implementation. Human expertise remains essential in guiding model training, interpreting results, overseeing the ethical deployment of AI technologies, and deepening the fundamental understanding of catalytic mechanisms, thereby fostering a dynamic human-machine collaborative framework. The integration of these three components effectively overcomes traditional R&D bottlenecks, propelling catalysis research toward intelligent, and high-precision advancements.

In the context of high-throughput experimentation applications, Zhao et al. [13] developed a sophisticated system integrating machine learning, robotic automation, and big data, which was further employed for the synthesis and retrosynthesis of nanocrystals. This study successfully combined automation and machine learning to enhance the efficiency of colloidal nanocrystal synthesis while providing a comprehensive analysis of HTE-generated big data. By demonstrating a data-driven robotic synthesis framework, this research represents a significant advancement in digital synthesis, enabling the transition from data to nanocrystals with precise morphological control. Moreover, it is expected to reduce reliance on traditional trial-and-error experimentation and labor-intensive characterization, thereby streamlining catalyst development and accelerating innovation.

Another notable example is the study by Pan et al. [14], which introduced a high-throughput methodology for the development of high-entropy alloy (HEA) electrocatalysts. Given the vast number of possible elemental combinations in quinary HEAs and their profound impact on catalytic performance, traditional experimental approaches are insufficient for efficiently exploring and identifying high-performance HEAs. To address this challenge, the research team integrated microscale precursor array printing with pulsed high-temperature synthesis technology, enabling the high-throughput fabrication of multi-element HEA arrays. For catalytic activity assessment, scanning electrochemical liquid cell microscopy was employed to conduct high-throughput and precise evalua-

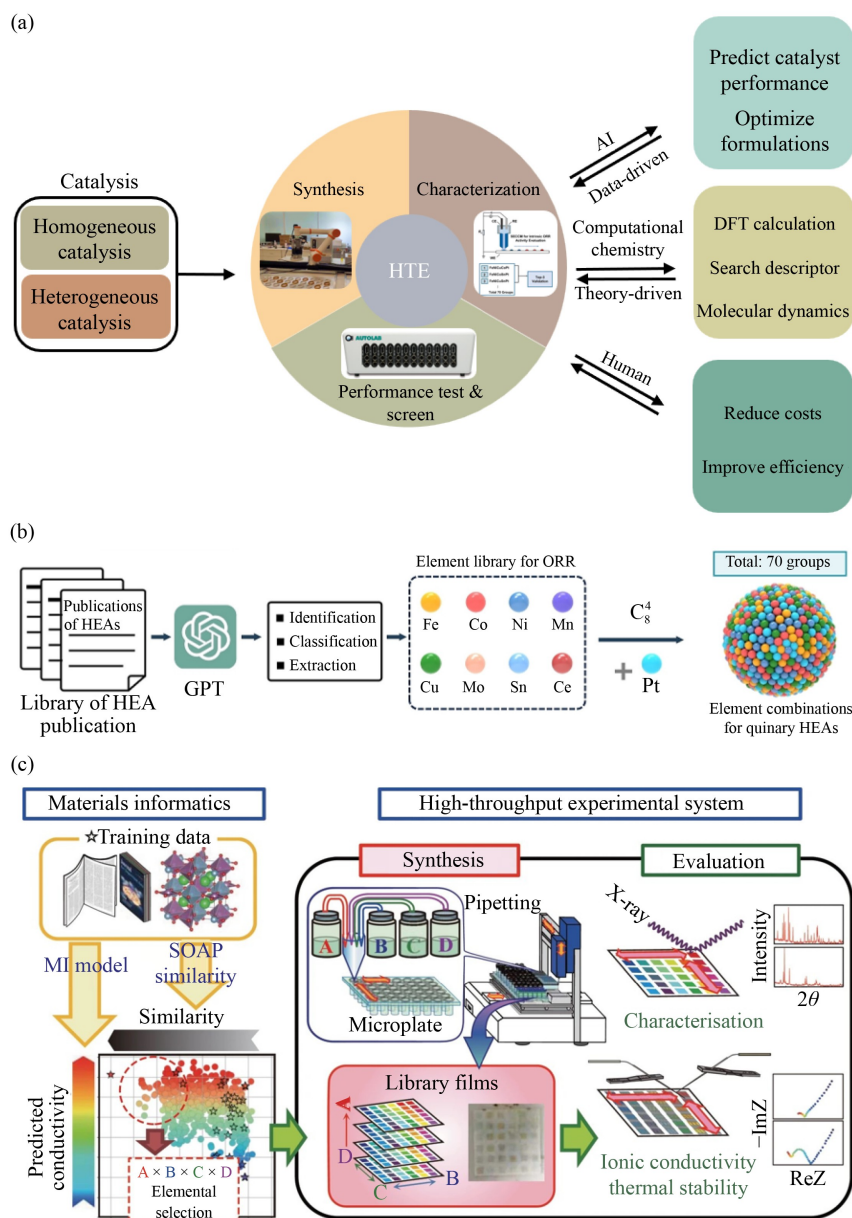


Fig. 2 (a) General application architecture of HTE. (b) Element combination design based on LLM. Reprinted with permission from Ref. [13], copyright 2024, Wiley. (c) Scheme of investigations of catalysts based on informatics tools and HTE. Reprinted with permission from Ref. [5], copyright 2020, Springer Nature.

tions of the intrinsic activity of HEA arrays in the oxygen reduction reaction, providing critical insights for the synthesis of high-performance catalytic materials. Additionally, DFT calculations were utilized to analyze element synergy effects and variations in active sites within high-activity element combinations, revealing common characteristics of advantageous elemental interactions. By combining HTE, experimental catalyst validation, and computational modeling, this approach offers a novel and efficient pathway for the accelerated discovery of high-performance multi-element materials in energy catalysis.

High-throughput synthesis and screening methodologies in materials science and catalysis were initially met

with resistance, primarily due to the complexity and high costs associated with their implementation. However, these approaches are now increasingly accepted and widely adopted in the field. These advanced high-throughput techniques are applicable across both homogeneous and heterogeneous catalysis, facilitating progress into more complex and challenging areas of catalyst chemistry.

3 Application of AI in catalytic research

As chemical research becomes increasingly complex and

high-dimensional, traditional research paradigms—primarily based on exhaustive exploration and trial-and-error approaches—are facing significant challenges. AI has emerged as an indispensable tool in catalyst design, providing innovative solutions to navigate the vast and complex parameter space that conventional methods often fail to address. This section examines the diverse applications of AI in catalytic research, emphasizing its role in predicting catalyst properties, mining experimental data, enhancing computational chemistry, and advancing spectroscopic techniques.

Predicting catalyst properties such as activity, selectivity, conversion rate, and acidity or basicity is a complex challenge that necessitates the integration of multiple factors, including elemental composition, structural characteristics, preparation conditions, and experimental data. Machine learning has emerged as a powerful tool in this domain, enabling the construction of predictive models by synthesizing these diverse data sources. Initially, a comprehensive data set of catalyst parameters is compiled from databases and the scientific literature. Elemental compositions are converted into physicochemical descriptors, microstructural features are extracted, and preparation conditions are systematically encoded. To address challenges such as data gaps, dimensional inconsistencies, and class imbalances, techniques like imputation, standardization, and sampling are applied. Appropriate algorithms are selected and trained using cross-validation, with performance metrics such as accuracy guiding the tuning of hyperparameters to improve prediction precision. Shapley additive explanations (SHAP) value analysis is employed to identify key influencing factors, which are subsequently validated experimentally. Iterative incorporation of feedback data facilitates model refinement, creating a “prediction-validation-update” closed-loop system that supports the rational design of catalysts. This workflow has been successfully applied in several studies. For instance, Guo et al. [15] employed a similar approach to predict catalytic ammonia decomposition for hydrogen production, revealing a strong positive correlation between ammonia decomposition and reaction temperature. They found that with a total metal loading below 20%, enhanced ammonia decomposition and hydrogen generation rates could be achieved. Additionally, Karthikeyan et al. [16] reviewed the use of machine learning to predict adsorption energy, reaction descriptors, structure-property relationships, and the synthesis conditions and reaction design of catalysts for hydrogen evolution reactions.

To date, the integration of AI with experimental data has led to significant breakthroughs in catalyst design. Recently, Wang et al. [7] delved deep into the data mining of “metal-support interaction” in supported metal catalysts. They summarized a large amount of experimental data from multiple core literatures, and

through interpretable AI algorithms, constructed a feature space composed of up to 30 billion expressions based on material properties as basic features. Using compressed sensing algorithms, combined with domain knowledge and theoretical derivation, they selected physically clear and numerically accurate descriptors and successfully established a control equation between “metal-support interaction” and material properties. This work provides a new perspective for the development of explainable AI technology for catalysis research.

Furthermore, the application of AI in computational chemistry has also shown the potential of enhancing the efficiency of catalyst design. Wu et al. [17] focused on challenges brought by complex reactions and dynamic changes in catalyst morphology during catalysis and developed a machine learning-based universal interatomic potential model named Catalytic Large Atomic Model (CLAM). The CLAM model combines advanced neural network architectures, including the deep potential model with a gated attention mechanism and the graph neural networks for large and diverse molecular simulation data sets, to achieve improved accuracy in energy and force predictions, which are crucial for catalytic applications. Moreover, the framework also incorporates a “local fine-tuning” algorithm to accelerate the process of transition state prediction and structure optimization without sacrificing accuracy. This method not only accelerates the construction of reaction networks but also promotes detailed kinetic analysis required for rational catalyst design.

Meanwhile, the application of AI in spectroscopy has emerged as a powerful tool for catalytic structure design. By randomly generating a large number of spectra to rapidly predict adsorption energy and iterating the spectral generation process to identify spectra corresponding to desired properties, AI can facilitate the continuous design of catalytic structures with optimized performance. Wang et al. [18] investigated how machine learning techniques can quantitatively determine surface-adsorbate properties from vibrational spectroscopy. They explored the use of machine learning to analyze vibrational spectroscopy, establishing a direct quantitative relationship between spectral signals and microscopic properties. This approach avoids the accumulation of errors inherent in traditional indirect methods, which typically infer structure from spectral signals before deducing properties. Additionally, they developed a quantitative spectrum-property relationship capable of directly determining key interaction properties of the substrate-adsorbate system, such as adsorption energy and charge transfer, from infrared and Raman spectral signals of the adsorbates. This highlights the potential of machine learning-assisted spectroscopy in accurately describing surface-adsorbate interactions. Building on this, Chong et al. [19] advanced the field by establishing an interpretable spectrum-property relationship through the

combination of physics-based spectral descriptors and symbolic regression methods. This study predicted the adsorption energy and charge transfer of carbon monoxide adsorbed on copper-based metal-organic frameworks, demonstrating that the model could maintain prediction accuracy even when trained on data sets containing as few as 20 samples or with up to 50% questionable labels. Furthermore, the models showed high confidence in identifying mislabeled data, achieving nearly 100% precision. The robust, transferable, and fault-tolerant spectrum-property relationship developed in this study enables real-time analysis of actual chemical data and offers a promising approach for future monitoring of catalytic processes or catalyst design. This relationship simplifies these complex tasks, making them as straightforward as performing basic mathematical computations.

In recent years, LLMs have also made rapid advancements and have been successfully applied in the field of chemistry. A notable example is ChemCrow, developed by M.Bran et al. [20]. This innovative LLM-based chemistry agent integrates 18 expert-designed tools with the capabilities of GPT-4. It autonomously plans and executes complex chemical syntheses, significantly enhancing the efficiency of organic synthesis and materials design. ChemCrow's successful synthesis of three known organocatalysts after receiving the prompt "Find a thiourea organocatalyst which accelerates the Diels-Alder reaction. After you find it, please plan and execute a synthesis for this organocatalyst" demonstrates its potential in catalysis-related applications. Another example of LLM application in chemistry is the AI workflow proposed by Lai et al. [9]. This workflow integrates LLMs with Bayesian optimization and an active learning loop to facilitate catalyst design. The LLM extracts catalytic recipes and reaction information from diverse literature, which are then used as inputs for model training and optimization. These examples underscore how advanced language models can be combined with traditional experimental design and robust optimization strategies, heralding a new paradigm in catalyst design.

4 The rise of self-driving laboratories in catalysis research

As discussed in previous sections, the intelligent transformation of chemical catalysis laboratories has undergone significant advancements. This evolution has progressed from the foundational applications of AI and ML for data analysis and basic automation, to HTE, and ultimately to the development of fully integrated, closed-loop intelligent systems. This progression has culminated in the emergence of self-driving laboratories, which

represent the latest paradigm in research automation.

A key milestone in this evolution is the work of Cooper and his team [21], who have made significant progress by integrating synthetic laboratories with autonomous mobile robots. This innovation enables the laboratory to operate equipment and make decisions autonomously, mimicking human-like decision-making processes, thus driving a paradigm shift toward self-driving laboratories. The modular workflow proposed in this study integrates mobile robots, automated synthesis platforms, liquid chromatography-mass spectrometry, and desktop nuclear magnetic resonance spectroscopy. This integration allows robots to process orthogonal measurement data, select successful reactions for further investigation, and automatically verify the reproducibility of screening results. By combining mobile robots with distributed synthesis and analysis platforms, this approach streamlines laboratory workflows, significantly enhancing the efficiency and automation of chemical research. The study demonstrated the application of this methodology in three areas: structurally diverse chemistry, supramolecular host-guest chemistry, and photochemical synthesis, particularly in exploratory chemistry that may yield multiple potential products. This work highlights the application of autonomous mobile robots in catalysis, particularly in advancing automated measurement, decision-making, and the understanding of supramolecular host-guest interaction mechanisms. Through this approach, researchers can gain deeper insights into the fundamental interactions in catalytic processes, thereby optimizing catalyst design.

As depicted in Fig. 3, the future of laboratories is envisioned to extend beyond standalone testing modules; instead, they will be integrated into automated workflows that not only provide real-time feedback on experimental results but also enhance the accuracy and efficiency of these processes. The ultimate vision is the realization of unmanned "dark labs", where researchers can input their concepts into a platform, and AI agents, utilizing advanced AI and HTE, will autonomously execute the tasks. This vision represents a future where the gap between conceptualization and experimental outcomes is significantly reduced, enabling researchers to more intuitively and rapidly translate their ideas into tangible scientific discoveries.

Zhu et al. [22] also proposed an all-round AI-Chemist equipped with scientific data intelligence, capable of performing fundamental tasks typically required in chemical research. This platform, based on a service-oriented architecture, can autonomously access and interpret literature from cloud databases to propose experimental plans accordingly. The AI-Chemist significantly reduced both the total experimental time and the waiting time for robotic operations through multi-task dynamic optimization. For instance, in the dye-sensitized photocatalytic water splitting experiment, the total time was reduced from 1810 to 980 min following

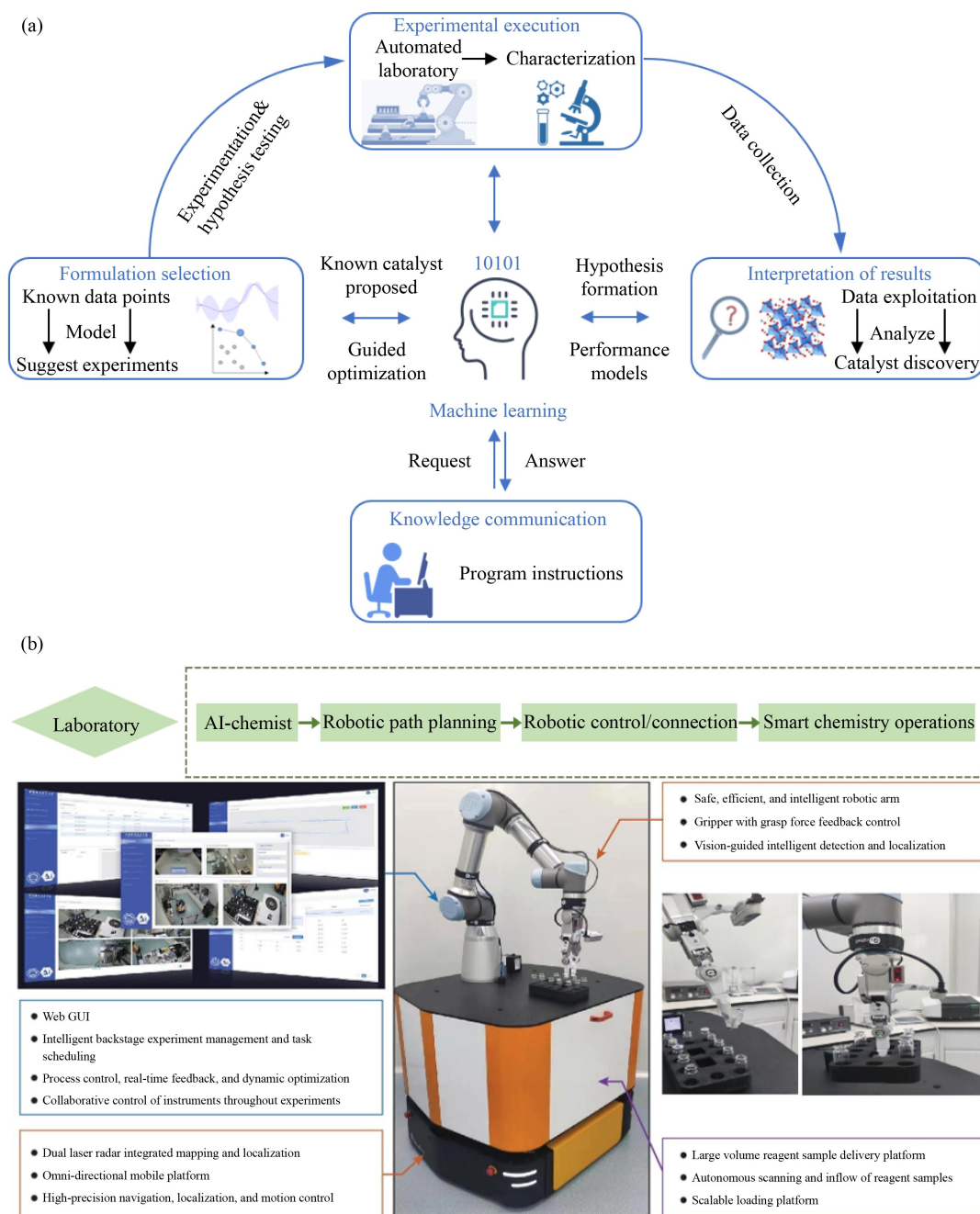


Fig. 3 (a) The role of AI and researchers in self-driving laboratories. (b) The robot and the workstations form a holistic smart laboratory for the AI-Chemist. Reprinted with permission from Ref. [22], copyright 2022, Oxford University Press.

optimization. Additionally, using non-noble metal oxygen evolution reaction electrocatalysts as an example, the AI-Chemist demonstrated its utility in the synthesis of high-entropy alloy nanoparticles. High-entropy alloys are promising candidates for future electrocatalysts due to their structural stability, diversity of adsorption sites, and high catalytic activity. This study showcased an AI system integrated with machine reading, automated experimentation, and intelligent data analysis, capable of simulating the cognitive processes of human scientists by independently generating scientific hypotheses, designing experiments, and analyzing results. These findings

underscore the vast potential for AI applications in chemical research.

Furthermore, the LLM agent ChemCrow, discussed in the previous section, can be further integrated with robotic synthesis platforms, such as IBM's cloud-connected RoboRXN, to further advance the development of self-driving laboratories. As illustrated in Fig. 4, this workflow provides the LLM with a suite of tools derived from various chemistry-related software packages and programs, along with user inputs. ChemCrow operates through an automated, iterative chain-of-thought process, in which it determines its path, selects tools,

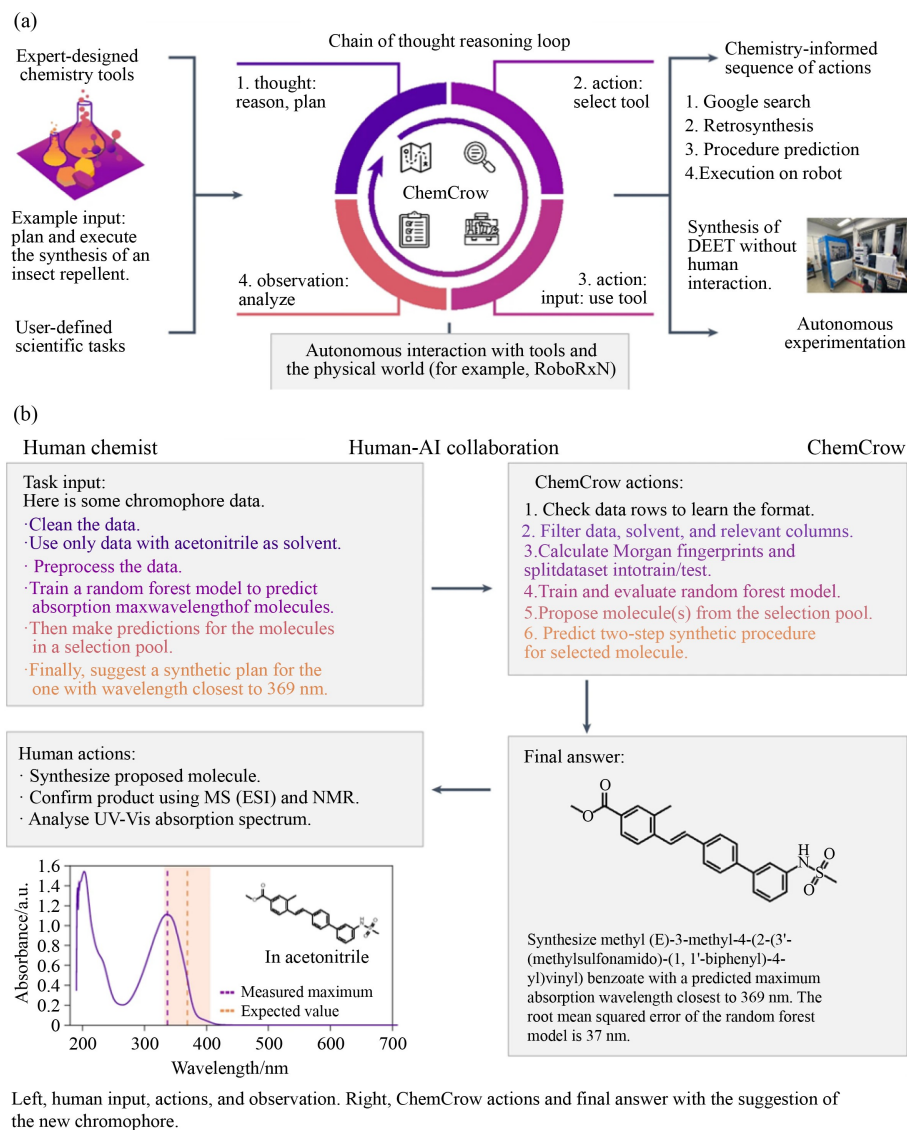


Fig. 4 (a) An overview of ChemCrow. (b) Human-model interaction leading to the discovery of a new chromophore. Reprinted with permission from Ref. [20], copyright 2024, Springer Nature.

specifies actions, and inputs, ultimately arriving at a final answer. This seamless translation of theoretical plans into practical outcomes demonstrates its ability to interact effectively with the physical world. In addition, ChemCrow facilitates the discovery of novel chromophores through its advanced machine learning capabilities. Evaluations indicate that ChemCrow outperforms GPT-4 in both chemical accuracy and task completion, thereby bridging the gap between experimental and computational chemistry. This advancement not only assists expert chemists by streamlining complex tasks but also democratizes access to advanced chemical knowledge, rendering it a valuable tool for both professionals and non-experts.

In summary, self-driving laboratories have demonstrated significant potential in the field of chemical catalysis, facilitating the systematic exploration of chemical and material spaces while enabling researchers

to conduct experiments in a more reproducible and efficient manner [23]. By integrating AI-driven data analysis, these automated systems not only increase the quantity of experiments but also enhance the success rate of each experiment through intelligent decision-making and adaptive experimentation. This integration improves research efficiency, reduces resource consumption, and accelerates the development of new catalysts. The future of laboratories is envisioned as being fully integrated into automated workflows, providing real-time, effective feedback on test results. Ultimately, this will lead to the realization of self-driving laboratories, wherein the complexities of laboratory operations are abstracted from the researcher. Researchers will be able to input their scientific hypotheses directly into an intuitive platform, and the advanced AI system will autonomously conduct the experiments and return the results, thereby streamlining the research process from concept to discovery.

5 Navigating the AI-driven catalyst pipeline from research and development to sustainable industrial application

In industrial catalysis, key metrics such as activity, selectivity, conversion rate, and economic viability must be considered and optimized in an integrated manner. The convergence of AI and HTE technologies provides new pathways to address these challenges and accelerate industrial scale-up. AI can predict catalyst performance, support cross-scale simulations, and elucidate reaction mechanisms, while HTE enables the rapid screening of candidate materials. This synergy not only accelerates the discovery and optimization of catalysts but also refines process parameters, enhances production efficiency, and improves product quality. Consequently, it reduces costs and fosters the development of more sustainable and environmentally friendly chemical production processes.

Recent advancements in AI have demonstrated substantial potential in accelerating the development and optimization of catalysts, from laboratory-scale research to industrial applications. For instance, a study by Ruan et al. [24] introduced an end-to-end chemical synthesis development platform powered by LLMs, highlighting the versatility of AI in catalysis research and industrial applications. They developed a unified LLM-based reaction development framework (LLM-RDF) designed to streamline the entire chemical synthesis process, encompassing literature review, substrate scope screening, reaction kinetics analysis, condition optimization, and scale-up.

As illustrated in Fig. 5, the LLM-RDF framework utilized an agent named Literature Scouter to search for synthetic methods capable of oxidizing alcohols to aldehydes using air as the oxidant. By employing vector search technologies and accessing a database comprising over 20 million academic articles, the agent identified a highly sustainable copper/TEMPO dual catalytic system. This system was recommended due to its superior environmental sustainability, simplicity, safety, chemoselectivity, and substrate compatibility in comparison to alternative methods. The framework further incorporated agents such as Experiment Designer, Hardware Executor, Spectrum Analyzer, and Result Interpreter to automate high-throughput screening of substrates and reaction conditions. Using an automated liquid handling platform, the system conducted experiments to determine optimal reaction conditions for various alcohol substrates. The agents analyzed gas chromatography data to evaluate yields and provided insights into the reactivity patterns of different substrates. To investigate the solvent effects on oxidation kinetics, the LLM-RDF framework orchestrated a series of automated experiments and data analysis tasks. The agents designed and executed kinetic experiments, analyzed proton nuclear magnetic resonance, and fitted

the data to kinetic models. The results revealed that dimethyl sulfoxide solvent provided superior chemoselectivity compared to acetonitrile, emphasizing the critical role of solvent choice in industrial applications. Additionally, the framework demonstrated automated optimization of reaction conditions using Bayesian optimization algorithms. The agents interfaced with an automated synthesis platform to iteratively adjust reaction parameters and maximize product yield. This closed-loop optimization process identified high-yield conditions with minimal human intervention, showcasing the potential of AI-driven optimization in industrial settings. Finally, the LLM-RDF framework facilitated the scale-up of the optimized reaction to a 1 g scale. The agents calculated stoichiometries, proposed a two-stage scale-up strategy, and provided guidance on reactor design and purification methods. This comprehensive approach ensured efficient scale-up while maintaining high yield and purity.

The LLM-RDF framework developed by Ruan et al. [24] represents a transformative approach to catalyst development and optimization. By integrating AI agents with automated experimental platforms, the framework substantially reduces the time and labor associated with traditional synthesis development. The technology not only accelerates the discovery of new catalysts but also enhances their industrial applicability by optimizing reaction conditions and facilitating efficient scale-up. However, challenges persist in fully realizing the potential of AI in industrial catalysis. These challenges include ensuring the reliability of AI-generated responses, incorporating domain-specific knowledge into LLMs, and addressing the limitations of current models in handling complex chemical mechanisms. Future advancements should focus on enhancing the robustness and adaptability of AI frameworks, as well as exploring open-source alternatives to improve transparency and reproducibility.

In summary, the integration of AI and HTE technologies, as exemplified by the LLM-RDF framework, provides a promising approach for advancing the catalyst development pipeline from research and development to sustainable industrial applications. This methodology not only streamlines the development process but also improves the efficiency and sustainability of chemical production, positioning AI as a pivotal element in the future of catalysis.

6 Challenges and research directions in AI-enabled catalyst applications

The integration of AI and HTE has already demonstrated substantial potential in accelerating catalyst development and optimizing industrial processes. Despite the considerable opportunities presented, several distinct challenges remain that must be addressed in order to fully leverage their potential in catalytic applications.

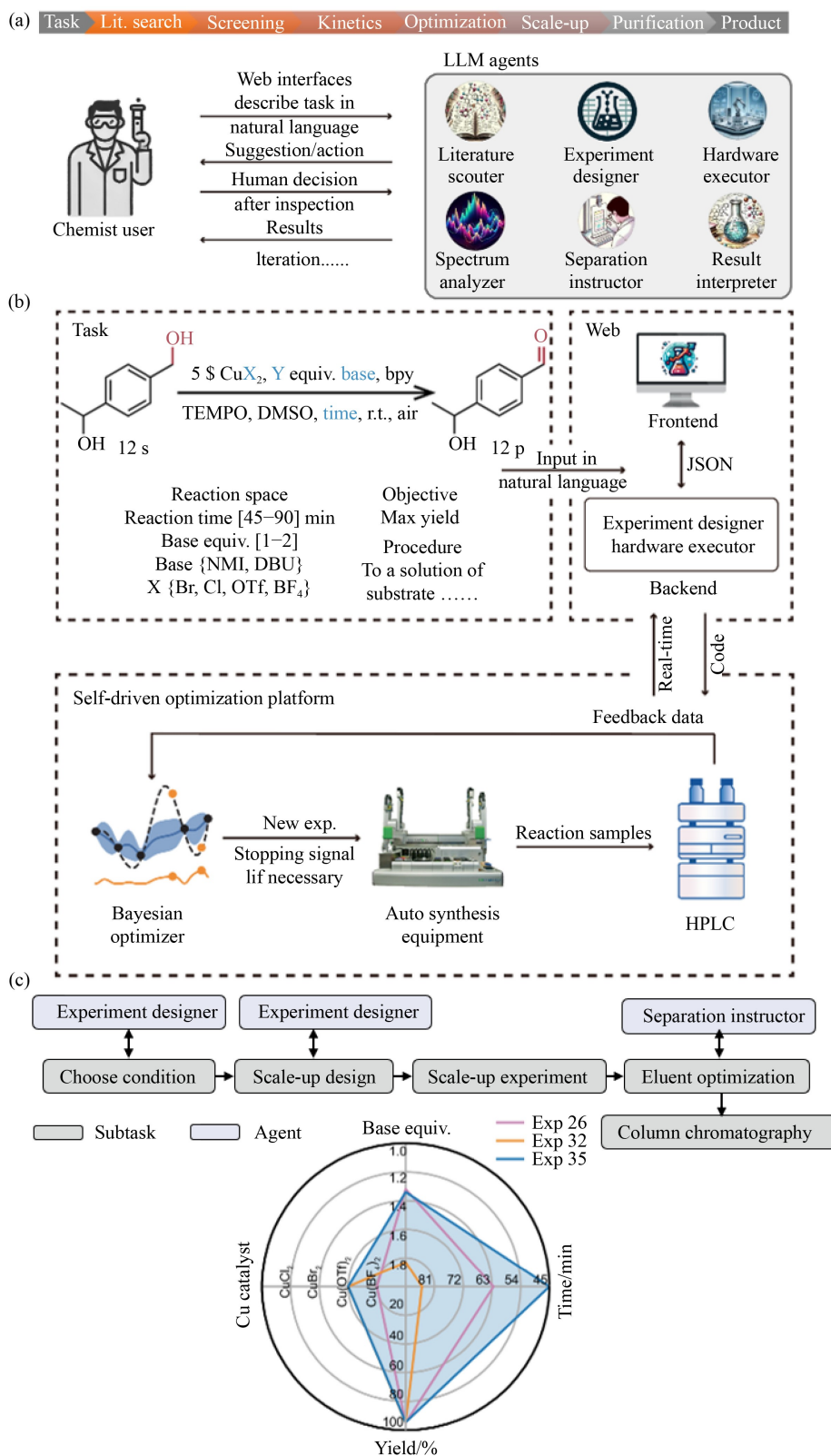


Fig. 5 (a) Workflow for LLM-RDF. (b) Self-driven reaction optimization system in LLM-RDF. (c) Workflow for reaction scale-up and the radar chart for experimental result comparison. Reprinted with permission from Ref. [24], copyright 2024, Springer Nature.

The key challenges that must be addressed include: (a) data quality and accessibility: the effectiveness of AI in catalyst design and optimization heavily depends on

the quality and breadth of the available data. The limited availability of comprehensive data sets that encompass a wide range of catalyst behaviors under various laboratory

conditions can hinder the optimization capabilities of AI workflows; (b) model interpretability and reliability: AI models must be interpretable to enhance user confidence in their reliability for real-world applications and to facilitate the design of experiments and trials. Current models, often characterized by their complexity and “black-box” nature, need to be made more transparent to improve their practical applicability; (c) resource optimization for HTE and AI: while HTE and AI technologies offer significant efficiency gains, they require substantial resources, including financial investment and large data sets. It is essential to develop more resource-efficient AI methodologies, especially for low-data scenarios, where simpler models can yield meaningful results; (d) integration of multi-scale simulations: catalyst design necessitates simulations across multiple scales, from the atomic level to the macroscopic. The difficulty of achieving effective cross-scale modeling limits the ability to evaluate catalyst performance in real-world applications; (e) scalability to industrial applications: catalysts that perform well under laboratory conditions may encounter different challenges when scaled up for industrial use, such as mass transfer limitations. The transition from laboratory to industrial scale requires the integration of domain-specific knowledge, multi-step trial-and-error processes, and calibration.

Future research may focus on the following areas: (a) data quality and database construction: there is a pressing need to construct comprehensive catalyst databases to support the training and validation of AI models. Integrating experimental data with computational simulation results can enhance the diversity and coverage of data sets. It is crucial to ensure the quality and consistency of data when merging data sets from various sources. LLMs could be leveraged to extract key information and develop standardized data formats, thereby improving data curation practices. Additionally, data augmentation techniques could be employed to increase dataset diversity; (b) interpretability and validation of AI models: there is a concerted effort to develop transparent and verifiable AI models, thereby enhancing their credibility in both scientific research and industrial applications. This includes integrating AI with “white-box” first-principle models, elucidating AI decision-making processes, and validating model predictions to improve reliability. Techniques such as SHAP values and feature importance analysis can provide deeper insights into model behavior. Moreover, the use of interpretable descriptors and symbolic regression methods can enhance the explainability of AI models in catalysis; (c) innovation in high-throughput technologies and data-efficient AI: advancements in HTE are essential for improving experimental efficiency and reducing costs. Well-designed automation and robotics can significantly reduce manual labor and shorten experimental timelines. Furthermore, the development of intelligent self-driving

laboratories, integrating HTE with data-efficient AI prediction and optimization techniques, is critical for minimizing non-directed experiments and exhaustive enumeration of formulas or process conditions. Accurate AI prediction models can improve the precision of search directions, while optimization methods can reduce the need for exhaustive enumeration, thus accelerating the identification of optimal formulas or conditions. Multi-objective optimization is often encountered in catalyst performance evaluation. Optimization techniques, such as Bayesian optimization, must balance exploration (searching untested areas) and exploitation (capitalizing on known effective parameters), ensuring an efficient approach to catalyst optimization. Moreover, training and deploying AI models, particularly LLMs, requires significant computational resources. Developing more efficient algorithms and utilizing cloud computing and distributed platforms can help address these challenges; (d) multi-scale modeling approaches: developing multi-scale modeling approaches is essential for obtaining a comprehensive understanding of catalyst properties and performance. Combining quantum mechanics, molecular dynamics, and continuum models can yield more holistic predictions of catalyst performance. HTE can assist in validating these models, while AI can be utilized to correlate multi-scale data and refine modeling details based on real-world observations; (e) industrial scale-up strategies: research into scaling strategies from laboratory to industrial settings is essential to ensure catalyst consistency and efficiency. Simulating industrial conditions and optimizing the scale-up pathway can mitigate risks and costs associated with this process. This can be achieved by incorporating domain-specific knowledge via LLMs and utilizing hybrid models that combine AI with classical computational chemistry.

In summary, while AI offers substantial opportunities for advancing catalyst research and development, it also presents a series of challenges that require innovative solutions. The integration of these technologies, particularly through the use of LLMs and intelligent closed-loop laboratories, is at the forefront of AI and HTE development. This convergence plays a pivotal role by extracting knowledge from scientific literature and automating the experimental process. In conjunction with ongoing research in explainable machine learning models, active learning, and Bayesian optimization, these advancements are essential for addressing current challenges and unlocking the full potential of AI and HTE in catalysis.

7 Conclusions

The integration of AI and HTE has emerged as a transformative force in the field of catalysis, providing

unprecedented opportunities to accelerate the development and optimization of catalysts from laboratory research to industrial applications. This convergence has the potential to revolutionize traditional research paradigms, which have historically relied on trial-and-error approaches and extensive experimentation. By harnessing the predictive capabilities of AI and the rapid screening potential of HTE, researchers can now navigate the complex landscape of catalyst development with greater efficiency, reducing time-to-market and enhancing the sustainability of chemical production processes.

The emergence of self-driving laboratories, exemplified by the development of robotic AI-chemists and LLM-enabled AI agents such as ChemCrow, highlights the potential for fully automated and intelligent systems capable of autonomously executing complex workflows, thereby accelerating the pace of scientific discovery. Furthermore, existing frameworks, such as LLM-RDF, which streamline the entire catalyst development pipeline, demonstrate the potential of AI-driven systems that span literature search, substrate screening, reaction optimization, and scale-up. By automating repetitive tasks and providing real-time feedback, these frameworks enable researchers to focus on innovation rather than routine experimentation. Despite these advancements, several challenges remain. Enhancing the quality and quantity of data, ensuring the reliability and interpretability of AI models, optimizing resource utilization, and developing robust multi-scale modeling approaches are critical areas requiring continued research. Additionally, the successful transition from laboratory-scale to industrial-scale applications demands a deeper integration of domain-specific knowledge and the development of scalable, cost-effective strategies.

In conclusion, the AI-catalyst pipeline represents a novel paradigm in catalysis research and development. By addressing the challenges discussed in this perspective and leveraging the synergies between AI, HTE, and traditional computational and experimental methodologies, the field can progress toward more efficient, sustainable, and innovative catalytic processes.

Competing interests The authors declare that they have no competing interests.

Acknowledgements This work is supported by the Special Project of National Natural Science Foundation (Grant Nos. W2411009 and 42341204).

References

1. Hagemeyer A, Jandeleit B, Liu Y, Poojary D M, Turner H W, Volpe A F Jr, Henry Weinberg W. Applications of combinatorial methods in catalysis. *Applied Catalysis A: General*, 2001, 221(1–2): 23–43

2. Kumar G, Bossert H, McDonald D, Chatzidimitriou A, Ardagh M A, Pang Y, Lee C, Tsapatsis M, Abdelrahman O A, Dauenhauer P J. Catalysis-in-a-box: robotic screening of catalytic materials in the time of covid-19 and beyond. *Matter*, 2020, 3(3): 805–823
3. Weidenhof B, Reiser M, Stowe K, Maier W F, Kim M, Azurdia J, Gulari E, Seker E, Barks A, Laine R M. High-throughput screening of nanoparticle catalysts made by flame spray pyrolysis as hydrocarbon/NO oxidation catalysts. *Journal of the American Chemical Society*, 2009, 131(26): 9207–9219
4. Liu X, Liu B, Ding J, Deng Y, Han X, Zhong C, Hu W. Building a library for catalysts research using high-throughput approaches. *Advanced Functional Materials*, 2022, 32(1): 2107862
5. Matsubara M, Suzumura A, Ohba N, Asahi R. Identifying superionic conductors by materials informatics and high-throughput synthesis. *Communications Materials*, 2020, 1(1): 5
6. Zhang X, Wang L, Helwig J, Luo Y, Fu C, Xie Y, Liu M, Lin Y, Xu Z, Yan K, et al. Artificial intelligence for science in quantum, atomistic, and continuum systems. *arXiv*, 2023
7. Wang T, Hu J, Ouyang R, Wang Y, Huang Y, Hu S, Li W. Nature of metal-support interaction for metal catalysts on oxide supports. *Science*, 2024, 386(6724): 915–920
8. Toyao T, Maeno Z, Takakusagi S, Kamachi T, Takigawa I, Shimizu K. Machine learning for catalysis informatics: recent applications and prospects. *ACS Catalysis*, 2020, 10(3): 2260–2297
9. Lai N S, Tew Y S, Zhong X, Yin J, Li J, Yan B, Wang X. Artificial intelligence (AI) workflow for catalyst design and optimization. *Industrial & Engineering Chemistry Research*, 2023, 62(43): 17835–17848
10. Gao B, Hu J, Tang S, Xiao X, Chen H, Zuo Z, Qi Q, Peng Z, Wen J, Zou D. Organic-inorganic perovskite films and efficient planar heterojunction solar cells by magnetron sputtering. *Advanced Science*, 2021, 8(22): 2102081
11. Guan Q, Guo X, Fan L, Meng B, Sha J. High-throughput preparation and quick characterization of oxidation behaviors of complex Al-Cr compositional gradient coatings on a novel Co-Al-W-based superalloy prepared using multi-arc ion plating technology. *Intermetallics*, 2024, 174: 108464
12. Sommer N, Bauer A, Kahlmeyer M, Wegener T, Degener S, Liehr A, Bolender A, Vollmer M, Holz H, Zeiler S, et al. High-throughput alloy development using advanced characterization techniques during directed energy deposition additive manufacturing. *Advanced Engineering Materials*, 2023, 25(15): 2300030
13. Zhao H, Chen W, Huang H, Sun Z, Chen Z, Wu L, Zhang B, Lai F, Wang Z, Adam M L, et al. A robotic platform for the synthesis of colloidal nanocrystals. *Nature Synthesis*, 2023, 2(6): 505–514
14. Pan Y, Shan X, Cai F, Gao H, Xu J, Zhou M. Accelerating the discovery of oxygen reduction electrocatalysts: high-throughput screening of element combinations in Pt-based high-entropy alloys. *Angewandte Chemie International Edition*, 2024, 63(37): e202407116
15. Guo W, Shafizadeh A, Shahbeik H, Rafiee S, Motamedi S, Ghafarian Nia S A, Nadian M H, Li F, Pan J, Tabatabaei M, et al. Machine learning for predicting catalytic ammonia decomposition: an approach for catalyst design and performance

- prediction. *Journal of Energy Storage*, 2024, 89: 111688
16. Karthikeyan M, Mahapatra D M, Razak A S A, Abahussain A A, Ethiraj B, Singh L. Machine learning aided synthesis and screening of HER catalyst: present developments and prospects. *Catalysis Reviews: Science and Engineering*, 2024, 66(4): 997–1027
 17. Wu Z, Zhou L, Hou P, Liu Y, Guo T, Liu J. Catalytic large atomic model (CLAM): a machine-learning-based interatomic potential universal model. *ChemRxiv*, 2024
 18. Wang X, Jiang S, Hu W, Ye S, Wang T, Wu F, Yang L, Li X, Zhang G, Chen X, et al. Quantitatively determining surface-adsorbate properties from vibrational spectroscopy with interpretable machine learning. *Journal of the American Chemical Society*, 2022, 144(35): 16069–16076
 19. Chong Y, Huo Y, Jiang S, Wang X, Zhang B, Liu T, Chen X, Han T, Smith P E S, Wang S, et al. Machine learning of spectral-property relationship for imperfect and small chemistry data. *Proceedings of the National Academy of Sciences of the United States of America*, 2023, 120(20): e2220789120
 20. M.Bran A, Cox S, Schilter O, Baldassari C, White A D, Schwaller P. Augmenting large language models with chemistry tools. *Nature Machine Intelligence*, 2024, 6(5): 525–535
 21. Dai T, Vijaykrishnan S, Szczypiński F T, Ayme J, Simaei E, Fellowes T, Clowes R, Kotopantov L, Shields C E, Zhou Z, et al. Autonomous mobile robots for exploratory synthetic chemistry. *Nature*, 2024, 635(8040): 890–897
 22. Zhu Q, Zhang F, Huang Y, Xiao H, Zhao L, Zhang X, Song T, Tang X, Li X, He G, et al. An all-round AI-chemist with a scientific mind. *National Science Review*, 2022, 9(10): nwac190
 23. Abolhasani M, Kumacheva E. The rise of self-driving labs in chemical and materials sciences. *Nature Synthesis*, 2023, 2(6): 483–492
 24. Ruan Y, Lu C, Xu N, He Y, Chen Y, Zhang J, Xuan J, Pan J, Fang Q, Gao H, et al. An automatic end-to-end chemical synthesis development platform powered by large language models. *Nature Communications*, 2024, 15(1): 10160