

Cite this: *Digital Discovery*, 2026, 5, 497

# Accelerating catalytic advancements through the precision of high-throughput experiments & calculations

Jenny G. Vitillo,<sup>1</sup> Alán Aspuru-Guzik,<sup>2</sup> Eric Duskocil,<sup>3</sup> Omar K. Farha,<sup>4</sup> Timur Islamoglu,<sup>5</sup> Heather J. Kulik,<sup>6</sup> Peter M. Margl,<sup>7</sup> Stuart Miller,<sup>8</sup> Jordan Reddel,<sup>9</sup> Aayush R. Singh<sup>10</sup> and Varinia Bernales<sup>11</sup>

The growing demand for energy-efficient processes to support a sustainable future drives the need for research to rapidly explore chemical and material space through accelerated catalyst discovery initiatives. Recent breakthroughs in high-throughput experimental and computational methods are transforming the catalysis field, surpassing traditional approaches to manipulating variables in catalytic processes. Key advancements in innovation include the integration of machine learning for efficient catalyst screening, high-throughput experimentation, data-driven methodologies employing comprehensive databases, and *in situ* and *in operando* techniques for realistic observations. This progress has undoubtedly been intertwined with a collaborative framework across disciplines, reshaping catalyst discovery methods in both industry and academia. This Opinion article presents a multifaceted perspective from coauthors with expertise spanning various stages of the Technology Readiness Level spectrum, highlighting both opportunities and persistent challenges in integrating computational and experimental approaches in catalysis. These challenges span from obtaining high-quality experimental data, scaling simulations to industrially relevant materials and process conditions to navigating the complexity and predictive accuracy of computational models.

Received 25th November 2025  
Accepted 20th January 2026

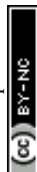
DOI: 10.1039/d5dd00524h

rsc.li/digitaldiscovery

## 1. Introduction

The pressing global challenges of today urge disruptive technological innovations to accelerate and broaden our exploration of the fundamental principles governing chemical reactions. In addressing the need for faster discovery and testing, high-throughput (HT) methods in catalysis have revolutionized the field by enabling the rapid and comprehensive exploration of vast chemical spaces previously out of reach.<sup>1</sup> This allows an unprecedented acceleration in discovering and optimizing new

catalytic systems and identifying optimized reaction conditions.<sup>2–4</sup> High-throughput experimentation (HTE) has emerged as a transformative approach to addressing the need for faster discovery and testing. First adopted in 1984 in the life sciences for creating peptide libraries *via* parallel synthesis on microtiter plates,<sup>5</sup> HTE's application in the physical sciences came later. An early example of its application to catalysis involved using infrared (IR) imaging thermography to screen libraries of heterogeneous catalysts for hydrogen oxidation activity by detecting heat release.<sup>6</sup>

<sup>1</sup>Department of Science and High Technology and INSTM, Università Degli Studi Dell'Insubria, Via Valleggio 9, Como 22100, Italy<sup>2</sup>Department of Chemistry, University of Toronto, 80 St. George St., Toronto, ON M5S 3H6, Canada<sup>3</sup>Acceleration Consortium, 700 University Ave., Toronto, ON M7A 2S4, Canada<sup>4</sup>Department of Computer Science, University of Toronto, 40 St George St., Toronto, ON M5S 2E4, Canada<sup>5</sup>Vector Institute for Artificial Intelligence, W1140-108 College St., Schwartz Reisman Innovation Campus, Toronto, ON M5G 0C6, Canada<sup>6</sup>Department of Materials Science & Engineering, University of Toronto, 184 College St., Toronto, ON M5S 3E4, Canada<sup>7</sup>Department of Chemical Engineering & Applied Chemistry, University of Toronto, 200 College St., Toronto, ON M5S 3E5, Canada<sup>8</sup>Canadian Institute for Advanced Research (CIFAR), 661 University Ave., Toronto, ON M5G 1M1, Canada<sup>9</sup>NVIDIA, 431 King St W #6th, M5V 1K4, Toronto, Canada<sup>10</sup>Applied Sciences & Technology, BP, 30 S Wacker Drive, Suite 900, Chicago 60606, IL, USA<sup>11</sup>Department of Chemistry and International Institute for Nanotechnology, Northwestern University, Evanston 60208, IL, USA<sup>12</sup>Paula M. Trienens Institute for Sustainability and Energy, Northwestern University, Evanston 60208, IL, USA<sup>13</sup>Materials Discovery Research Institute, UL Research Institutes, Skokie 60077, Illinois, USA<sup>14</sup>Departments of Chemical Engineering and Chemistry, Massachusetts Institute of Technology, Cambridge, MA, USA<sup>15</sup>Dow, Inc., Core R&D, 1776 Building, Midland 48667, Michigan, USA<sup>16</sup>SandboxAQ, Palo Alto, 94301, CA, USA

Building on the significance of HTE, it is essential to clarify key terminology, as some terms are often used interchangeably despite their nuanced differences. High-throughput synthesis (HTS) refers to systematically exploring chemical space to produce desired materials, utilizing tools that enable parallel reaction setups or simultaneous characterization processes.<sup>7</sup> On the other hand, “accelerated materials discovery,” which includes the expanding field of self-driving laboratories,<sup>8</sup> builds upon HTS by incorporating advanced data analysis techniques—with simple methods such as Design of Experiments to modern approaches such as Bayesian optimization—to create a feedback loop that optimizes and streamlines the synthesis and characterization processes for targeted compounds.<sup>9–11</sup> This distinction underscores the progression of HTE from a tool for rapid experimentation to an integrated framework driving innovation in material science.

Although HT techniques are becoming indispensable for remaining competitive in academia and industry, significant barriers prevent both high-throughput experimentation (HTE) and computation (HTC) in catalysis from reaching their full potential. Key areas for improvement include (i) enhancing data management, standardization, and integration, (ii) accelerating characterization processes, and (iii) fostering transparent and efficient interdisciplinary collaboration—each of these requiring ongoing focus and innovation. The large amount of data produced by HT methods in short periods must be analyzed to extract useful information. This results in a substantial manual workload, diminishing the advantages of high-throughput methods and hindering their broader adoption.<sup>1,12</sup>

Lately, most of the solutions on how to automate the handling of these large datasets have come from artificial intelligence (AI) methods, which are particularly powerful in solving this issue.<sup>1,12,13</sup> Collaboration between chemists, materials scientists, engineers, and data scientists becomes essential to leveraging the full potential of these technologies. It goes without saying that maintaining depth in subject-matter expertise is essential for success in this field, as in any scientific endeavor. A frequent bottleneck in HTE is the inherent timescale mismatch between characterization and other critical processes, such as formulation, reaction, and synthesis methods. In many cases, HT characterization becomes the most crucial and primary limiting factor in accelerating discovery workflows. In catalysis—the focus of this opinion piece—characterizing key catalytic intermediates is particularly important. However, current experimental methods are still not amenable to HT approaches. To overcome this limitation, it is important to complement experimental measurements with accurate and inexpensive predictions from first principles. Several HTC workflows have been designed to achieve these needs,<sup>14–17</sup> allowing for the parallel acquisition of mechanistic data, including intermediate identification. Combined with theoretical predictions, such workflows help bridge the gap between detailed mechanistic understanding and scalable discovery. This convergence of HTE and computational chemistry, driven by advances in robotics and AI, has been further augmented with deep learning, generative models, and machine learning

potentials to help address data gaps.<sup>18</sup> The topic has been the subject of several recent reviews and perspectives,<sup>19–25</sup> and some positive examples of predictions,<sup>26</sup> proving the potential of data science in catalysis to optimize reaction conditions and improve yields.<sup>27,28</sup>

This Opinion article condensed the ideas discussed in the framework of a symposium organized at the ACS Fall of 2024 by the Catalysis Division with a homonymous title as the present article, where the authors, experts in the field from academia and industry, participated in a panel discussion to cover this topic.<sup>29</sup> The aim of this Opinion article is to expose the topics explored during this event, covering the current state of high-throughput catalysis, highlighting recent breakthroughs, and discussing the opportunities and challenges in this rapidly evolving field. We specifically focus on three interconnected aspects critical to accelerating catalytic advancements: (i) high-throughput characterization for balancing data precision with experimental efficiency (Section 3); (ii) standardization of data and metadata to improve data sharing and reproducibility (Section 4); and (iii) integrated workflows connecting experimental and computational systems for autonomous discovery (Section 5). By concentrating on these key challenges and opportunities, this work aims to provide actionable recommendations for the catalysis community.

In this paper, catalytic discovery refers to the identification of new catalysts or reaction systems enabled by data-driven and high-throughput methods, whereas catalytic advancement encompasses the broader acceleration of catalyst understanding, optimization, and deployment. It is important to note that homogeneous (molecular) and heterogeneous (solid-based) catalysis differ substantially in synthesis control, reaction environment, and characterization complexity. While the data-driven and high-throughput strategies discussed here are broadly applicable, their implementation requires domain-specific adaptation to account for these intrinsic differences.

## 2. Current landscape and advances in accelerated catalytic discovery

To our knowledge, the first high-throughput study dates to 1909–1912, when Alwyn Mittash, in collaboration with Georg Stern and Hans Wolf and under the direction of Carl Bosch, ran thousands of systematic reactions to find an economically viable alternative to osmium and uranium-based catalysts for the ammonia production.<sup>30</sup> This resulted in the identification of metallic iron promoted by small alumina and potassium oxide admixtures. Nowadays, this catalyst is used with few modifications for the Haber-Bosch process.<sup>30</sup> Although this pioneering study took place more than 100 years ago, a search on Google Patent using (high throughput AND catalysis) as parameters highlights that from the > 100 000 results, most are dated from 2010 onwards (see Fig. 1a).<sup>31</sup> This graph indicates that, only about 14 years ago, HT methods transitioned from niche approaches to widely adopted techniques due to the automation of the process. A key factor in this transition has been the development of robotic platforms capable of parallel synthesis





a catalyst in relative isolation but also of painstaking formulation, scaling, and tuning of the catalytic process. For instance, in heterogeneous catalytic systems, carefully formulated, highly disordered systems are often key to sustained catalytic performance. These systems are difficult to model with current DFT methods because they require large scales and long simulation times. Here, continued development of simulation methods and machine learning methods to augment traditional, physical models will be required to match HTC to HTE. In homogeneous catalysis, similar challenges pertain, for instance, to reactions involving ion pairs, systems with significant conformational flexibility, and systems where the solvent actively participates in catalysis—especially in combination with complex electronic structures. Nonetheless, strategies that leverage qualitative trends, like energy descriptors, for systems with complex electronic structures have successfully guided the design of new iron-based molecular catalysts for methane oxidation.<sup>13</sup> For solid-state simulations in heterogeneous catalysis, the choice of the electronic structure method is often dictated by a trade-off between accuracy and computational feasibility, particularly in high-throughput or large-scale screening studies. While local DFT functionals remain widely used due to their favorable scaling and robustness, they are increasingly complemented by more advanced approaches. State-of-the-art modeling now frequently incorporates dispersion-corrected functionals, meta-GGA formulations, hybrid functionals, or multilevel and embedding strategies to improve accuracy and transferability.<sup>51</sup> Hybrid functionals and beyond-DFT methods can provide significant improvements for systems involving localized electronic states, such as transition metal oxides, non-zero oxidation states, or open-shell configurations, albeit at a substantially higher computational cost.<sup>52</sup> Conversely, for metallic or gapless systems, the benefits of hybrid functionals are often limited, and carefully validated local approaches may still yield reliable trends. In all cases, systematic benchmarking and validation against experimental data or higher-level calculations on representative models are essential to ensure predictive reliability.<sup>52</sup>

The coupling of AI tools and automation for data analysis is also becoming more and more common in HTE, with the development of mobile robotic chemists and self-driving laboratories (SDLs), bringing significant advantages in process efficiency and decision-making (see Section 5).<sup>12,53,54</sup> The high investment necessary for HTE equipment remains an important discussion point for its scientific and social repercussions, limiting opportunities for innovation and development. Integrating these advanced technologies with sustainable practices and cost-effective solutions remains a critical goal. This has brought to the development of “frugal twins” or low-cost SDLs,<sup>55</sup> that are low-cost, functionally simplified versions of sophisticated SDLs. Built using off-the-shelf components and open-source software, these systems require a capital investment up to three orders of magnitude lower than their high-end counterparts. Originally conceived to address the accessibility gap in HTE technology, frugal twins enable cost-effective automation of experimental workflows. They offer a pragmatic solution for academic labs with limited resources, supporting increased

throughput and experimental reproducibility while preserving key features such as scalability and modularity.<sup>55</sup>

### 3. High-throughput characterization: balancing precision with efficiency

The panel unanimously supported the view that advancing characterization techniques is essential to fully realize the potential of high-throughput (HT) catalysis. Catalysis, particularly in its heterogeneous form, exhibits intricate behavior due to dynamic active sites, evolving reaction environments, and the complex interplay between structure and function. These features create significant challenges for data generation, interpretation, and modeling. In heterogeneous systems, additional factors such as multiple concurrent surface reaction pathways and mass or energy transfer limitations further complicate the correlation between catalyst structure, composition, and performance. AI-assisted HT approaches should therefore aim to incorporate, or at least recognize, these intrinsic phenomena. On the computational side, limited and potentially biased training data, reliance on DFT-derived datasets, and the need for rigorous uncertainty quantification remain key obstacles to developing robust, transferable predictive models. To maximize learning from HT catalysis experiments, the experimenter must have accurate knowledge not only of the catalyst itself but also of the products and side-products formed in the reactor. The former is often comparatively trivial, *e.g.*, when using a well-defined homogeneous catalyst, but sometimes extraordinarily difficult, such as when the catalyst is a product-by-process in which only the preparation process, but not the microstructure of the catalyst, is known. In addition, real-life reactions frequently involve complex multicomponent mixtures that pose significant analytical challenges. To achieve the maximum impact of HTE, the high-throughput analytical methodology needs innovation for HT analytical not to remain a bottleneck. The limitations of analytical characterization methods have collateral implications for any HTC intended to accompany HTE, because computational approaches perform poorly when required to infer chemical outcomes in an inherently combinatorial space under weak experimental constraints, a challenge that platforms such as AutoRXN attempt to mitigate but ultimately cannot resolve. To maximize the power of HTE, it is also important to have good data integration because HTE characterization generates oodles of data that require an inhuman effort to work up, so ML should be used, but that can only work if the data is fully integrated from the experimental design all the way forward to characterization.

While HTE's parallel screening accelerates exploration across a broader chemical and material space, its effectiveness ultimately depends on carefully (1) selecting the initial parameters and (2) integrating characterization and testing techniques to ensure accuracy. For instance, at the symposium, Bozal-Ginesta *et al.* showed an excellent example of HT synthesis and characterization of perovskite thin film libraries.<sup>56</sup> This work characterized compositional maps using



various HT techniques, such as XRF, XRD, Raman, and EIS. Herein, the authors stressed the importance of a sufficiently large and consistent experimental dataset to extract relevant oxygen reduction reaction properties alongside appropriate data processing and analysis routines. Gagliardi and Del Ferro also showcased at the symposium a HTE campaign, where researchers evaluated the performance of decorated MOFs in propylene dimerization to hexadienes.<sup>4,57</sup> Initial exploration utilizing Bayesian optimization techniques required over six months and 721 experiments, unexpectedly resulting in low yield (<5%) reactions. Consequently, the reevaluation of the initial parameters led the authors to redesign the experiments, increasing the H<sub>2</sub> concentration to 40 vol%, allowing significant yield improvements.<sup>4,57</sup> These improvements were attributed to the formation of larger nanoparticles under higher H<sub>2</sub> levels. This example illustrates the relevance of taking a human-in-the-loop approach, *i.e.*, subject matter expertise to guide parameter choices effectively, as screening all possible conditions is often impractical due to time and cost.<sup>4,54,57,58</sup> Crucially, the success of this campaign depended on careful post-hoc characterization of the catalysts using high-angle annular dark-field scanning transmission electron microscopy (HAADF-STEM) and DFT modeling, which revealed that larger copper nanoparticles formed under high H<sub>2</sub> conditions were responsible for the improved catalytic activity. This underscores that high-throughput experimentation alone is insufficient and effective HT catalysis requires integrating detailed characterization and mechanistic understanding to interpret results accurately and guide subsequent experimental design.

For accelerated catalyst discovery platforms relying on a high-throughput framework, a tight coupling between synthesis and characterization is fundamental. For instance, high-throughput PXRD is an invaluable tool for identifying materials or validating synthesis processes. Patterns can be matched against databases or analyzed directly, with parameters then compared to target materials. However, relying solely on statistical measures such as goodness-of-fit for analysis is insufficient. Poor fits may arise from unfitted diffraction features or slight inaccuracies in peak shapes.<sup>59,60</sup> Trained experts should evaluate these results and consider alternative models, such as structural disorder or missing elements. Understanding the effects of disorder and defects is essential for differentiating between genuinely novel structures and those arising from imperfections. Incorporating these considerations into material characterization can provide critical guidance in defining exploration boundaries.

There is an essential balance between data collection speed and quality. When structural information is paramount, prioritizing quality over speed is essential to avoid misinterpretation, which could derail the discovery process. For example, in HTC, the balance between precision and speed is achieved by adequately selecting the computational method and limiting the dimension of the cluster models.<sup>34</sup> Although it is evident that methods with prohibitive resource scaling are unsuitable for large-scale screening, it is critical in HT catalysis to prioritize data quality over computational speed. This ensures that the resulting models can capture the underlying chemical

complexity and preserve the integrity and predictive power of the analysis.

Therefore, devoting HTE and HTC resources to investigate crystalline structures plays a relevant role in exploring regions with the potential for discovering new catalysts, even if these are model systems that may not fully capture the complexity of industrial-scale processes. During HT catalyst activity screening, it is worth noting that using small-scale testing and characterization rarely guarantees success in real-world conditions. Nonetheless, HTE provides a unique advantage by enabling the rapid pre-screening of vast experimental spaces using minimal material and time, thereby guiding more targeted optimization efforts at larger scales. While miniaturized setups may not capture all kinetic or thermodynamic aspects relevant for scale-up, they provide valuable early-stage insight that accelerates and informs subsequent development. However, relatively few studies report systematic follow-up even at the laboratory scale, reflecting both logistical constraints and the prevailing gap between screening platforms and scale-up infrastructure. Despite this limitation, the ability of HTE to rapidly triage large sets of possibilities enables more efficient use of experimental resources. Notable efforts toward bridging this gap include the work by Ruan *et al.*,<sup>61</sup> where large language model (LLM)-powered agents were integrated with automated synthesis and testing platforms to support a complete workflow from literature mining through kinetic analysis and optimization to scale-up, for various catalytic processes. Such end-to-end approaches highlight the growing potential for routine translation of HTE findings into scalable technologies.<sup>62</sup> To fully realize this potential, it is essential to establish consistent protocols for HT experiments, including characterization and standardized data reporting (as discussed in Section 4). These practices are crucial to enhancing reproducibility and minimizing duplication of effort across multiple and diverse research environments, significantly advancing the pace of catalysts discovery.<sup>63</sup>

## 4. Enhancing data integration and interpretation: learning from collaboration and failure

Most HT screening campaigns involve multiple steps, from library design and characterization to catalytic performance evaluation. In addition to the inherent complexity of HT design, the vast volumes of data generated at the HT speed pose significant challenges for data integration. These challenges are intensified in cross-disciplinary collaborations, where inconsistent data and metadata formats lead to misinterpretations, stalling research progress.<sup>64</sup> Consistent metadata collection in catalysis is uniquely important, as slight variations in experimental conditions can significantly alter a catalyst's structure and activity, adding further intricacies to data harmonization. Moreover, the absence of data stewardship protocols for reporting data and metadata, including critical experimental parameters—such as particle size, material preparation,



catalyst loading, temperature, and flow rate—hinders comparative studies across different experiments and scales.

Multiple databases and repositories, such as the CatApp,<sup>65</sup> Catalysis-hub,<sup>66</sup> Open Catalyst,<sup>35</sup> ioChemBD,<sup>67</sup> and NOMAD,<sup>68</sup> are publicly available, aiming to supplement in-house data resources and support data management tasks. Although the goal of these databases is to ensure that data adheres to the FAIR principles for data management—being Findable, Accessible, Interoperable, and Reusable,<sup>69</sup> this is still not always attainable in practice.<sup>51</sup> This limited practical adherence is reflected in the persistently low dataset-to-article ratio observed in chemistry, which remains below the scientific average and is even lower in catalysis, where approximately one dataset is available per one hundred articles, indicating that experimental and computational data are still rarely shared at scale.<sup>51</sup> One of the root causes may be primarily linked to resistance to change on the part of researchers and institutions, resulting in long-standing inconsistencies in database design and repository standards. Additionally, HTC datasets frequently omit key experimental conditions such as reaction temperature, pH, solvent, or chemical environment for simplicity.<sup>36,70–72</sup> The lack of uniformity in the database causes difficulties in applying AI tools to advance the discovery process through complex analysis.<sup>73,74</sup> Thus, standardized datasets cannot only improve the training of machine learning (ML) models but also deepen our understanding of how experimental variables influence material performance and catalytic activity.

In the past decades, scientific research has also shifted towards projects that rely on multi-institutional efforts. This often requires scientists to share and combine datasets from multiple sources while ensuring consistency to land research efforts into meaningful insights.<sup>75,76</sup> However, the lack of standardized protocols across institutions and the complexity of diverse experimental setups introduces additional challenges. This becomes particularly complicated when data is obtained from various equipment, and each is operated by software developed by different vendors, often resulting in fragmented and difficult-to-integrate data. Although several efforts have been made to overcome these barriers,<sup>77,78</sup> it is crucial to reassess the data collection mechanism as a community. This represents an essential step towards making data machine- and human-readable to drive discovery, which also accounts for AI agents.<sup>79</sup> Thus, the discussion around data integration and interpretation requires a focus on (1) technical approaches, such as new algorithms or methodologies for combining diverse data types, and (2) pragmatic organizational strategies to improve workflows and communication across disciplines and equipment. To this end, interdisciplinary collaborations involving data scientists in research teams and enhancing data stewardship educational programs are critical in advancing effective and collaborative data analysis.

There is no doubt that the conjunction of HTE and HTC provides enormous opportunities to transform the field of catalysis. The integration of HTC with HTE offers synergistic advantages: HTC can serve both as a pre-screening tool, guiding the design of focused HTE campaigns, and as a post-analysis aid, helping to rationalize trends and mechanistic insights

from experimental data. At the symposium, this was exemplified by Ser and Hao *et al.*, which demonstrated the ligand-dependence of palladium-catalyzed protodeboronation.<sup>40</sup> The HTE setup facilitated the parallel reaction of 27 unique phosphine ligands and their respective deboronation yields were subsequently analyzed *via* HPLC in one workflow. HTC *via* DFT was used to investigate the proposed reaction mechanism containing 23 intermediates and transition states for all 27 ligands and determining that bulkier ligands favor the formation of an unstable post-transmetalation product that subsequently undergoes facile protodeboronation with water, supporting the experimental observations. This example demonstrates how the synergy between HTE and HTC can facilitate the elucidation of more general trends. This interplay between simulation and experimentation is also central to the development of autonomous, closed-loop platforms for catalytic discovery.

HT research unifies aspects of many traditionally disparate fields, such as synthesis, analytical chemistry, statistics, machine learning, robotics, cheminformatics, and quantum chemistry, to name a few. Properly capitalizing on these opportunities requires scientists who have been trained to take advantage of them. For education, both for new generations of scientists as well as those with long experience, it follows that a certain amount of technical knowledge outside of one's core competency domain is beneficial. This holds true both if one is working in comparative isolation or as part of a large, multi-disciplinary team. For team leads, the ability to translate between experts in different knowledge domains becomes paramount to fostering good teamwork. Especially large organizations, such as industrial research laboratories, will need to adjust hiring and internal upskilling efforts to facilitate cross-disciplinary communication, particularly accounting for a gradual shift to a more significant share of digital skill sets. Both research managers and researchers must adopt an open-minded yet practical approach to the rapidly advancing fields within the digital space. Unrealistic expectations and the indiscriminate use of AI for the sake of following the latest trend have the potential to undermine trust, while also leading to a significant waste of resources, including time, funding, and energy, particularly in large-scale screening efforts.

Another critical point in the discussion of data integration occurs when HT catalysis campaigns involving theory and experiment are considered. Here, one key challenge is the lack of knowledge of the atomistic structure of a catalyst under catalytic conditions, which is in line with the discussion covered in Section 3. This often results in simplified and not necessarily representative catalyst models.<sup>57,80–82</sup> The successful integration of HTC into catalysis research frequently requires multiple iterations and feedback from experimental results to validate further or refine existing hypotheses around the plausible active sites and corresponding catalytic behavior. Thus, planning how metadata should be stored in HTC campaigns is not always straightforward, as this requires building prior knowledge and development time from both computational models and experiments.



A frequently overlooked opportunity in scientific research arises when model predictions and experimental data do not align. Rather than discarding such results or forcing models to fit, these discrepancies should be seen as valuable opportunities for new insights. The same applies to results that do not yield the expected material or catalytic activity; such data is often excluded from publications or patents as it may be perceived as a distraction. Such results can lead to a better understanding of what drives or inhibits a catalytic system. In this sense, learning from 'negative' results is crucial to advance scientific research. Thus, creating a culture of documenting negative results is mandatory. To this end, journals and funding agencies should promote and stimulate a change in the paradigm of publication bias and incentivize proper documentation of comprehensive datasets, avoiding selective reporting that shapes findings to fit a preferred narrative.<sup>83</sup> Novel publication forms such as tutorials and articles that provide quick incremental updates, such as the Commits article recently released by Digital Discovery,<sup>84</sup> are much-needed for the effective co-development of software-hardware workflows.

## 5. Optimizing high-throughput methods: insights from experience and innovation

The optimization of HT methods relies on the design of efficient and well-integrated experimental and computational frameworks. In this regard, HT workflows should seamlessly connect data from materials' synthesis, characterization, and catalytic performance with a feedback loop.<sup>75</sup> For example, a practical HTE framework should enable the rapid screening of materials, while an effective HTC should accurately predict materials' properties before testing and optimizing reaction conditions. The integration of both ideally should result in a parallel and iterative workflow where computational predictions inform the experimental design and *vice versa*, which accelerates the identification of new candidates. By systematically bridging these frameworks, we can establish a cycle that minimizes redundancy and maximizes discovery efficiency.

Similarly, leveraging historical data is essential to enhancing the precision and efficacy of HT workflows.<sup>76</sup> Existing datasets, including those from less successful or "negative" experiments, are invaluable for elucidating material properties, optimizing reaction conditions, and refining methodologies. The preservation and conditioning of historical data and metadata can guide researchers in identifying patterns, preventing redundant experiments, and enhancing predictive models. Recent advances in literature mining, particularly those leveraging LLM agents,<sup>85</sup> have shown that information from historical sources can be systematically extracted and structured, offering a cost-effective and scalable way to expand training datasets and reveal trends that would otherwise be overlooked. Innovation in HT research requires both adapting existing technologies and creating novel ones. For instance, emerging techniques like AI-driven optimization algorithms and automated synthesis platforms can streamline discovery.<sup>15,86,87</sup> Along these lines,

integrating negative results is critical for diversifying AI datasets. As mentioned in Section 4, most databases document and published literature predominantly report successful experiments, while failed attempts or non-productive conditions are underrepresented due to longstanding publication biases. This lack of negative data leads to biased training sets for machine learning models, including LLMs,<sup>88–92</sup> which are often fine-tuned on scientific texts. Although these models are not trained on nonsensical or erroneous data, their exposure is limited to what is reported, typically positive results. This can result in overly optimistic or incomplete suggestions when LLMs are used in hypothesis generation or retrosynthetic planning. Still, despite lacking fine-grained experimental details, historical datasets offer valuable patterns and trends that support both human and AI-driven discovery.

Adopting effective workflows and datasets must go hand-in-hand with developing new tools, such as next-gen HT spectroscopic techniques or reactors, which can provide more precise data and enable new insights into molecular and material behavior under different conditions. One should also democratize HT research by combining these principles with emerging technologies such as AI and robotics to optimize HT research further. This can only be accomplished by openly sharing data, methodologies, and tools so that researchers worldwide can easily access and build from well-funded research institutions, bringing new perspectives and applications to the table and new solutions. The seamless integration of new automated data analysis and human expertise ensures that each experimental run contributes to creating an incremental and continuous learning feedback loop.

Finally, a crucial element in HT optimization is the commitment to continuous, critical thinking to transition from high-throughput to smart-throughput experimentation.<sup>93</sup> Artificial intelligence and machine learning approaches are increasingly integral to both HTE and HTC workflows. Bayesian optimization and active learning can efficiently guide experimental campaigns, ML interatomic potentials accelerate atomistic simulations at near-DFT accuracy, and large-language models can mine literature to reveal catalytic patterns. However, realizing their full potential requires improving data diversity and reliability, refining uncertainty quantification, and developing tighter integration between computational and experimental pipelines. Researchers must rigorously assess and refine experimental protocols, challenge assumptions, and embrace new tools and methodologies for improved data interpretation. This includes reevaluating experimental setups, adjusting protocols, and adopting new programming skills alongside knowledge of statistical or machine-learning models to enhance data interpretation. Embracing feedback from successes and failures allows for iterative growth in HT research, keeping in mind that unexpected results can spark new hypotheses, ideas, and discoveries.

## 6. Outlook

High-throughput catalysis stands at a transformative crossroads. As the demand for sustainable energy, efficient chemical



processes, and advanced materials increases to support the immediate and long-term changes needed, the catalysis community must move beyond current bottlenecks, aiming for bold innovation and strategic collaboration. This requires the community to move beyond incremental improvements; it calls for transformative shifts in how catalytic research is envisioned, executed, and analyzed. Through insights from our workshop speakers and panel discussions, we have identified five critical focus areas for driving this transformation.

### 6.1. Embracing digital transformation in catalysis

The development of enhanced automated workflows goes hand-in-hand with integration platforms that combine synthesis, characterization, and data analysis—in such platforms, experimental and computational data directly inform machine learning models and redefine efficiency and precision. We expect the field to go smoothly beyond traditional procedures by closing the HT catalysis loop, balancing human/computer interventions, and removing the HT characterization time-scale mismatch. A wilder vision of this future would expect to achieve full autonomy on HT platforms with systems capable of fully designing, executing, and refining experiments without many interruptions and interventions, drastically accelerating discovery and innovation and then focusing on incoming challenges such as scalability and commercialization. It should also be recognized that the study of heterogeneous catalysis is intrinsically multiscale and influenced by numerous interdependent factors, including surface reaction networks, and mass and energy transfer phenomena. Incorporating such complexities into digital and data-driven frameworks remains a key challenge for the field.

### 6.2. Harnessing advanced data analytics for deeper insights

While high-throughput methods generate massive, complex datasets, the real challenge will always remain the extraction of meaningful insights. Using advanced analytics tools for data integration and machine learning offers immense potential for identifying hidden patterns and forecasting catalytic behaviors. Despite the evident progress in these fields, inconsistent data quality and limited accessibility continue to block the path to discoveries. A clear example is that besides the advances in natural language processing in mining scientific literature, obstacles such as non-digitized data, paywalls, and unstructured reporting standards hinder comprehensive integration. We encourage a shift towards open-access, machine-readable databases with standardized data and metadata, which is critical in removing these hurdles. Nevertheless, it should be noted that while there is widespread aspiration for open data, its implementation raises legitimate concerns around intellectual property, particularly in fields like catalysis, where data sharing may conflict with commercial interests and shareholder obligations. This tension also extends to academia, which finds itself at a crossroads: on one hand, promoting transparency and collaboration; on the other, being increasingly pushed to “valorize” its outputs. Acknowledging this dichotomy between idealism and economic pragmatism, we believe that advancing

open data initiatives will require new incentive structures, legal frameworks, and cultural shifts that make data sharing both safe and valuable for all stakeholders. Moreover, while current ML models prioritize incremental improvements within known chemical spaces, their ability to uncover non-obvious, high-performing candidates is expected to grow as datasets and algorithms evolve, as recently discussed by Schrier *et al.*<sup>94</sup> As a result, the boundary between optimization and discovery may increasingly blur, enhancing the potential of ML-HTE to surface exceptional catalysts hidden in complex design landscapes. To achieve this, the field should move toward the development of multimodal foundation models capable of learning transferable representations from heterogeneous data—such as atomic structures, spectra, and text—thereby bridging the gap between *in silico* predictions and experimental reality.<sup>58</sup>

### 6.3. Promoting standardization, benchmarking, and reproducibility

To benefit from advanced data analytics, researchers must adopt global efforts to standardize data collection methods, experimental procedures, and characterization protocols to address existing reproducibility challenges. For example, detailed reporting templates should capture critical synthesis conditions (*e.g.*, temperature, reagent ratios, pH), catalytic performance metrics (*e.g.*, activity, selectivity, stability), and catalyst/system characteristics (*e.g.*, particle size, surface area, reactor design). By adopting this approach, the community will establish much-needed benchmarks for data quality, experimental protocols, and characterization standards, ensuring consistency across the field. We invite the community to embrace innovative publication formats, such as data journals and platforms for negative results. These approaches will enable new insights into material performance and enhance transparency and scientific rigor in catalyst research.

### 6.4. Fostering a unified, interdisciplinary research landscape

Groundbreaking discoveries in catalysis often occur within cross-disciplinary collaboration. Catalysis involves many disciplines, and breakthroughs will occur when chemists, engineers, data scientists, and theorists work together to solve complex problems. A key to successful collaboration is establishing a shared scientific vocabulary—for instance, harmonizing terms like “descriptor” in chemistry with “label” in computer science enhances clarity and knowledge transfer. Thus, fostering an interdisciplinary research landscape will require creating and participating in interdisciplinary consortia with shared research facilities and collaborative training initiatives to solidify this approach.

### 6.5. Nurturing the next generation of catalysis researchers

As high-throughput tools become more sophisticated, it becomes increasingly critical for training programs to equip scientists with the necessary skills to meet this challenge. Today, catalysis researchers are required to possess knowledge and skills beyond traditional chemistry and materials science,



including data science, automation, and computational modeling. Thus, investing in educational programs and professional development initiatives is mandatory to ensure researchers can leverage new technologies to their fullest potential. It is also crucial to reassure emerging scientists that advanced technologies are not here to constrain their scientific development or creativity. Instead, they should be regarded as powerful tools that free them from tedious routine tasks, enabling deeper intellectual exploration.

Beyond these overarching directions, it is important to recognize that catalysis presents a unique set of challenges distinct from those in broader chemistry or materials science. Catalytic systems are inherently dynamic, operating under nonequilibrium conditions and involving complex reaction networks, multiscale transport phenomena, and evolving active sites that are difficult to capture experimentally or computationally. Furthermore, bridging the gap between model and real catalysts—across pressures, timescales, and reactor environments—remains a fundamental obstacle to predictive understanding. The interplay between catalyst structure, reaction mechanism, and process conditions also introduces feedback loops that complicate optimization efforts. Addressing these catalysis-specific complexities will be fundamental for realizing the transformative potential outlined above, ensuring that high-throughput, data-driven approaches deliver meaningful insights into real-world catalytic behavior.

## Author contributions

J. G. V. and V. B. conceptualized, organized, wrote the first draft, and edited the manuscript. T. I. contributed to writing the first draft and edited the manuscript. A. A. -G., E. D., O. K. F., H. J. K., P. M. M., S. M., J. R., and A. R. S. edited the manuscript. All authors have read and agreed to the published version of the manuscript.

## Conflicts of interest

There are no conflicts to declare.

## Data availability

This Opinion article does not involve any data or code. Accordingly, there are no datasets or software to share or cite. All ideas and conclusions presented herein are based on open discussions among the co-authors in the context of the symposium held at the ACS Fall 2024 meeting.

## Acknowledgements

The authors thank Prof. Christopher J. Cramer for useful discussions and the Material Discovery Research Institute from Underwriter Laboratories (UL) Research Institutes, Acceleration Consortium, and Digital Discovery for sponsoring the event. O.K.F. gratefully acknowledges support from the Catalyst Design for Decarbonization Center, an Energy Frontier Research Center funded by the U.S. Department of Energy,

Office of Science, Basic Energy Sciences (award number DE-SC0023383). A.A.-G. and V. B. acknowledge the University of Toronto's Acceleration Consortium, which receives funding from the CFREF-2022-00042 Canada First Research Excellence Fund.

## References

- 1 B. Burger, P. M. Maffettone, V. V. Gusev, C. M. Aitchison, Y. Bai, X. Wang, X. Li, B. M. Alston, B. Li, R. Clowes, *et al.*, A mobile robotic chemist, *Nature*, 2020, **583**(7815), 237–241, DOI: [10.1038/s41586-020-2442-2](https://doi.org/10.1038/s41586-020-2442-2).
- 2 D. D. Devore and R. M. Jenkins, Impact of High Throughput Experimentation on Homogeneous Catalysis Research, *Comments Inorg. Chem.*, 2014, **34**(1–2), 17–41, DOI: [10.1080/02603594.2014.947027](https://doi.org/10.1080/02603594.2014.947027).
- 3 A. Hashemi, S. Bougueroua, M.-P. Gaigeot and E. A. Pidko, HiREX: High-Throughput Reactivity Exploration for Extended Databases of Transition-Metal Catalysts, *J. Chem. Inf. Model.*, 2023, **63**(19), 6081–6094, DOI: [10.1021/acs.jcim.3c00660](https://doi.org/10.1021/acs.jcim.3c00660).
- 4 K. E. McCullough, D. S. King, S. P. Chheda, M. S. Ferrandon, T. A. Goetjen, Z. H. Syed, T. R. Graham, N. M. Washton, O. K. Farha, L. Gagliardi, *et al.*, High-Throughput Experimentation, Theoretical Modeling, and Human Intuition: Lessons Learned in Metal–Organic–Framework-Supported Catalyst Design, *ACS Cent. Sci.*, 2023, **9**(2), 266–276, DOI: [10.1021/acscentsci.2c01422](https://doi.org/10.1021/acscentsci.2c01422).
- 5 H. M. Geysen, R. H. Meloen and S. J. Barteling, Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid, *Proc. Natl. Acad. Sci. U. S. A.*, 1984, **81**(13), 3998–4002, DOI: [10.1073/pnas.81.13.3998](https://doi.org/10.1073/pnas.81.13.3998).
- 6 A. Holzwarth, H.-W. Schmidt and W. F. Maier, Detection of Catalytic Activity in Combinatorial Libraries of Heterogeneous Catalysts by IR Thermography, *Angew. Chem., Int. Ed.*, 1998, **37**(19), 2644–2647, DOI: [10.1002/\(SICI\)1521-3773\(19981016\)37:19<2644::AID-ANIE2644>3.0.CO;2-#](https://doi.org/10.1002/(SICI)1521-3773(19981016)37:19<2644::AID-ANIE2644>3.0.CO;2-#).
- 7 W. F. Maier, K. Stöwe and S. Sieg, Combinatorial and High-Throughput Materials Science, *Angew. Chem., Int. Ed.*, 2007, **46**(32), 6016–6067, DOI: [10.1002/anie.200603675](https://doi.org/10.1002/anie.200603675).
- 8 G. Tom, S. P. Schmid, S. G. Baird, Y. Cao, K. Darvish, H. Hao, S. Lo, S. Pablo-García, E. M. Rajaonson, M. Skreta, *et al.*, Self-Driving Laboratories for Chemistry and Materials Science, *Chem. Rev.*, 2024, **124**(16), 9633–9732, DOI: [10.1021/acs.chemrev.4c00055](https://doi.org/10.1021/acs.chemrev.4c00055).
- 9 R. Pollice, G. dos Passos Gomes, M. Aldeghi, R. J. Hickman, M. Krenn, C. Lavigne, M. Lindner-D'Addario, A. Nigam, C. T. Ser, Z. Yao, *et al.*, Data-Driven Strategies for Accelerated Materials Design, *Acc. Chem. Res.*, 2021, **54**(4), 849–860, DOI: [10.1021/acs.accounts.0c00785](https://doi.org/10.1021/acs.accounts.0c00785).
- 10 *Accelerated Materials Discovery - How to Use Artificial Intelligence to Speed Up Development*, ed. P. De Luna, De Gruyter Brill, 2022, DOI: [10.1515/9783110738087](https://doi.org/10.1515/9783110738087).
- 11 E. O. Pyzer-Knapp, J. W. Pitera, P. W. J. Staar, S. Takeda, T. Laino, D. P. Sanders, J. Sexton, J. R. Smith and A. Curioni, Accelerating materials discovery using artificial



- intelligence, high performance computing and robotics, *npj Comput. Mater.*, 2022, **8**(1), 84, DOI: [10.1038/s41524-022-00765-z](https://doi.org/10.1038/s41524-022-00765-z).
- 12 S. Back, A. Aspuru-Guzik, M. Ceriotti, G. Gryn'ova, B. Grzybowski, G. H. Gu, J. Hein, K. Hippalgaonkar, R. Hormázabal, Y. Jung, *et al.*, Accelerated chemical science with AI, *Digital Discovery*, 2024, **3**(1), 23–33, DOI: [10.1039/D3DD00213F](https://doi.org/10.1039/D3DD00213F).
  - 13 A. Nandy, C. Duan, M. G. Taylor, F. Liu, A. H. Steeves and H. J. Kulik, Computational Discovery of Transition-metal Complexes: From High-throughput Screening to Machine Learning, *Chem. Rev.*, 2021, **121**(16), 9927–10000, DOI: [10.1021/acs.chemrev.1c00347](https://doi.org/10.1021/acs.chemrev.1c00347).
  - 14 A. Hjorth Larsen, J. Jørgen Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, *et al.*, The atomic simulation environment—a Python library for working with atoms, *J. Phys.: Condens. Matter*, 2017, **29**(27), 273002, DOI: [10.1088/1361-648X/aa680e](https://doi.org/10.1088/1361-648X/aa680e).
  - 15 J. P. Unsleber, H. Liu, L. Talirz, T. Weymuth, M. Morchen, A. Grofe, D. Wecker, C. J. Stein, A. Panyala and B. Peng, High-throughput ab initio reaction mechanism exploration in the cloud with automated multi-reference validation, *J. Chem. Phys.*, 2023, **158**, 084803, DOI: [10.1063/5.0136526](https://doi.org/10.1063/5.0136526).
  - 16 S. P. Huber, S. Zoupanos, M. Uhrin, L. Talirz, L. Kahle, R. Häuselmann, D. Gresch, T. Müller, A. V. Yakutovich, C. W. Andersen, *et al.*, AiiDA 1.0, a scalable computational infrastructure for automated reproducible workflows and data provenance, *Sci. Data*, 2020, **7**(1), 300, DOI: [10.1038/s41597-020-00638-4](https://doi.org/10.1038/s41597-020-00638-4).
  - 17 E. I. Ioannidis, T. Z. H. Gani and H. J. Kulik, molSimplify: A toolkit for automating discovery in inorganic chemistry, *J. Comput. Chem.*, 2016, **37**(22), 2106–2117, DOI: [10.1002/jcc.24437](https://doi.org/10.1002/jcc.24437).
  - 18 K. Li, A. N. Rubungo, X. Lei, D. Persaud, K. Choudhary, B. DeCost, A. B. Dieng and J. Hattrick-Simpers, Probing out-of-distribution generalization in machine learning for materials, *Commun. Mater.*, 2025, **6**(1), 9, DOI: [10.1038/s43246-024-00731-w](https://doi.org/10.1038/s43246-024-00731-w).
  - 19 C. Bozal-Ginesta, S. Pablo-García, C. Choi, A. Tarancón and A. Aspuru-Guzik, Developing machine learning for heterogeneous catalysis with experimental and computational data, *Nat. Rev. Chem.*, 2025, **9**(9), 601–616, DOI: [10.1038/s41570-025-00740-4](https://doi.org/10.1038/s41570-025-00740-4).
  - 20 M. Abolhasani and E. Kumacheva, The rise of self-driving labs in chemical and materials sciences, *Nat. Synth.*, 2023, **2**(6), 483–492, DOI: [10.1038/s44160-022-00231-0](https://doi.org/10.1038/s44160-022-00231-0).
  - 21 G. dos Passos Gomes, R. Pollice and A. Aspuru-Guzik, Navigating through the Maze of Homogeneous Catalyst Design with Machine Learning, *Trends Chem.*, 2021, **3**(2), 96–110, DOI: [10.1016/j.trechm.2020.12.006](https://doi.org/10.1016/j.trechm.2020.12.006).
  - 22 M. Erdem Günay and R. Yildırım, Recent advances in knowledge discovery for heterogeneous catalysis using machine learning, *Catal. Rev.*, 2020, **63**, 120–164.
  - 23 K. McCullough, T. Williams, K. Mingle, P. Jamshidi and J. Lauterbach, High-throughput experimentation meets artificial intelligence: a new pathway to catalyst discovery, *Phys. Chem. Chem. Phys.*, 2020, **22**(20), 11174–11196, DOI: [10.1039/D0CP00972E](https://doi.org/10.1039/D0CP00972E).
  - 24 B. R. Goldsmith, J. Esterhuizen, J.-X. Liu, C. J. Bartel and C. Sutton, Machine learning for heterogeneous catalyst design and discovery, *AIChE J.*, 2018, **64**(7), 2311–2323, DOI: [10.1002/aic.16198](https://doi.org/10.1002/aic.16198).
  - 25 J. Benavides-Hernández and F. Dumeignil, From Characterization to Discovery: Artificial Intelligence, Machine Learning and High-Throughput Experiments for Heterogeneous Catalyst Design, *ACS Catal.*, 2024, **14**(15), 11749–11779, DOI: [10.1021/acscatal.3c06293](https://doi.org/10.1021/acscatal.3c06293).
  - 26 N. H. Angello, V. Rathore, W. Beker, A. Wołos, E. R. Jira, R. Roszak, T. C. Wu, C. M. Schroeder, A. Aspuru-Guzik, B. A. Grzybowski, *et al.*, Closed-loop optimization of general reaction conditions for heteroaryl Suzuki-Miyaura coupling, *Science*, 2022, **378**(6618), 399–405, DOI: [10.1126/science.adc8743](https://doi.org/10.1126/science.adc8743).
  - 27 C. Chen and S. P. Ong, A universal graph deep learning interatomic potential for the periodic table, *Nat. Comput. Sci.*, 2022, **2**(11), 718–728, DOI: [10.1038/s43588-022-00349-3](https://doi.org/10.1038/s43588-022-00349-3).
  - 28 D. P. Kovács, I. Batatia, E. S. Arany and G. Csányi, Evaluation of the MACE force field architecture: From medicinal chemistry to materials science, *J. Chem. Phys.*, 2023, **159**(4), 044118, DOI: [10.1063/5.0155322](https://doi.org/10.1063/5.0155322).
  - 29 J. G. Vitillo, V. Bernales, A. Singh, O. K. Farha, S. Miller, J. Reddel, A. Aspuru-Guzik, P. M. Margl, E. Dskocil and H. J. Kulik, Panel discussion: Accelerating catalytic advancements through the precision of high-throughput experiments and calculations, [CATL] Division of Catalysis Science & Technology, American Chemical Society Denver, Colorado, USA, 2024, 2024.
  - 30 R. E. Oesper, Alwin Mittasch, *J. Chem. Educ.*, 1948, **25**(10), 531, DOI: [10.1021/ed025p531](https://doi.org/10.1021/ed025p531).
  - 31 Google Patents, <https://patents.google.com/> (accessed 18 October 2024).
  - 32 L. Buglioni, F. Raymenants, A. Slattery, S. D. A. Zondag and T. Noël, Technological Innovations in Photochemistry for Organic Synthesis: Flow Chemistry, High-Throughput Experimentation, Scale-up, and Photoelectrochemistry, *Chem. Rev.*, 2022, **122**(2), 2752–2906, DOI: [10.1021/acs.chemrev.1c00332](https://doi.org/10.1021/acs.chemrev.1c00332).
  - 33 G. Lyall-Brookes, A. C. Padgham and A. G. Slater, Flow chemistry as a tool for high throughput experimentation, *Digital Discovery*, 2025, **4**(9), 2364–2400, DOI: [10.1039/D5DD00129C](https://doi.org/10.1039/D5DD00129C).
  - 34 F. Ju, X. Wei, L. Huang, A. J. Jenkins, L. Xia, J. Zhang, J. Zhu, H. Yang, B. Shao, P. Dai, *et al.*, Acceleration without Disruption: DFT Software as a Service, *arXiv*, 2024, preprint arXiv:2406.11185, DOI: [10.48550/arXiv.2406.11185](https://doi.org/10.48550/arXiv.2406.11185).
  - 35 L. Chanussot, A. Das, S. Goyal, T. Lavril, M. Shuaibi, M. Riviere, K. Tran, J. Heras-Domingo, C. Ho, W. Hu, *et al.*, Open Catalyst 2020 (OC20) Dataset and Community Challenges, *ACS Catal.*, 2021, **11**(10), 6059–6072, DOI: [10.1021/acscatal.0c04525](https://doi.org/10.1021/acscatal.0c04525).
  - 36 J. Abed, J. Kim, M. Shuaibi, B. Wander, B. Duijf, S. Mahesh, H. Lee, V. Gharakhanyan, S. Hoogland, E. Irtem, *et al.*, Open Catalyst Experiments 2024 (OCx24): Bridging Experiments



- and Computational Models, *arXiv*, 2024, preprint, arXiv:2411.11783, DOI: [10.48550/arXiv.2411.11783](https://doi.org/10.48550/arXiv.2411.11783).
- 37 Y. Su, X. Wang, Y. Ye, Y. Xie, Y. Xu, Y. Jiang and C. Wang, Automation and machine learning augmented by large language models in a catalysis study, *Chem. Sci.*, 2024, **15**(31), 12200–12233, DOI: [10.1039/D3SC07012C](https://doi.org/10.1039/D3SC07012C).
- 38 A. Aspuru-Guzik, Suzuki reaction: A gift that keeps on giving, in *[CATL] Division of Catalysis Science & Technology*, American Chemical Society, Denver, Colorado, 2024.
- 39 F. Strieth-Kalthoff, H. Hao, V. Rathore, J. Derasp, T. Gaudin, N. H. Angello, M. Seifrid, E. Trushina, M. Guy, J. Liu, *et al.*, Delocalized, asynchronous, closed-loop discovery of organic laser emitters, *Science*, 2024, **384**(6697), eadk9227, DOI: [10.1126/science.adk9227](https://doi.org/10.1126/science.adk9227).
- 40 C. T. Ser, H. Hao, S. Pablo-García, K. Jorner, S. Li, R. Pollice and A. Aspuru-Guzik, Bulky phosphine ligands promote palladium-catalyzed protodeboronation, *J. Am. Chem. Soc.*, 2025, **147**(47), 43884–43901, DOI: [10.1021/jacs.5c14153](https://doi.org/10.1021/jacs.5c14153).
- 41 J. Abed, J. Heras-Domingo, R. Y. Sanspeur, M. Luo, W. Alnoush, D. M. Meira, H. Wang, J. Wang, J. Zhou, D. Zhou, *et al.*, Pourbaix Machine Learning Framework Identifies Acidic Water Oxidation Catalysts Exhibiting Suppressed Ruthenium Dissolution, *J. Am. Chem. Soc.*, 2024, **146**(23), 15740–15750, DOI: [10.1021/jacs.4c01353](https://doi.org/10.1021/jacs.4c01353).
- 42 D. Desai, T. Karin and J. L. Baker Rivest, High-throughput experimentation for catalysis, *US Pat.*, US11981563B2, 2024.
- 43 X. Caldentey and E. Romero, High-Throughput Experimentation as an accessible technology for academic organic chemists in Europe and beyond, *ChemRxiv*, 2022, preprint, DOI: [10.26434/chemrxiv-2022-wf3kf](https://doi.org/10.26434/chemrxiv-2022-wf3kf).
- 44 S. N. Steinmann, A. Hermawan, M. Bin Jassar and Z. W. Seh, Autonomous high-throughput computations in catalysis, *Chem Catal.*, 2022, **2**(5), 940–956, DOI: [10.1016/j.cheecat.2022.02.009](https://doi.org/10.1016/j.cheecat.2022.02.009).
- 45 J. K. Nørskov, T. Bligaard, J. Rossmeisl and C. H. Christensen, Towards the computational design of solid catalysts, *Nat. Chem.*, 2009, **1**(1), 37–46, DOI: [10.1038/nchem.121](https://doi.org/10.1038/nchem.121).
- 46 J. G. Vitillo, C. J. Cramer and L. Gagliardi, Multireference Methods are Realistic and Useful Tools for Modeling Catalysis, *Isr. J. Chem.*, 2022, **62**(1–2), e202100136, DOI: [10.1002/ijch.202100136](https://doi.org/10.1002/ijch.202100136).
- 47 V. Bernales and R. D. Froese, Rhodium catalyzed hydroformylation of olefins, *J. Comput. Chem.*, 2019, **40**(2), 342–348, DOI: [10.1002/jcc.25605](https://doi.org/10.1002/jcc.25605).
- 48 J. P. Janet, S. Ramesh, C. Duan and H. J. Kulik, Accurate Multiobjective Design in a Space of Millions of Transition Metal Complexes with Neural-Network-Driven Efficient Global Optimization, *ACS Cent. Sci.*, 2020, **6**(4), 513–524, DOI: [10.1021/acscentsci.0c00026](https://doi.org/10.1021/acscentsci.0c00026).
- 49 J. G. Vitillo, C. C. Lu, A. Bhan and L. Gagliardi, Comparing the reaction profiles of single-iron catalytic sites in enzymes and in reticular frameworks for methane-to-methanol oxidation, *Cell Rep. Phys. Sci.*, 2023, **4**, 101422.
- 50 S. Seritan, C. Bannwarth, B. S. Fales, E. G. Hohenstein, S. I. L. Kokkila-Schumacher, N. Luehr, J. W. Jr Snyder, C. Song, A. V. Titov, I. S. Ufimtsev, *et al.*, TeraChem: Accelerating electronic structure and ab initio molecular dynamics with graphical processing units, *J. Chem. Phys.*, 2020, **152**(22), 224110, DOI: [10.1063/5.0007615](https://doi.org/10.1063/5.0007615).
- 51 P. S. F. Mendes, N. López, S. De, R. Fushimi, R. Yildirim and A. Trunschke, Data as a Key Resource in Catalysis: A Community Account, *ChemCatChem*, 2025, **17**(24), e01226, DOI: [10.1002/cctc.202501226](https://doi.org/10.1002/cctc.202501226).
- 52 J. G. Vitillo, Best practices for transition-metal catalysts with density functional methods, *Chem Catal.*, 2025, **5**(7), 101427, DOI: [10.1016/j.cheecat.2025.101427](https://doi.org/10.1016/j.cheecat.2025.101427).
- 53 A. I. Cooper, P. Courtney, K. Darvish, M. Eckhoff, H. Fakhruddin, A. Gabrielli, A. Garg, S. Haddadin, K. Harada and J. Hein, Accelerating Discovery in Natural Science Laboratories with AI and Robotics: Perspectives and Challenges from the 2024 IEEE ICRA Workshop, *arXiv*, 2025, preprint, arXiv:2501.06847, DOI: [10.48550/arXiv.2501.06847](https://doi.org/10.48550/arXiv.2501.06847), Yokohama, Japan.
- 54 N. Orouji, J. A. Bennett, R. B. Canty, L. Qi, S. Sun, P. Majumdar, C. Liu, N. López, N. M. Schweitzer, J. R. Kitchin, *et al.*, Autonomous catalysis research with human–AI–robot collaboration, *Nat. Catal.*, 2025, **8**(11), 1135–1145, DOI: [10.1038/s41929-025-01430-6](https://doi.org/10.1038/s41929-025-01430-6).
- 55 S. Lo, S. G. Baird, J. Schrier, B. Blaiszik, N. Carson, I. Foster, A. Aguilar-Granda, S. V. Kalinin, B. Maruyama, M. Politi, *et al.*, Review of low-cost self-driving laboratories in chemistry and materials science: the “frugal twin” concept, *Digital Discovery*, 2024, **3**(5), 842–868, DOI: [10.1039/D3DD00223C](https://doi.org/10.1039/D3DD00223C).
- 56 C. Bozal-Ginesta, J. Sirvent, G. Cordaro, S. Fearn, S. Pablo-García, F. Chiabrera, C. Choi, L. Laa, M. Núñez, A. Cavallaro, *et al.*, Performance Prediction of High-Entropy Perovskites La<sub>0.8</sub>Sr<sub>0.2</sub>MnxCo<sub>y</sub>Fe<sub>z</sub>O<sub>3</sub> with Automated High-Throughput Characterization of Combinatorial Libraries and Machine Learning, *Adv. Mater.*, 2024, **36**(50), 2407372, DOI: [10.1002/adma.202407372](https://doi.org/10.1002/adma.202407372).
- 57 L. Gagliardi, Computational rational high-throughput generation of novel MOF class based on iron-sulfur cluster, in *[CATL] Division of Catalysis Science & Technology*, American Chemical Society, Denver, Colorado, USA, 2024.
- 58 H. Xin, J. Kitchin, N. López, N. Schweitzer, N. Artrith, F. Che, L. Grabow, K. Gunasooriya, H. Kulik and T. Laino, Roadmap for transforming heterogeneous catalysis with artificial intelligence, *ChemRxiv*, 2025, DOI: [10.26434/chemrxiv-2025-chn4j](https://doi.org/10.26434/chemrxiv-2025-chn4j).
- 59 J. Leeman, Y. Liu, J. Stiles, S. B. Lee, P. Bhatt, L. M. Schoop and R. G. Palgrave, Challenges in High-Throughput Inorganic Materials Prediction and Autonomous Synthesis, *PRX Energy*, 2024, **3**(1), 011002, DOI: [10.1103/PRXEnergy.3.011002](https://doi.org/10.1103/PRXEnergy.3.011002).
- 60 P. M. Maffettone, L. Banko, P. Cui, Y. Lysogorskiy, M. A. Little, D. Olds, A. Ludwig and A. I. Cooper, Crystallography companion agent for high-throughput materials discovery, *Nat. Comput. Sci.*, 2021, **1**(4), 290–297, DOI: [10.1038/s43588-021-00059-2](https://doi.org/10.1038/s43588-021-00059-2).
- 61 Y. Ruan, C. Lu, N. Xu, Y. He, Y. Chen, J. Zhang, J. Xuan, J. Pan, Q. Fang, H. Gao, *et al.*, An automatic end-to-end



- chemical synthesis development platform powered by large language models, *Nat. Commun.*, 2024, 15(1), 10160, DOI: [10.1038/s41467-024-54457-x](https://doi.org/10.1038/s41467-024-54457-x).
- 62 A. Li, P. Cui, X. Wang, A. Fisher, L. Li and D. Cheng, The artificial intelligence-catalyst pipeline: accelerating catalyst innovation from laboratory to industry, *Front. Chem. Sci. Eng.*, 2025, 19(7), 55, DOI: [10.1007/s11705-025-2560-3](https://doi.org/10.1007/s11705-025-2560-3).
- 63 A. C. Alba-Rubio, P. Christopher, M. L. Personick and K. J. Stowers, Recommendations to standardize reporting on the synthesis of heterogeneous catalysts, *J. Catal.*, 2024, 429, 115259, DOI: [10.1016/j.jcat.2023.115259](https://doi.org/10.1016/j.jcat.2023.115259).
- 64 A. White, M. Gibaldi, J. Burner, R. A. Mayo and T. Woo, Alarming structural error rates in MOF databases used in data driven workflows identified via a novel metal oxidation state-based method, *J. Am. Chem. Soc.*, 2025, 147(21), 17579–17583, DOI: [10.1021/jacs.5c04914](https://doi.org/10.1021/jacs.5c04914).
- 65 J. S. Hummelshøj, F. Abild-Pedersen, F. Studt, T. Bligaard and J. K. Nørskov, CatApp: A Web Application for Surface Chemistry and Heterogeneous Catalysis, *Angew. Chem., Int. Ed.*, 2012, 51(1), 272–274, DOI: [10.1002/anie.201107947](https://doi.org/10.1002/anie.201107947).
- 66 K. T. Winther, M. J. Hoffmann, J. R. Boes, O. Mamun, M. Bajdich and T. Bligaard, Catalysis-Hub.org, an open electronic structure database for surface reactions, *Sci. Data*, 2019, 6(1), 75, DOI: [10.1038/s41597-019-0081-y](https://doi.org/10.1038/s41597-019-0081-y).
- 67 M. Álvarez-Moreno, C. de Graaf, N. López, F. Maseras, J. M. Poblet and C. Bo, Managing the Computational Chemistry Big Data Problem: The ioChem-BD Platform, *J. Chem. Inf. Model.*, 2015, 55(1), 95–103, DOI: [10.1021/ci500593j](https://doi.org/10.1021/ci500593j).
- 68 M. Scheidgen, L. Himanen, A. N. Ladines, D. Sikter, M. Nakhaee, Á. Fekete, T. Chang, A. Golparvar, J. A. Márquez and S. Brockhauser, NOMAD: A distributed web-based platform for managing materials science research data, *J. Open Source Softw.*, 2023, 8(90), 5388.
- 69 M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, *et al.*, The FAIR Guiding Principles for scientific data management and stewardship, *Sci. Data*, 2016, 3(1), 160018, DOI: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18).
- 70 J. H. Montoya, C. Grimley, M. Aykol, C. Ophus, H. Sternlicht, B. H. Savitzky, A. M. Minor, S. B. Torrisi, J. Goedjen, C.-C. Chung, *et al.*, How the AI-assisted discovery and synthesis of a ternary oxide highlights capability gaps in materials science, *Chem. Sci.*, 2024, 15(15), 5660–5673, DOI: [10.1039/D3SC04823C](https://doi.org/10.1039/D3SC04823C).
- 71 J. Montoya, AI-guided synthesis of inorganic materials, in *[CATL] Division of Catalysis Science & Technology*, American Chemical Society, Denver, Colorado, USA, 2024.
- 72 J. Abed, Towards accelerating catalyst discovery: bridging the gap between modeling, experiments, and large-scale integration, in *[CATL] Division of Catalysis Science & Technology*, American Chemical Society, Denver, Colorado, USA, 2024.
- 73 L. Takahashi, I. Miyazato and K. Takahashi, Redesigning the Materials and Catalysts Database Construction Process Using Ontologies, *J. Chem. Inf. Model.*, 2018, 58(9), 1742–1754, DOI: [10.1021/acs.jcim.8b00165](https://doi.org/10.1021/acs.jcim.8b00165).
- 74 E. A. Pfeif and K. Kroenlein, Perspective: Data infrastructure for high throughput materials discovery, *APL Mater.*, 2016, 4(5), 053203, DOI: [10.1063/1.4942634](https://doi.org/10.1063/1.4942634).
- 75 A. J. Medford, M. R. Kunz, S. M. Ewing, T. Borders and R. Fushimi, Extracting Knowledge from Data through Catalysis Informatics, *ACS Catal.*, 2018, 8(8), 7403–7429, DOI: [10.1021/acscatal.8b01708](https://doi.org/10.1021/acscatal.8b01708).
- 76 Z. Jensen, S. Kwon, D. Schwalbe-Koda, C. Paris, R. Gómez-Bombarelli, Y. Román-Leshkov, A. Corma, M. Moliner and E. A. Olivetti, Discovering Relationships between OSDAs and Zeolites through Data Mining and Generative Neural Networks, *ACS Cent. Sci.*, 2021, 7(5), 858–867, DOI: [10.1021/acscentsci.1c00024](https://doi.org/10.1021/acscentsci.1c00024).
- 77 D. Kuehl, AC Webworks: An XML Model for Analytical Instrument Data, *Anal. Chem.*, 2003, 75(5), 125A–127A, DOI: [10.1021/ac031260c](https://doi.org/10.1021/ac031260c).
- 78 B. A. Schäfer, D. Poetz and G. W. Kramer, Documenting Laboratory Workflows Using the Analytical Information Markup Language, *J. Assoc. Lab. Autom.*, 2004, 9(6), 375–381, DOI: [10.1016/j.jala.2004.10.003](https://doi.org/10.1016/j.jala.2004.10.003).
- 79 J. West, Harnessing the power of data: the importance of quality, accessibility, and design in AI and ML, in *[CATL] Division of Catalysis Science & Technology*, American Chemical Society, Denver, Colorado, USA, 2024.
- 80 K. Reuter, C. P. Plaisance, H. Oberhofer and M. Andersen, Perspective: On the active site model in computational catalyst screening, *J. Chem. Phys.*, 2017, 146(4), 040901, DOI: [10.1063/1.4974931](https://doi.org/10.1063/1.4974931).
- 81 A. Darù, J. S. Anderson, D. M. Proserpio and L. Gagliardi, Symmetry is the Key to the Design of Reticular Frameworks, *Adv. Mater.*, 2025, 37(52), 2414617, DOI: [10.1002/adma.202414617](https://doi.org/10.1002/adma.202414617).
- 82 A. Darù, Computational rational high-throughput generation of novel MOF class based on iron-sulfur cluster, in *[CATL] Division of Catalysis Science & Technology*, American Chemical Society, Denver, Colorado, USA, 2024.
- 83 S. L. Scott, T. B. Gunnoe, P. Fornasiero and C. M. Crudden, To Err is Human; To Reproduce Takes Time, *ACS Catal.*, 2022, 12(6), 3644–3650, DOI: [10.1021/acscatal.2c00967](https://doi.org/10.1021/acscatal.2c00967).
- 84 A. Aspuru-Guzik, J. E. Hein and J. Schrier, Commit: Mini article for dynamic reporting of incremental improvements to previous scholarly work, *Digital Discovery*, 2024, 4, 301–302, DOI: [10.1039/D4DD90053G](https://doi.org/10.1039/D4DD90053G).
- 85 T. M. Pruyun, A. Aswad, S. T. Khan, R. Black and S. M. Moosavi, MOF-ChemUnity: Unifying metal-organic framework data using large language models, *J. Am. Chem. Soc.*, 2025, 147(47), 43474–43486, DOI: [10.1021/jacs.5c11789](https://doi.org/10.1021/jacs.5c11789).
- 86 J. P. Unsleber, Accelerating Reaction Network Explorations with Automated Reaction Template Extraction and Application, *J. Chem. Inf. Model.*, 2023, 63(11), 3392–3403, DOI: [10.1021/acs.jcim.3c00102](https://doi.org/10.1021/acs.jcim.3c00102).
- 87 J. P. Unsleber, AutoRXN: Cloud-based investigations of catalysis with automated reaction explorations, in *[CATL] Division of Catalysis Science & Technology*, American Chemical Society, Denver, Colorado, USA, 2024.



- 88 I. Kevlishvili, R. G. St Michel, A. G. Garrison, J. W. Toney, H. Adamji, H. Jia, Y. Román-Leshkov and H. J. Kulik, Leveraging natural language processing to curate the tmCAT, tmPHOTO, tmBIO, and tmSCO datasets of functional transition metal complexes, *Faraday Discuss.*, 2025, **256**, 275–303, DOI: [10.1039/D4FD00087K](https://doi.org/10.1039/D4FD00087K).
- 89 D. Zhang, B. Smith, H. Wu, M.-T. Nguyen, R. Rousseau and V.-A. Glezakou, Data Analytics for Catalysis Predictions: Are We Ready Yet?, *ACS Catal.*, 2024, **14**(10), 8073–8086, DOI: [10.1021/acscatal.3c05285](https://doi.org/10.1021/acscatal.3c05285).
- 90 A. Nandy, C. Duan, J. P. Janet, S. Gugler and H. J. Kulik, Strategies and Software for Machine Learning Accelerated Discovery in Transition Metal Chemistry, *Ind. Eng. Chem. Res.*, 2018, **57**(42), 13973–13986, DOI: [10.1021/acs.iecr.8b04015](https://doi.org/10.1021/acs.iecr.8b04015).
- 91 H. J. Kulik, Accelerating the discovery of novel transition metal catalysts through divide and conquer analytical and machine learning strategies, in *[CATL] Division of Catalysis Science & Technology*, American Chemical Society, Denver, Colorado, USA, 2024.
- 92 D. Chu, Ligand additivity models for accelerated precision molecular catalyst screening, in *[CATL] Division of Catalysis Science & Technology*, American Chemical Society, Denver, Colorado, USA, 2024.
- 93 S. Saikin, On Autonomy in Materials Discovery, *Matter Lab Seminar Series*, University of Toronto, 2025.
- 94 J. Schrier, A. J. Norquist, T. Buonassisi and J. Brgoch, Pursuit of the Exceptional: Research Directions for Machine Learning in Chemical and Materials Science, *J. Am. Chem. Soc.*, 2023, **145**(40), 21699–21716, DOI: [10.1021/jacs.3c04783](https://doi.org/10.1021/jacs.3c04783).

