Yang *et al. J. Mater. Inf.* **2025**, *5*, 46
DOI: 10.20517/jmi.2025.38

Journal of Materials Informatics

Perspective

Open Access

Heterogeneous catalyst design by generative models

Chao Yang, Lulu Wang, Jinbo Zhu, Pengfei Ou*

Department of Chemistry, National University of Singapore, Singapore 117543, Singapore.

***Correspondence to:** Asst. Prof. Pengfei Ou, Department of Chemistry, National University of Singapore, 3 Science Drive 3, Singapore 117543, Singapore. E-mail: pengf.ou@nus.edu.sg

How to cite this article: Yang, C.; Wang, L.; Zhu, J.; Ou, P. Heterogeneous catalyst design by generative models. *J. Mater. Inf.* **2025**, *5*, 46. https://dx.doi.org/10.20517/jmi.2025.38

Received: 14 May 2025 **First Decision:** 19 Jun 2025 **Revised:** 26 Jun 2025 **Accepted:** 2 Jul 2025 **Published:** 24 Jul 2025

Academic Editors: Hao Li, Yaqiong Su **Copy Editor:** Pei-Yun Wang **Production Editor:** Pei-Yun Wang

Abstract

Understanding interactions between reactive species and surfaces remains a fundamental challenge in materials science and heterogeneous catalysis. Central to this challenge is the efficient and accurate generation of realistic surfaces and intermediate structures. Despite growing efforts, a universal and systematic approach to surface structure generation is still lacking, particularly for complex interfaces. Existing automated protocols often require extensive computations to identify stable configurations. Recent advances in dataset availability and machine learning techniques, especially in generative models, are beginning to show promise for tasks such as catalyst structure generation. In this perspective, we highlight the emerging capabilities of generative models in catalytic research and outline future directions for their applications. These include property-guided surface structure generation, efficient sampling of adsorption geometries, and the generation of complex transition-state structures. We aim to provide catalysis researchers with a clear view of current progress, outline key challenges, and identify opportunities for integrating generative models into the design and discovery of heterogeneous catalysts.

Keywords: Generative model, machine learning, heterogeneous catalyst, surface science

INTRODUCTION

Catalysis has been central not only to industrial production but also to the advancement of fundamental scientific disciplines. Among various types of catalysis, heterogeneous catalysis accounts for a significant

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

proportion of practical applications[1], from traditional thermocatalytic reactions (such as the Haber-Bosch process) to emerging electrochemical technologies (such as water splitting)[2]. Catalyst remains the core of chemical transformations; it decreases the reaction activation energy, offers alternative reaction pathways, and, as a result, promotes the reaction rate. An essential question in catalysis is identifying the active sites and understanding the reaction mechanism. Significant insights into catalytic reaction mechanisms have been gained through surface science techniques[3] and Kohn-Sham density functional theory (DFT)[4]. The integration of experimental and theoretical approaches has emerged as a powerful approach, facilitating the rational design of catalysts. Complementing this traditional paradigm, artificial intelligence (AI) has recently emerged as a transformative tool, offering new capabilities for modelling, predicting, and designing catalysts[5].

As shown in Figure 1, the number of publications on generative models in materials science and chemistry has steadily increased over recent years, reaching 476 in 2024. A similar trend is observed in surface catalysis. Despite the increasing applications, adoption remains comparatively limited, at approximately 0.4% (data from Web of Science). The scarcity of domain-specific datasets, such as those capturing adsorption configurations and complex interfacial environments on catalytic surfaces, limits the generalizability of generative models beyond well-studied systems. Furthermore, the inherent gap between theoretical simulations and experimental validation remains a critical bottleneck limiting broader adoption in this field. In this context, this perspective focuses on the use of generative models to explore catalytic reactions to further advance this emerging field, with an emphasis on diffusion-based and transformer-based approaches. We begin by outlining the evolution of heterogeneous catalyst design - from traditional paradigms to AI-driven approaches - highlighting the transformative potential of generative models. Following this, we provide an overview of the main generative architectures and their applications to understand their advancements and capabilities [Figure 2]. Through this, we aim to contextualize the emerging role of generative models in heterogeneous catalyst design.

THE EMERGENCE OF GENERATIVE MODELS

In computational catalysis, a conventional research workflow typically involves several key steps: enumerating possible surface structures and corresponding adsorbates[6,7], optimizing the surface models to obtain geometries and energies[8], calculating the reaction free energy profiles that connect a series of intermediates[9], and evaluating reaction barriers to estimate the overall reaction rates[10]. This process may not fully capture chemical intuition or the underlying complexity of realistic surfaces. In addition, the increasing complexity of the reaction network from reactant to product requires extensive DFT calculations. Moreover, dynamic catalytic processes, such as surface reconstruction and phase transformation, are also inevitable[3]. These challenges underscore the limitations of conventional modeling approaches and highlight the necessity for a paradigm shift in how catalysts are designed.

In addressing complex catalytic problems, machine learning interatomic potentials (MLIPs) have advanced rapidly[11], offering a promising approach to bridge the gap between atomistic-level structure and DFT-level energy. Serving as surrogate models, MLIPs accelerate the evaluation of a specific structure, including energy, force, stability, phonon spectrum, and others. However, the diversity of catalyst structures and reaction intermediates is constrained by the initial feeding structure and the extrapolation ability[12]. Optimization-based approaches, such as global minimal hopping[13], genetic algorithm[14], and meta-dynamics[15], combined with MLIP, have enabled the discovery of new active surfaces and reaction pathways.

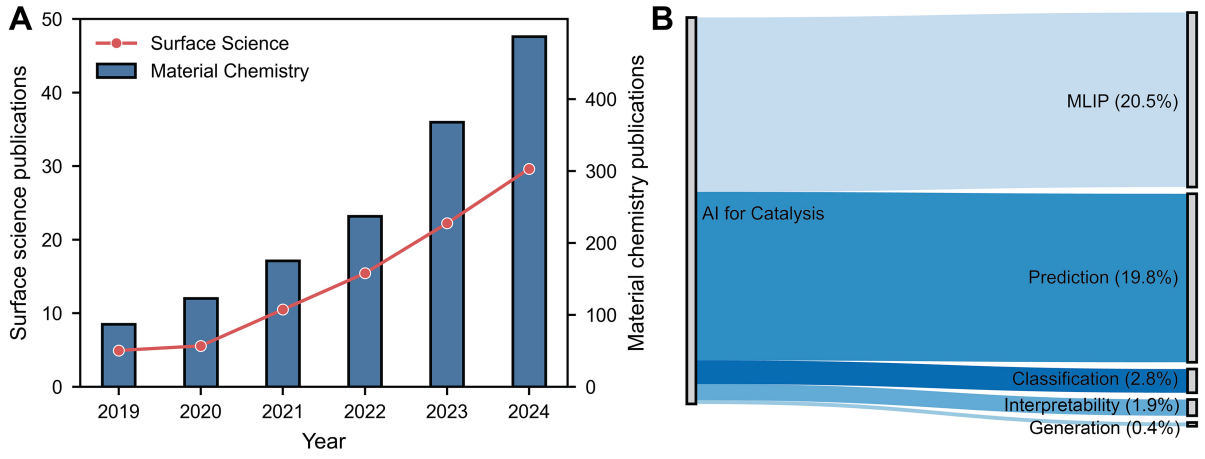


Figure 1. (A) Trends in the number of publications on generative models applied in materials science and chemistry over recent years; (B) Sankey diagram illustrating the categorization and relative proportions of AI applications in catalyst research, based on data extracted from the Web of Science database. AI: Artificial intelligence.

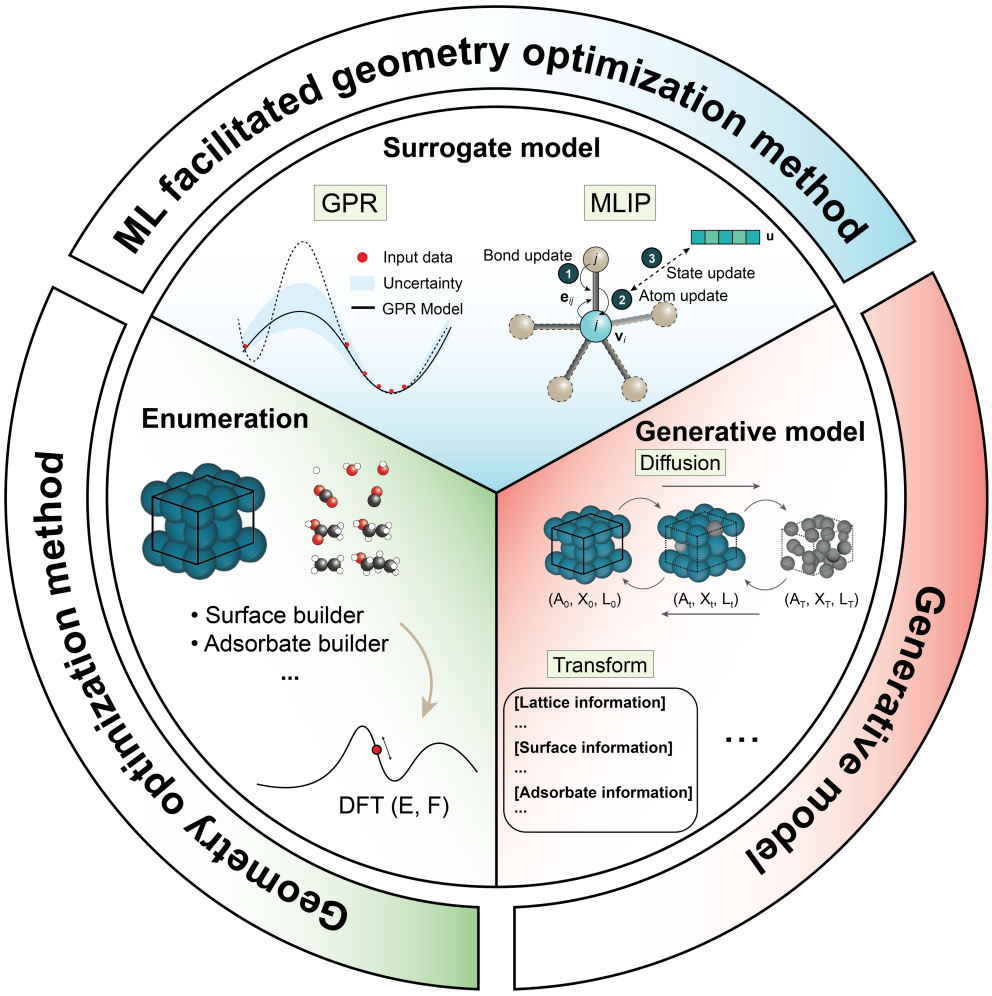


Figure 2. Overview of machine learning techniques related to heterogeneous catalysis. DFT: Density functional theory; GPR: Gaussian process regression; MLIP: machine learning interatomic potential.

In recent years, due to the generation and availability of massive data sets, there has been a meteoric rise in the employment of generative models. These models have emerged as promising candidates for tackling the inverse problem of structure–property prediction and overcoming the drawbacks of forward design in discovering new molecules and materials[16-18]. They are particularly important, as they allow extrapolative exploration in the material or configuration space, thus enabling discoveries beyond existing datasets. Moreover, properties-oriented structure generation could be beneficial for recognizing active sites and reaction mechanisms.

OVERVIEW OF GENERATIVE MODEL ARCHITECTURES

A historical roadmap of the generative models is illustrated in Figure 3, categorizing representative models based on their application domains and underlying architectures. Early applications focused on molecular generation[19-24], followed by the extension to crystal structure prediction (CSP)[25-28], which incorporated periodicity and symmetry constraints. Building upon these developments, generative methods were further applied to surface structures and catalytic intermediates[29-32], treating these tasks as analogs to molecular generation and CSP. Two types of generative models are extensively developed [Figure 4A and B]: variational autoencoders (VAEs)[33], and generative adversarial networks (GANs)[34]. A VAE consists of two neural networks: an encoder and a decoder. The encoder approximates a posterior distribution $q_\theta(z|x)$ over the latent space z given input data x, while the decoder estimates a likelihood distribution $p_\phi(x|z)$ to reconstruct the data from the latent space[35].

Recent advances, particularly in transformer models[36] [Figure 4C] and diffusion models[37] [Figure 4D], have demonstrated unprecedented capabilities in generating coherent, diverse, and realistic content. Diffusion model, inspired by non-equilibrium statistical physics, corrupts the structure in a data distribution through an iterative forward diffusion process and then restores the structure through a reversed process[37].

In contrast to diffusion models, transformer-based architectures rely on multi-head attention mechanisms[36]. They process discrete tokens through an embedding layer and use attention blocks to model contextual dependencies between input elements. Beyond their exceptional multimodal capabilities, they have also been employed in CSP tasks by tokenizing crystal structures[38].

GENERATIVE MODELS IN HETEROGENEOUS CATALYSIS

Benefiting from extensive exploration of the chemical space of crystal structures, many previously unknown yet thermodynamically stable materials have been successfully discovered and synthesized[28]. These newly discovered materials can be further investigated for catalytic reactions using traditional strategies. However, directly generating surface structures and adsorbate configurations offers a more promising pathway for property-guided design (a summary of relevant methods is presented in Table 1).

Surface structure generation

From a global perspective, the goal of generating sufficient surface structures is to identify the most thermodynamically stable or catalytically active crystal facets for a given material composition. This approach is particularly promising for discovering unexpected candidates in underexplored chemical spaces. Song *et al.* demonstrated the ability to generate novel surface structures from latent vectors in CO_2 reduction[26,32]. They combined the crystal diffusion variational autoencoder (CDVAE) model with a bird swarm optimization algorithm, produced over 250,000 candidate structures, 35% of which were predicted to exhibit high catalytic activity. Among these, five alloy compositions, i.e., CuAl, AlPd, Sn_2Pd_5, Sn_9Pd_7, and $CuAlSe_2$, were synthesized and characterized, with two achieving Faradaic efficiencies of ~90% for CO_2 reduction.

Table 1. Summary of representative generative models for catalyst design

Name	Modeling principle	Complexity	Applications	Advantages
VAE	Latent space distribution (hidden representation)	Stable to train	CO_2RR on alloy catalysts[32]	Good interpretability and efficient latent sampling
GAN	Realistic data distribution via adversarial feedback	Difficult to train	Ammonia synthesis with alloy catalysts[29]	High-resolution generation
Diffusion	Reverse-time denoising from noise	Computationally expensive but stable	Surface structure generation[30]	Strong exploration capability and accurate generation
Transformer	Probabilistic token dependencies in sequences		CatGPT for $2e^-$ ORR reaction[31]	Conditional and multi-modal generation

VAE: Variational autoencoder; CO_2RR: carbon dioxide reduction reaction; GAN: generative adversarial network; ORR: oxygen reduction reaction.

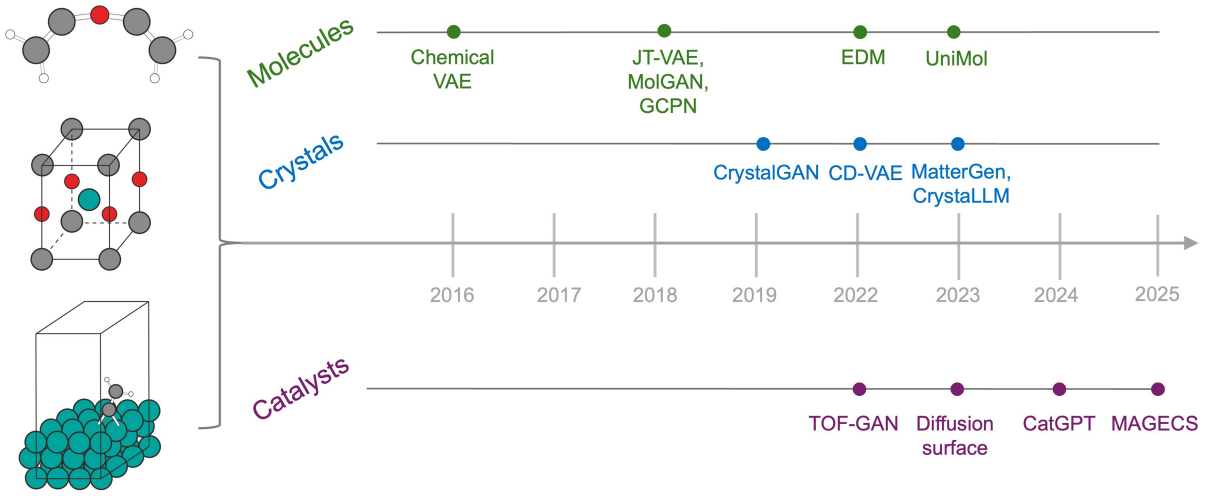


Figure 3. Historical roadmap of generative models for molecules (e.g., ChemicalVAE[19], MolGAN[21], UniMol[24]), crystals (e.g., CD-VAE[26], CrystaLLM[27], MatterGen[28]), and catalysts (e.g., TOF-GAN[29], Diffusion-surface[30], CatGPT[31], MAGECS[32]), covering VAE, GAN, diffusion, and Transformer paradigms. VAE: Variational autoencoder; GAN: generative adversarial network.

From a localized perspective, the focus shifts to identifying atomic-scale active site motifs and strategies to increase their density or effectiveness. Rather than relying on public databases, Rønne *et al.* constructed a custom dataset of surface structures through a global structure search, which was used to train a diffusion model[30]. Tailored to confined surface systems and guided by learned forces, the model generates diverse and stable thin-film structures atop fixed substrates, outperforming random searches in resolving complex domain boundaries.

By integrating strategies both globally and locally - utilizing global search to discover diverse surface motifs and local sampling to fine-tune atomic-scale active sites - one can construct a wide spectrum of surface structures with complex compositions. To ensure an effective balance between exploration and exploitation, latent-space-based generative models can be combined with both global search heuristics and local gradient-based optimizers. The continuous nature of latent representations enables fine-tuned optimization guided by property gradients (e.g., adsorption energy). Furthermore, the incorporation of MLIP for energy and force predictions reduces the cost of multi-scale structure evaluations, making this integrated approach scalable for realistic catalyst systems.

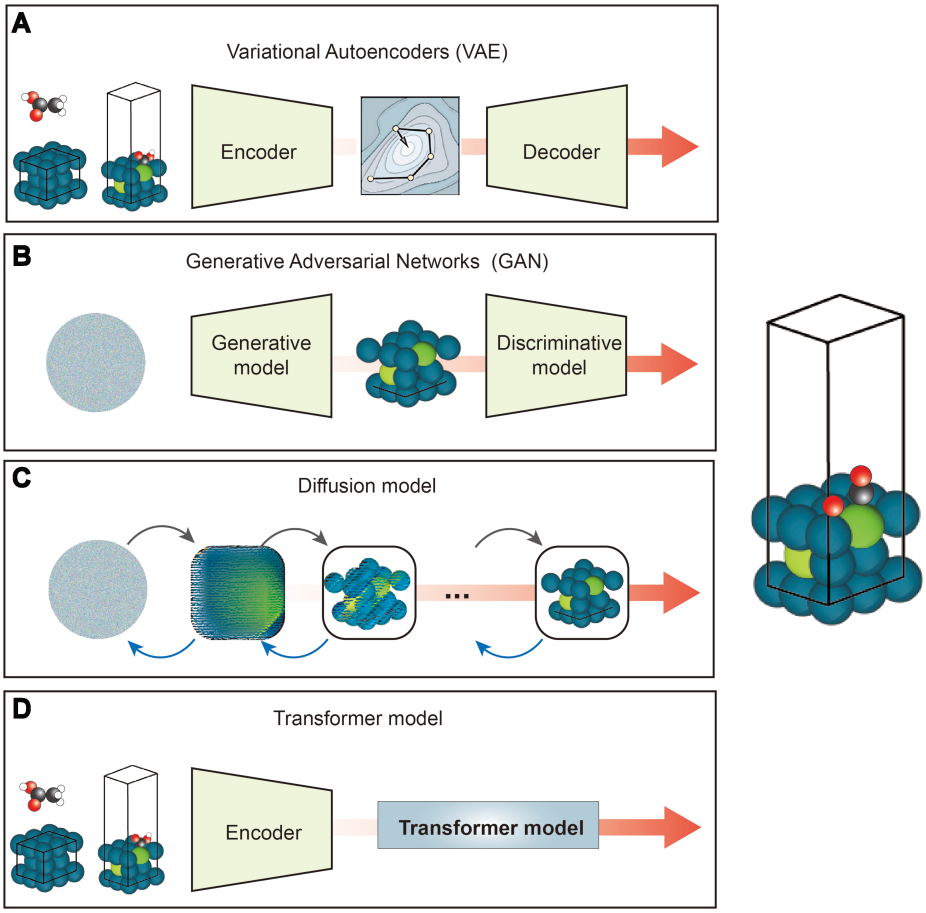


Figure 4. Representative architectures for generative models regarding surface structure generation: (A) VAEs; (B) GANs; (C) diffusion model; and (D) transformer model. VAEs: Variational autoencoders; GANs: generative adversarial networks.

Direct surface structure generation can also be realized using transformer models. For instance, Mok *et al.* trained a GPT-2-based generative model on the OC20 dataset, which generated catalyst structures that closely resemble real surfaces, although challenges such as atomic overlap remain[31]. Building on this early success, the recent expansion of the open-source ecosystem has made a growing number of pretrained transformer models publicly available. These models, with significantly improved reasoning and prediction capabilities, are poised to further accelerate the development of structure generation workflows in catalysis[39].

Adsorbate generation

Another critical challenge is generating surface models with adsorbates that account for multiple structural configurations, each corresponding to a potential thermodynamic local minimum. These alternative structures can differ not only in energy but also in reactivity and mechanistic pathways. From a kinetic perspective, variations in configuration can lead to significantly different activation energies and reaction mechanisms.

Given the high structural fidelity achievable by generative models, they offer a promising alternative to brute-force enumeration. By introducing property biases - such as adsorption energy or local stability - generative approaches may enable the inverse design of intermediate configurations. These models can be

trained directly on a dataset containing the molecule adsorbed surface structure or through molecular docking of adsorbates to surfaces[40]. To ensure the physically realistic structures, several diagnostic criteria can be applied, such as interatomic distance thresholds to flag overlapping atoms, and coordination number analysis to detect under- or over-coordinated species. Furthermore, introducing surface-aware constraints - such as adsorption site templates, dipole orientation checks, or lattice symmetry operations - can substantially improve generation fidelity. Advanced anomaly detection techniques, leveraging descriptors such as adsorption energy, binding motifs, or electronic fingerprints, can further help filter out unrealistic configurations and enhance the catalytic relevance.

Beyond stable intermediate states, the transition state (TS) structures between adsorbed intermediates are also essential in catalytic research. TS search could be reframed as a constrained structure generation problem, where the desired output satisfies specific criteria, such as possessing only one negative eigenvalue in the Hessian matrix. This idea has been demonstrated in molecular systems by 2D reaction graphs[41] or joint distributions of multiple molecules that form an elementary reaction[42]. For these tasks, preserving physical symmetries such as SE(3) equivariant or object-aware SE(3) equivariant is likely necessary.

OUTLOOK

The rapid growth of computing power - driven by breakthroughs in hardware, GPU architectures, and optimization algorithms - has propelled us into the era of AI. In scientific research, AI serves as a transformative tool, capable of solving complex mathematical equations, automating high-throughput experimentation, and accelerating data-driven discovery across various disciplines.

Among AI approaches, generative models, in particular, offer extrapolative capabilities that are valuable across various fields of materials science and heterogeneous catalysis. The evolution from early architectures such as VAE and GAN to more recent diffusion and transformer models has significantly improved the validity and versatility of generated outputs, while also making training processes more manageable and efficient. These models have proven to be applicable far beyond bulk CSP, making them highly relevant to research problems involving complex materials.

In catalysis, generative models are particularly compelling. Structurally, the field has increasingly turned to complex systems such as high-entropy alloys and oxides, which often involve four or more elements with multiple coexisting phases. In terms of reactivity, adsorbed intermediates frequently contain large and diverse molecular species, including elements such as C, N, O, and S. Furthermore, from a mechanistic perspective, dynamic catalytic behavior - including surface reconstruction and transient active sites - requires detailed modeling that goes beyond static configurations. During generation, sufficient sampling of diverse surface–adsorbate configurations can be enhanced by combining data augmentation, heuristic latent space search, and active learning strategies.

Property-guided generative modeling stands out as a particularly promising yet underexplored direction in catalyst design. Both diffusion and transformer models are well-suited for this task. Moreover, combining the multimodal representation capabilities of transformers with the high-fidelity generative power of diffusion models presents a compelling opportunity to advance property-driven inverse design of catalytic materials.

To fully unlock the potential of generative models in catalysis, data-related challenges must also be addressed. To support generative models, high-fidelity datasets that reflect realistic surface structures and reaction conditions are necessary. In parallel, data-efficient strategies such as transfer learning, active learning, and self-supervised learning can reduce labeling costs and accelerate model development.

DECLARATIONS

Authors' contributions

Wrote this manuscript: Yang, C.; Ou, P.
Conceptualized this project: Yang, C.; Wang, L.; Zhu, J.
All authors discussed and contributed to the final version of the manuscript.

Availability of data and materials

The data supporting the findings can be obtained from the corresponding author upon reasonable request.

Financial support and sponsorship

Ou, P. acknowledges support from the National University of Singapore through the Presidential Young Professorship (PYP) Start-Up Grant (A-0010024-00-00) and the White Space Fund (A-0010024-01-00).

Conflicts of interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2025.

REFERENCES

1. Vogt, C.; Weckhuysen, B. M. The concept of active site in heterogeneous catalysis. *Nat. Rev. Chem.* **2022**, *6*, 89-111. DOI PubMed
2. Friend, C. M.; Xu, B. Heterogeneous catalysis: a central science for a sustainable future. *Acc. Chem. Res.* **2017**, *50*, 517-21. DOI PubMed
3. Huang, T.; Cong, X.; Wu, S.; et al. Visualizing the structural evolution of individual active sites in MoS_2 during electrocatalytic hydrogen evolution reaction. *Nat. Catal.* **2024**, *7*, 646-54. DOI
4. Nørskov, J. K.; Studt, F.; Abild-Pedersen, F.; Bligaard, T. Fundamental Concepts in Heterogeneous Catalysis. John Wiley & Sons; 2014. DOI
5. Lai, N. S.; Tew, Y. S.; Zhong, X.; et al. Artificial intelligence (AI) workflow for catalyst design and optimization. *Ind. Eng. Chem. Res.* **2023**, *62*, 17835-48. DOI
6. Sharma, U.; Nguyen, A.; Kitchin, J. R.; Ulissi, Z. W.; Janik, M. J. Enumeration of surface site nuclearity and shape in a database of intermetallic low-index surface facets. *J. Catal.* **2024**, *440*, 115795. DOI
7. Tran, K.; Ulissi, Z. W. Active learning across intermetallics to guide discovery of electrocatalysts for CO_2 reduction and H_2 evolution. *Nat. Catal.* **2018**, *1*, 696-703. DOI
8. Reveles, J. U.; Köster, A. M. Geometry optimization in density functional methods. *J. Comput. Chem.* **2004**, *25*, 1109-16. DOI PubMed
9. Ye, J.; Liu, C.; Mei, D.; Ge, Q. Active oxygen vacancy site for methanol synthesis from CO_2 hydrogenation on In_2O_3(110): a DFT study. *ACS. Catal.* **2013**, *3*, 1296-306. DOI
10. Liu, Z. P.; Hu, P. General rules for predicting where a catalytic reaction should occur on metal surfaces: a density functional theory study of C-H and C-O bond breaking/making on flat, stepped, and kinked metal surfaces. *J. Am. Chem. Soc.* **2003**, *125*, 1958-67. DOI PubMed
11. Ko, T. W.; Ong, S. P. Recent advances and outstanding challenges for machine learning interatomic potentials. *Nat. Comput. Sci.* **2023**, *3*, 998-1000. DOI PubMed
12. Pitfield, J.; Brix, F.; Tang, Z.; et al. Augmentation of universal potentials for broad applications. *Phys. Rev. Lett.* **2025**, *134*, 056201. DOI PubMed
13. Jung, H.; Sauerland, L.; Stocker, S.; Reuter, K.; Margraf, J. T. Machine-learning driven global optimization of surface adsorbate

geometries. *npj. Comput. Mater.* **2023**, *9*, 1065. DOI
14. Xu, J.; Xie, W.; Han, Y.; Hu, P. Atomistic insights into the oxidation of flat and stepped platinum surfaces using large-scale machine learning potential-based grand-canonical Monte Carlo. *ACS. Catal.* **2022**, *12*, 14812-24. DOI
15. Yang, M.; Bonati, L.; Polino, D.; Parrinello, M. Using metadynamics to build neural network potentials for reactive events: the case of urea decomposition in water. *Catal. Today.* **2022**, *387*, 143-9. DOI
16. Sanchez-Lengeling, B.; Aspuru-Guzik, A. Inverse molecular design using machine learning: generative models for matter engineering. *Science* **2018**, *361*, 360-5. DOI PubMed
17. Park, H.; Li, Z.; Walsh, A. Has generative artificial intelligence solved inverse materials design? *Matter* **2024**, *7*, 2355-67. DOI
18. Chenebuah, E. T.; Nganbe, M.; Tchagang, A. B. A deep generative modeling architecture for designing lattice-constrained perovskite materials. *npj. Comput. Mater.* **2024**, *10*, 1381. DOI
19. Gómez-Bombarelli, R.; Wei, J. N.; Duvenaud, D.; et al. Automatic chemical design using a data-driven continuous representation of molecules. *ACS. Cent. Sci.* **2018**, *4*, 268-76. DOI PubMed PMC
20. Jin, W.; Barzilay, R.; Jaakkola, T. Junction tree variational autoencoder for molecular graph generation. *arXiv* **2018**, arXiv:1802.04364. https://doi.org/10.48550/arXiv.1802.04364. (accessed 4 Jul 2025)
21. De Cao, N.; Kipf, T. MolGAN: an implicit generative model for small molecular graphs. *arXiv* **2018**, arXiv:1805.11973. https://doi.org/10.48550/arXiv.1805.11973. (accessed 4 Jul 2025)
22. You, J.; Liu, B.; Ying, R.; Pande, V.; Leskovec, J. Graph convolutional policy network for goal-directed molecular graph generation. *arXiv* **2018**, arXiv:1806.02473. https://doi.org/10.48550/arXiv.1806.02473. (accessed 4 Jul 2025)
23. Hoogeboom, E.; Satorras, V. G.; Vignac, C.; Welling, M. Equivariant diffusion for molecule generation in 3D. *arXiv* **2022**, arXiv:2203.17003. https://doi.org/10.48550/arXiv.2203.17003. (accessed 4 Jul 2025)
24. Zhou, G.; Gao, Z.; Ding, Q.; et al. Uni-Mol: a universal 3D molecular representation learning framework. *ChemRxiv* **2023**,. DOI
25. Nouira, A.; Sokolovska, N.; Crivello, J. C. CrystalGAN: learning to discover crystallographic structures with generative adversarial networks. *arXiv* **2018**, arXiv:1810.11203. https://doi.org/10.48550/arXiv.1810.11203. (accessed 4 Jul 2025)
26. Xie, T.; Fu, X.; Ganea, O. E.; Barzilay, R.; Jaakkola, T. Crystal diffusion variational autoencoder for periodic material generation. *arXiv* **2021**, arXiv:2110.06197. https://doi.org/10.48550/arXiv.2110.06197. (accessed 4 Jul 2025)
27. Antunes, L. M.; Butler, K. T.; Grau-Crespo, R. Crystal structure generation with autoregressive large language modeling. *Nat. Commun.* **2024**, *15*, 10570. DOI PubMed PMC
28. Zeni, C.; Pinsler, R.; Zügner, D.; et al. A generative model for inorganic materials design. *Nature* **2025**, *639*, 624-32. DOI PubMed PMC
29. Ishikawa, A. Heterogeneous catalyst design by generative adversarial network and first-principles based microkinetics. *Sci. Rep.* **2022**, *12*, 11657. DOI PubMed PMC
30. Rønne, N.; Aspuru-Guzik, A.; Hammer, B. Generative diffusion model for surface structure discovery. *Phys. Rev. B.* **2024**, *110*, 235427. DOI
31. Mok, D. H.; Back, S. Generative pretrained transformer for heterogeneous catalysts. *J. Am. Chem. Soc.* **2024**, *146*, 33712-22. DOI PubMed
32. Song, Z.; Fan, L.; Lu, S.; Ling, C.; Zhou, Q.; Wang, J. Inverse design of promising electrocatalysts for CO_2 reduction via generative models and bird swarm algorithm. *Nat. Commun.* **2025**, *16*, 1053. DOI PubMed PMC
33. Kingma, D. P.; Welling, M. Auto-encoding variational bayes. *arXiv* **2013**, arXiv:1312.6114. https://doi.org/10.48550/arXiv.1312.6114. (accessed 4 Jul 2025)
34. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; et al. Generative adversarial networks. *Commun. ACM.* **2020**, *63*, 139-44. DOI
35. Menon, D.; Ranganathan, R. A generative approach to materials discovery, design, and optimization. *ACS. Omega.* **2022**, *7*, 25958-73. DOI PubMed PMC
36. Vaswani, A.; Shazeer, N.; Parmar, N.; et al. Attention is all you need. *arXiv* **2017**, arXiv:1706.03762. https://doi.org/10.48550/arXiv.1706.03762. (accessed 4 Jul 2025)
37. Ho, J.; Jain, A.; Abbeel, P. Denoising diffusion probabilistic models. *arXiv* **2020**, arXiv:2006.11239. https://doi.org/10.48550/arXiv.2006.11239. (accessed 4 Jul 2025)
38. Yan, K.; Li, X.; Ling, H.; et al. Invariant tokenization of crystalline materials for language model enabled generation. *arXiv* **2025**, arXiv:2503.00152. https://doi.org/10.48550/arXiv.2503.00152. (accessed 4 Jul 2025)
39. DeepSeek-AI; Guo, D.; Yang, D.; et al. DeepSeek-R1: incentivizing reasoning capability in LLMs via reinforcement learning. *arXiv* **2025**, arXiv:2501.12948. https://doi.org/10.48550/arXiv.2501.12948. (accessed 4 Jul 2025)
40. Martinelli, D. D. Generative machine learning for de novo drug discovery: a systematic review. *Comput. Biol. Med.* **2022**, *145*, 105403. DOI
41. Kim, S.; Woo, J.; Kim, W. Y. Diffusion-based generative AI for exploring transition states from 2D molecular graphs. *Nat. Commun.* **2024**, *15*, 341. DOI PubMed PMC
42. Duan, C.; Du, Y.; Jia, H.; Kulik, H. J. Accurate transition state generation with an object-aware equivariant elementary reaction diffusion model. *Nat. Comput. Sci.* **2023**, *3*, 1045-55. DOI PubMed