Review | Received 18 June 2025; Accepted 18 July 2025; Published 30 July 2025
https://doi.org/10.55092/aimat20250011

Reinforcement learning world models for catalyst surface reconstruction: state-of-the-art review

Aisha Samreen[1,2], Muhammad Azim[3] and Fuyi Chen[1,2,*]

[1] State Key Laboratory of Solidification Processing, Northwestern Polytechnical University, Xi'an 710072, China
[2] School of Materials Science and Engineering, Northwestern Polytechnical University, Xi'an 710072, China
[3] School of Management, Northwestern Polytechnical University, Xi'an 710072, China

* Correspondence author; E-mail: fuyichen@nwpu.edu.cn.

Highlights:

- DreamerV3 has been proposed to optimize dynamic catalyst surfaces at the atomic level while maintaining safety.
- The first conceptual RL world model framework for reconstructing AgPd nanoalloys in operando settings.
- Latent world models combine physics-informed simulations with operando data.
- World models significantly lower DFT computational burdens in catalytic reconstruction tasks.
- Using multi-objective reward balancing techniques in Dreamer-based latent world models, multi-objective catalyst surface control is conceptualized.

Abstract: Catalyst surfaces are dynamic objects that constantly rebuild in response to stimuli like temperature, electrochemical potentials, and adsorbates under reactive conditions. Conventional catalyst design paradigms, relying on static pre-catalyst structures, fail to account for this intrinsic dynamism, leading to imprecise predictions of catalytic activity and stability. This review critically analyzes the shortcomings of empirical and simulation-based design approaches while synthesizing basic surface restructuring phenomena (such as oxidation-state dynamics in PtCu single-atom alloys and adsorbate-induced phase transitions in AgPd nanoalloys). We investigate the possibilities of World Models guided Reinforcement Learning frameworks based on neural networks as a viable method for controlling and predicting surface reconstruction. These methods allow adaptive policy optimization in dynamic catalytic systems by combining experimental data with physics-informed atomistic simulations. The review describes important challenges in uncertainty quantification, reward balancing, and latent state interpretability for future catalyst-specific world models, while highlighting how AI frameworks trained on operando data, when paired with simulations informed by physics, opens new avenues for predictive catalyst design.

Keywords: reinforcement learning world models; catalyst surface reconstruction; model-based reinforcement learning; dreamer algorithm; adaptive catalyst surface control; computational catalysis; density functional theory; surrogate models; inverse design in catalysis

1. Introduction

The catalyst surface continuously rebuilds itself during catalytic reactions, making it a dynamic entity. Environmental stimuli like temperature, adsorbates, electrochemical potentials, and electrolyte composition cause changes in atomic structures, oxidation states, and defect concentrations [1,2]. Surface chemistry, active site topology, and ultimately catalytic function are all fundamentally redefined by these processes, which are by no means insignificant. Design paradigms for traditional catalysts, heavily focused on static pre-catalyst structure optimization and theoretical predictions under ideal scenario assumptions, fail to incorporate the dynamism found in reaction surroundings [3].

Extraordinary cases are illustrated by AgPd Nano alloys [4], where Pd surface segregation under oxidizing environments at variance with the surface before reduction, gives rise to active PdO_x sites for the formate oxidation activity, while suffering from sintering to inactive ensembles [5,6]. In PdO-Co_3O_4 nanocomposites, Hu *et al*. [6] also noted that surface oxide formation on RhO_2[110]/Rh has a significant impact on catalytic behavior, as reported by Over *et al*. [7]. However, such effects are not always beneficial. Buzkova *et al.* [8] found that surface PdO_x species may inhibit NO oxidation rate in some cases, which suggests a delicate interplay between beneficial and detrimental effects of surface restructuring.

Recent work highlights that, apart from chemical identity, crystal orientation and strain state are also important in influencing oxidation and catalysis. Adsorption behavior and reaction path depend on the reactivity of surface, which can adaptively reconstruct in reactive environment. This was shown by Duchesne *et al*. [9] and Zhang *et al*. [10] who demonstrated that PtAu and PdAg single atom alloys (SAAs) have enhanced formic acid oxidation (FAO) activity, indicating the importance of ultradilute noble metal atoms in the highly-ordered surface alloys for the facile reactant dissociation and the favorable intermediate adsorption.

An emerging paradigm that sees catalysts as dynamic, changing structures as opposed to static frameworks is being supported by mounting evidence. According to Feng *et al*. [1], spontaneous and occasionally irreversible rearrangements of the catalyst's initial structure are driven by interactions with the reaction environment (reactants, intermediates, and electrolyte). For instance, Pt clusters in CO oxidation are converted to single atoms under oxidative conditions, whereas Cu single atoms in Cu-N-C catalysts dynamically form clusters in CO_2 electro-reduction. The catalytic behavior is greatly influenced by the atomic-scale atom-to-nanocluster flows. Operando studies have shown that synergistic CO and H adsorption can weaken coordination bonds such as Cu-N, facilitating atomic migration and dynamically altering active site populations in real time. Conversely, oxygen vacancies and strong metal-support interactions, as exemplified by Pt-O-Ce bonds on CeO_2, can stabilize dispersed Pt single atoms in oxidative environments [2]. These dynamic equilibria continue to redefine long-held concepts of active sites and catalytic mechanisms.

Surface reconstruction is composed of a complex interface of extrinsic parameters, such as temperature, applied potential, electrolyte composition, and reaction duration, alongside intrinsic factors like composition, strain, and atomic arrangement [1,2]. In Cu-N-C catalysts, for instance, increasing cathodic potentials during CO_2 electroreduction destabilizes Cu-N bonds, speeding up the formation of

Cu clusters and ultimately increasing the production of ethanol [2]. These processes are also affected by physical forces like gas evolution and electrolyte convection, as well as chemical factors like adsorbate interactions and redox cycles (like Pd oxidation/reduction) [1,2].

Several important factors that control the dynamic structural changes under reactive conditions have an impact on surface reconstruction on catalytic surfaces. Phase transitions on the surface can be triggered by adsorbate-induced restructuring driven by the interaction between the adsorbates and the metal substrate, which is one of the main factors. For example, in the NO–H_2 reaction on Pt(100), NO adsorption causes the surface to change from a quasi-hexagonal arrangement of Pt atoms (stable phase) to a (1 × 1) metastable phase. Phase segregation and the creation of mesoscopic restructured islands with distinct borders are encouraged by this restructuring, which is fueled by an increase in adsorption energy that is connected with the number of nearby substrate atoms in the metastable state.

The surface dynamics are also significantly impacted by the rates of surface restructuring in relation to adsorption, reaction, and diffusion steps; surface restructuring typically happens more slowly than adsorption-reaction steps but significantly more slowly than adsorbate diffusion. Furthermore, even minor changes in the adsorbate arrangement can have a significant impact on the reaction kinetics and surface phase behavior, resulting in intricate phenomena like chaotic behavior and kinetic oscillations. By affecting the rates and stability of various surface phases, temperature and adsorbate partial pressure further modulate these effects. Therefore, the surface reconstruction processes in catalytic systems are controlled by the interaction of adsorbate interactions, surface atom states, and reaction conditions.

Modern operando characterization methods are now essential for decoding these fleeting and extremely reactive states. In order to correlate bond-length changes, oxidation state evolution, and atomic migration with catalytic performance under operando conditions, techniques like ab initio molecular dynamics (AIMD) simulations, identical-location scanning transmission electron microscopy (STEM), and synchrotron-based X-ray absorption spectroscopy (XAS) have proven crucial [1,2]. These tools close the long-standing gap between the dynamic realities of catalytic systems and static theoretical models.

Despite these developments, traditional catalyst design methodologies still primarily rely on empirical and iterative methods and have significant drawbacks. It can take weeks to synthesize and test a single catalyst variant, making traditional trial-and-error experimentation time-consuming and expensive. Even though they are helpful, physics-based simulations such as Density Functional Theory (DFT) calculations are limited by strict presumptions like idealized surfaces and constant temperature. The dynamic nature of reaction environments is also not accommodated by static defect engineering, and materials science's exploration of large configurational spaces is still ineffective and frequently ignores ideal structures. Moreover, current models have limited applicability because they are unable to generalize beyond particular chemistries and material systems [3]. In this regard, machine learning's (ML) primary strength is its capacity to generalize from training data to examples that have not yet been seen, which makes it an effective tool for speeding up processes like materials discovery, reaction outcome modeling, and molecular property prediction [11].

At the same time, reinforcement learning (RL) provides the ability to make adaptive decisions in complex reaction environments, has also become a promising tool for modeling catalyst surface reconstructions. However, there are several obstacles that prevent it from being used practically. In order to learn effective policies for predicting surface reconstructions in multi-component systems such as PtNiCu alloys, RL agents frequently need a large amount of simulation data, which imposes prohibitive

computational costs [12]. Recent studies have brought attention to sample inefficiency issues. Rapid screening and optimization across various catalyst spaces is further hampered by the fact that RL models trained on a specific alloy composition frequently lack transferability, requiring retraining for every new system [3]. Although RL has exciting potential for catalyst discovery, before these models can be widely used in real-world catalyst design, issues with efficiency, generalizability, uncertainty quantification, and physical interpretability need to be resolved.

Understanding the dynamic reconstruction behavior of AgPd catalyst surfaces under catalytic conditions is one particular research gap that still exists in adaptive control of these surfaces. In order to predict surface-active site structures, traditional models usually rely on bulk alloy properties, which may not accurately reflect actual catalytic performance [5]. Accurate predictions of several energy parameters necessary for comprehending surface reconstruction have been made possible by the creation of a generalizable deep learning potential for the Ag-Pd-F system, which has also allowed for deeper insights into interatomic interactions [5]. Further research on atomic diffusion and dislocation migration has demonstrated that stress relaxation during fluorination is a major factor in the structural development of AgPd nanoalloys.

The primary goals of this review are to address the following:

(1) To forecast and regulate surface reconstruction of AgPd catalysts under catalytic conditions, present a novel RL framework that is guided by a neural network-based world model simulator and trained on physics-informed atomistic simulations and experimental data.

(2) Highlight the benefits of incorporating RL agents into world models in catalyst optimization processes.

(3) Describe potential ways to combine AI-driven approaches with experimental validation techniques, while highlighting present and future constraints and opportunities for using world model-based RL frameworks in catalyst design, such as issues with model transferability, physical interpretability, and computational cost.

2. Reinforcement learning and deep reinforcement learning in catalyst surface reconstruction

2.1. Reinforcement learning: a framework for decision-making

Reinforcement learning has emerged as a powerful computational paradigm capable of enabling machines to solve complex, sequential decision-making tasks through interaction with their environment. This RL framework has demonstrated remarkable success in surpassing human performance in strategic games such as Go and Dota [13,14], and has been instrumental in advancing large language models beyond their pretraining data limitations [15,16]. Unlike traditional supervised learning approaches, which rely on labeled datasets, RL addresses problems by allowing an agent to interact with an environment, taking actions that lead to transitions between different states within a predefined state-space. In each step, the environment provides feedback in the form of scalar rewards or penalties, guiding the agent to learn policies that maximize cumulative rewards over time [17].

In the domain of computational chemistry, RL has demonstrated increasing promise across the varieties of domains [18]. Key applications include, molecule generation, geometry optimization, the search for retrosynthetic pathways and nonlinear chemical synthesis control. The role of RL in computational chemistry is graphically summarized in Figure 1, which offers a conceptual overview of these applications.

Molecule Generation: Designing novel molecules or nanoclusters [19] with optimal physicochemical or biological properties.

Geometry Optimization: Optimizing molecular geometries and nanoclusters [19] to achieve minimal energy configurations and enhanced stability.

Retrosynthetic Pathway Search: Discovering efficient, cost-effective synthetic pathways for complex chemical compounds.

Non-Linear Control of Chemical Synthesis: Dynamically adjusting reaction parameters such as temperature, pressure, and catalyst concentration to optimize chemical processes [20].

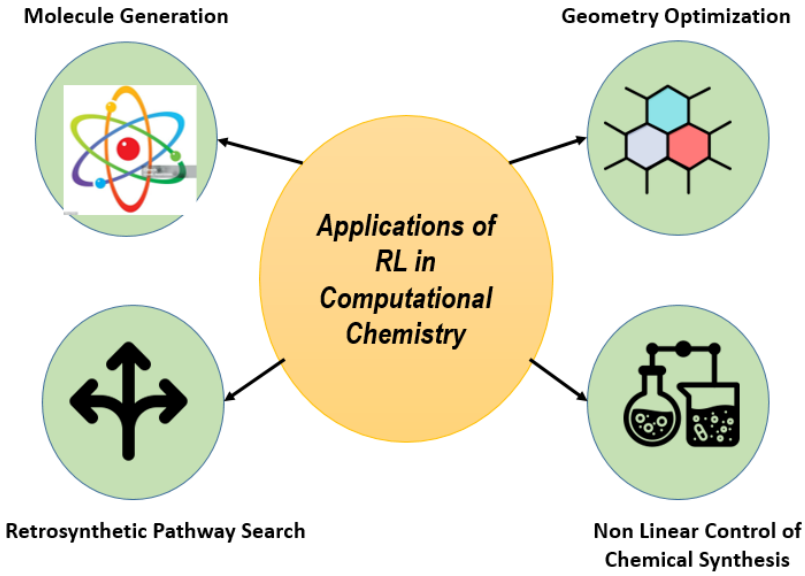

Figure 1. Applications of RL in computational chemistry.

Complex, high-dimensional environments or computationally costly simulations, like density functional theory calculations, present difficulties for reinforcement learning as it continues to find a variety of uses in computational chemistry. In order to overcome this, agents can use past experiences to learn streamlined, predictive models of their surroundings. Model-based reinforcement learning (MBRL) is a technique that enables agents to refine their policies or value functions by interacting with an internal model instead of the real environment. This method involves an agent learning a mechanism to describe the environment through experience, then interacting with the model to solve the Markov Decision Process (MDP). In order to improve its decision-making process iteratively, the agent then applies the learned policies to the actual environment, gaining fresh experience.

2.2. Deep reinforcement learning

Deep reinforcement Learning Model is an extension of traditional reinforcement learning that integrates deep neural networks to allow agents to handle high-dimensional state spaces and complex, sequential action sets. Depending on whether policy-based (e.g., policy gradient methods) or value-based (e.g., Q-learning) algorithms are used, neural networks in this framework act as function approximators for the agent's policy or value functions. This development has increased the applicability of RL to complex domains, such as catalyst design and material optimization, where the state and action spaces are large and computationally demanding.

The iterative modification of atomic configurations on catalyst surfaces through the use of Deep RL agents is one noteworthy application. In order to guide agents toward the discovery of new and potentially better surface structures, they interact with an environment that is modeled using physical or quantum mechanical simulations and receive rewards based on structural stability, energy minimization, or catalytic performance [21].

2.3. Applications of deep reinforcement learning in materials science

Deep RL is rapidly gaining traction as a transformative tool in materials science and computational chemistry, enabling important tasks like inverse design of crystalline materials, material microstructure optimization, nanocluster configuration and hypothesis generation. Figure 2 schematically depicts these various applications, demonstrating the range of ways DRL is advancing material design and modeling.

(a) Inverse Design of Crystalline Materials:

Deep RL frameworks have been applied to the inverse design of crystalline materials, allowing for the generation of novel compounds with targeted properties [22]. These approaches often rely on creating latent representations of crystal structures using generative models such as generative adversarial networks (GANs), variational autoencoders (VAEs), or diffusion-based models. By incorporating property constraints directly into the generative process, Deep RL agents can efficiently navigate the material design space to propose candidates optimized for characteristics such as thermoelectric efficiency, stability, or bandgap performance.

(b) Material Microstructure Optimization:

Optimizing material microstructures to attain specific extrinsic properties is another exciting application. By learning the intricate relationship between microstructural parameters and macroscopic material performance, deep reinforcement learning algorithms can fine-tune structural features to improve properties like thermal stability, electrical conductivity, and mechanical strength [23]. This approach opens new avenues for the targeted design of high-performance materials through microstructural engineering.

(c) Optimizing Nanocluster Configuration:

Autonomous prediction of low-energy configurations in multimetallic nanoclusters is a recent and significant use of Deep RL. In order to effectively traverse the high-dimensional potential energy surfaces (PES) of ternary $Ag_6Pd_5Cu_4$ nanoclusters, Mubeen [24] presented a DRL-based framework that integrates trust region policy optimization (TRPO). Their framework quickly predicted global minimum and several low-energy configurations by using atom-centered symmetry functions (ACSFs) to encode atomic environments and designed reward functions that rewarded energy minimization and penalized unstable geometries. Comparable to traditional optimization methods like Genetic Algorithms and Simulated Annealing, Usman and Chen [25] used a policy-based actor-critic DRL framework employing TRPO for autonomous structural optimization of Au_{13} nanoclusters, enabling efficient convergence to global minimum and multiple low-energy configurations with improved convergence speed and structural stability. These studies demonstrate DRL's increasing capacity to independently resolve combinatorial complexity in multimetallic systems, providing scalable routes for predicting the structure of nanoclusters with possible extensions to supported nanocatalyst applications.

(d) Hypothesis Generation in Computational Chemistry:

Deep RL has also proven valuable in autonomous hypothesis generation in fields such as drug discovery and catalyst design. By encoding domain-specific constraints into reward functions such as penalizing toxicophoric substructures in molecules or rewarding low-energy, high-activity catalyst surfaces—agents can efficiently explore vast combinatorial design spaces. This paradigm shifts the researcher's role from manual candidate evaluation to the design and validation of reward functions, dramatically accelerating discovery workflows [26].

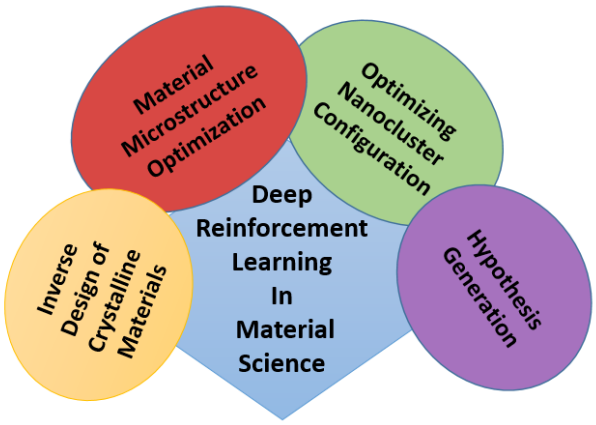

Figure 2. Applications of deep reinforcement learning.

2.4. Deep reinforcement learning for catalyst surface reconstruction: the ASLA approach

A significant recent advancement in applying Deep RL to materials science is represented by the Atomistic Structure Learning Algorithm (ASLA), introduced by Meldgaard *et al*. [27]. ASLA integrates image recognition with a deep RL framework to predict optimal surface reconstructions for crystalline materials. Utilizing a deep neural network as an RL, ASLA autonomously constructs chemically stable atomic structures by interacting with a first-principles quantum mechanical energy calculator. The model operates using only atomic position and type information, iteratively refining surface configurations based on energy evaluations. ASLA has demonstrated notable success in reconstructing surface structures for binary oxides such as anatase TiO_2(001)-(1 × 4) and rutile SnO_2(110)-(4 × 1) which illustrates the agent-environment interaction and reward-guided optimization strategy employed for atomic structure prediction [27].

2.5. Opportunities and challenges of ASLA for AgPd catalyst systems

Although ASLA shows promise, it presents a number of difficulties when applied to more complex systems, such as AgPd alloy catalysts:

Alloy Complexity: In contrast to binary oxide systems, AgPd alloys involve additional complexities related to surface segregation phenomena, alloy composition, and the dynamic interplay between silver and palladium atoms during surface reconstruction. Reward functions, atomic representations, and sampling strategies would require necessary modifications to extend ASLA to handle these factors.

Computational Cost: Although ASLA requires fewer density functional theory (DFT) evaluations than brute-force methods, it still requires 1,000–10,000 DFT calculations per optimization task, making it computationally costly for large, multi-element alloy systems with a large number of possible surface configurations [27].

Static Nature of Predictions: The primary goal of ASLA's current implementations is to forecast ground-state, static surface configurations. However, dynamic reconstructions, such as surface atom diffusion, phase transitions, and adsorbate-induced restructuring, occur in realistic catalytic systems under reaction conditions. To achieve predictive accuracy in real-world catalytic environments, integrating dynamic simulation capabilities into RL frameworks is still a crucial and unresolved challenge.

Reinforcement learning frameworks increasingly use model-based approaches to tackle these modeling and computational issues in complex catalyst systems. In these methods, agents learn predictive models that approximate the dynamics of their environment by drawing on their cumulative experience. By interacting with an internal model instead of the real world, agents can iteratively improve their policies or value functions using a technique called model-based reinforcement learning. The agent solves the Markov Decision Process (MDP) by interacting with a model that it has learned to describe the environment through experience. The improved policies are then implemented in the actual setting, where fresh insights are gained to improve decision-making even more. This method serves as the theoretical basis for the World Model frameworks covered in the next section and greatly lessens the computational load related to high-fidelity simulations like DFT calculations. These latent world model frameworks, especially the DreamerV3 framework suggested in Section 4, offer a scalable and uncertainty-aware modeling approach ideal for such complexities, given the highly dynamic behavior of systems like AgPd surfaces, which are characterized by restructuring, segregation, and adsorbate-induced transitions. As will be explained later, this serves as the rationale for using latent dynamics models in catalyst surface reconstruction tasks. The model-based reinforcement learning is schematically shown in Figure 3, where agents use simulated transitions from a learned environment model as well as real-world experiences to improve policies.

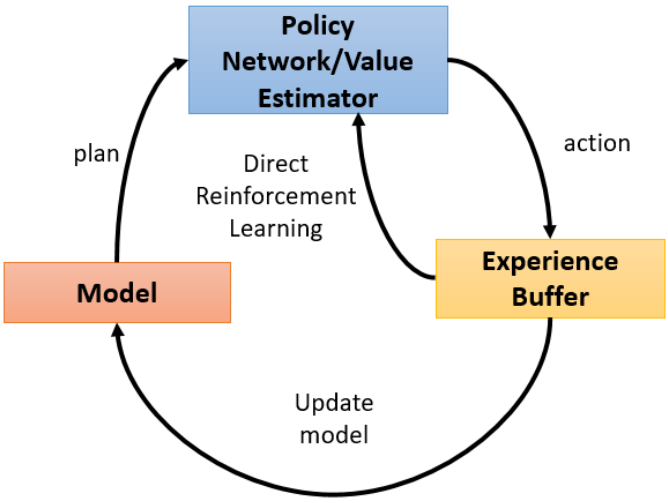

Figure 3. Schematic diagram of model based reinforcement learning.

2.6. The Dyna architecture as the basis of model-based RL

Reinforcement learning frameworks are increasingly using model-based approaches to tackle modeling and computational issues in complex catalyst systems. These techniques eliminate the need for expensive real-world calculations like DFT evaluations by having agents learn predictive models of their surroundings from experience and use them to model interactions.

In this regard, Sutton's Dyna architecture [28] serves as a fundamental framework. Dyna uses simulated experiences to combine planning, model learning, and direct reinforcement learning. Before interacting with the real world, agents in this method refine their policy or value functions by generating hypothetical experiences using an environment model they have learned from real interactions.

Direct Reinforcement Learning: Q-learning is used to update action-value estimates based on actual experiences.

$$Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma \max_{a'} Q(s',a') - Q(s,a)] \quad (1)$$

Model Learning: By using supervised learning from observed transitions, a transition model T is gradually built from $T(s \mid s,a)$ and a reward model $R(r \mid s,a)$.

Planning through Simulated Experience: It uses simulated transitions to carry out numerous updates:

Regarding k = 1 to K:

a. Take a memory sample of (s,a)

b. Make a prediction: $s' \sim T(s'|s,a), r \sim R(s,a)$

c. Use simulated (s,a,r,s') to update $Q(s,a)$

Contemporary model-based RL frameworks like ASLA and Dreamer, which provide effective policy learning in high-dimensional, computationally costly environments, are based on this hybrid structure of real and imagined experience.

3. World model in reinforcement learning

3.1. An overview of world models

In the past, reinforcement learning has optimized policies through direct trial-and-error interactions. However, in domains such as catalyst surface reconstruction, such real-world experimentation is expensive, time-consuming, and dangerous. In order to overcome this, world models, which act as internal simulators of environment dynamics, were introduced by model-based reinforcement learning MBRL frameworks. This allowed agents to predict state transitions and rewards without having to physically interact with the environment [29,30]. In order to increase efficiency and decrease data dependency, these models usually entail training a sizable unsupervised neural network to capture system dynamics in conjunction with a smaller RL-based controller for decision-making [31].

However, there are still issues with reinforcement learning's ability to adapt to new fields. Tasks involving board games [32], spatial reasoning [33], continuous control [34], discrete actions [35,36], sparse rewards [37], and vision-based systems [38] necessitate domain-specific adjustments for algorithms such as PPO [39], which frequently require significant hyper parameter tuning [40]. By allowing agents to internally simulate multi-step future states, world models, on the other hand, greatly improve policy optimization, lower computational costs, and increase sample efficiency [41]. In high-dimensional spaces,

their capacity to calculate analytic gradients instead of depending on derivative-free techniques further increases training efficiency [42–45].

Model-based RL frameworks use these world models to plan and learn "in imagination" [46,47], significantly reducing sample complexity, in contrast to conventional model-free RL techniques that necessitate copious environmental data [48–50]. The ability of world models to accurately simulate complex and continuous systems has improved due to developments in probabilistic modeling and deep learning, which have further improved latent state representations and decision-making capabilities [51–53]. The architecture of a world model-based reinforcement learning system is depicted in Figure 4.

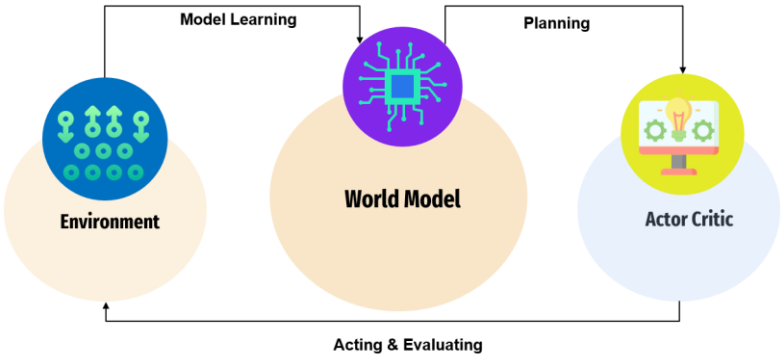

Figure 4. World model based RL architecture.

Three fundamental elements make up a typical world model architecture, which has been defined as;

Vision Model: It comprises of high-dimensional observations, like pictures of surface structures, are compressed into a lower-dimensional latent space by the Representation Learning Module (Encoder). Every image frame is encoded into a latent vector z using a Variational Autoencoder (VAE), which retains important details while eliminating extraneous ones as represented in Figure 5. The agent can more easily learn significant patterns and forecast future states using these latent representations thanks to the efficient processing made possible by this compression.

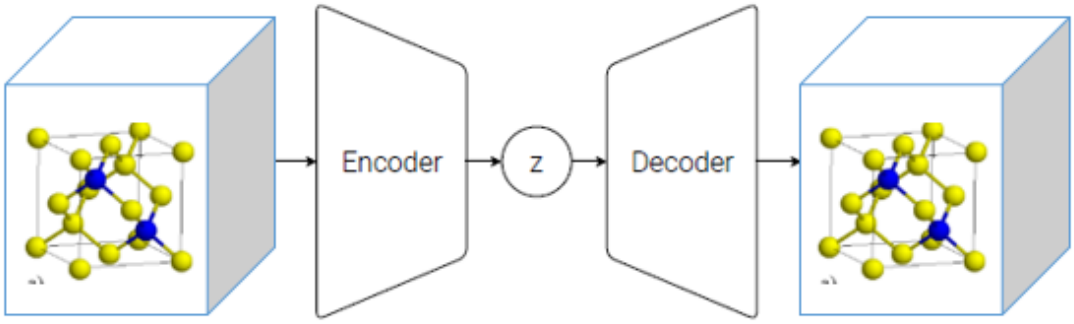

Figure 5. Vision model architecture of latent world models for processing atomic surface images.

Recurrent State Space Model: The Latent Dynamics Model helps create hypothetical trajectories by forecasting future latent states based on present states and actions. It divides the latent state into two components by combining stochastic and deterministic elements: a stochastic latent variable that

captures uncertainty and allows modeling multiple future possibilities, and a deterministic hidden state (RNN) that preserves long-term memory. By striking a balance between stability and uncertainty modeling, this method enables the model to generate multiple future predictions while preserving accurate long-term representations. Figure 6 illustrates the Recurrent State Space Model (RSSM), which is used in latent world models.

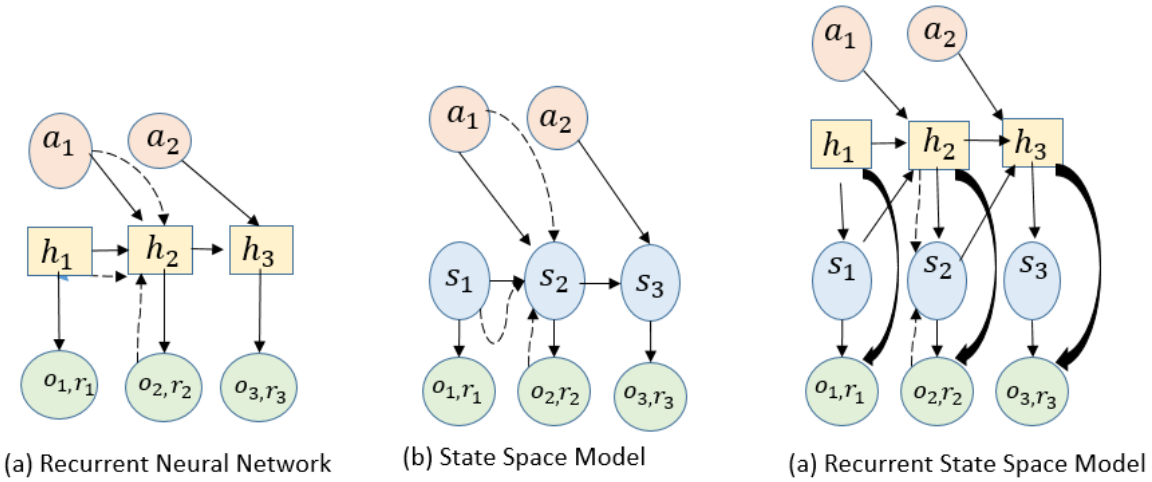

Figure 6. Latent dynamic model RSSM, a hybrid approach comprises RNN and SSM.

Actor-Critic: The Policy and Planning Module optimizes decision-making in the latent space by using hypothetical trajectories. It acquires behaviors that take into consideration rewards that are not immediately attainable. This is accomplished by training two important models in the world model's latent space: The Action Model (Policy Model), which carries out the policy by forecasting actions that maximize performance in the imagined environment, and the Value Model (Critic Model), which calculates the expected cumulative rewards from each imagined state and directs the action model toward the best course of action (Figure 7).

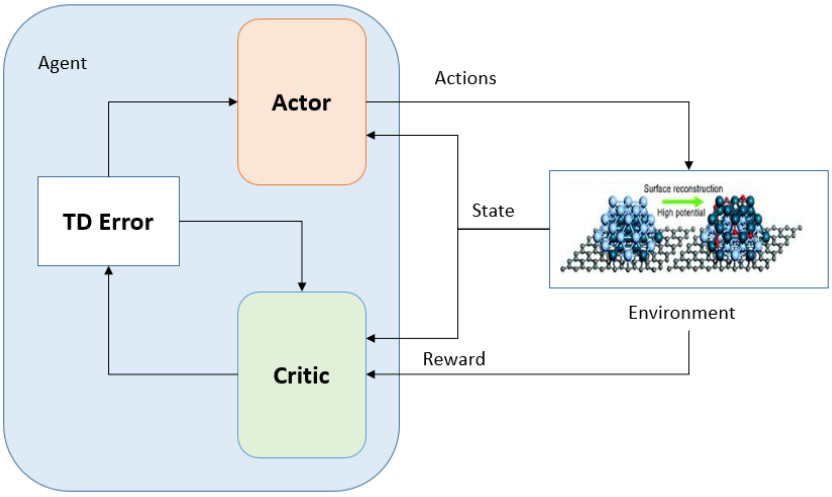

Figure 7. Action and value learning development through imaginary roll out in latent RL framework.

3.2. Historical development and progressive advances

Ha and Schmidhuber [31] significantly advanced the idea of world models in reinforcement learning by putting forth a novel framework that learns compact latent representations and models sequential transitions by combining a variational autoencoder (VAE) with a recurrent neural network (RNN)-based mixture density network (MDN-RNN). The viability of using latent imagination for control tasks in simulated environments was illustrated by this framework.

Capitalizing on this framework, Hafner *et al*. [28] introduced PlaNet, a dynamics model based on a recurrent state-space model combined with a planning mechanism that is strong in continuous control tasks. PlaNet opened the door for the Dreamer family, which improved the integration between reinforcement learning policy and world models.

3.3. Progress in latent world models: the dreamer series

With every iteration, the Dreamer series enhances the ability to learn, plan, and optimize policies in a learned latent space, marking a substantial breakthrough in latent world modeling.

DreamerV1: It introduced the basic technique of learning latent dynamics with a RSSM and then integrating it with actor-critic reinforcement learning. DreamerV1's ability to visualize trajectories in latent space and demonstrate remarkable sample efficiency on continuous control benchmarks set the groundwork for model-based reinforcement learning in complex environments [54].

DreamerV2: DreamerV2 [55] prioritized experience replay and added discrete latent variables to better capture environment stochasticity in order to improve learning focus. Its remarkable efficacy on discrete and continuous control tasks expanded its range of applications.

Continual-Dreamer: By incorporating selective experience replay techniques to avoid forgetting while learning new tasks one after the other, Continuous-Dreamer expanded DreamerV2 into the continual reinforcement learning (CRL) context. Through the introduction of task-agnostic learning capabilities, the world model was able to adjust to task changes without the need for explicit task labels. On Minigrid and Maniac benchmarks, Continual-Dreamer outperformed the most advanced task-agnostic CRL techniques, confirming the feasibility of model-based techniques in sequential multi-task settings [56].

DreamerV3: It combined large-scale training techniques and improved regularization, consolidating and expanding on earlier advancements to manage a range of action spaces and high-dimensional visual inputs. Therefore, DreamerV3 is particularly well-suited for intricate scientific applications that require precise atomic-scale interaction modeling, such as catalyst surface reconstruction [57].

The major advancements and characteristics of Dreamer variants from 2020 to 2023 are depicted in Figure 8.

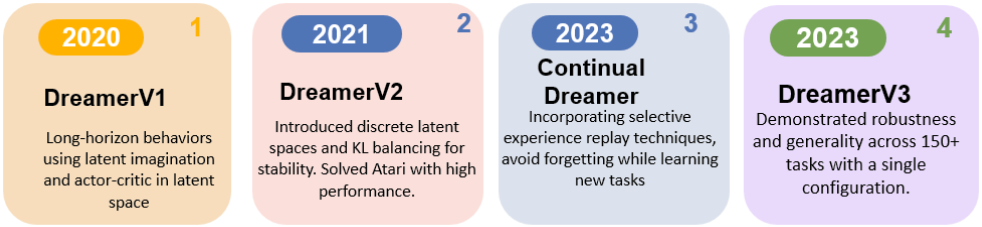

Figure 8. Key features, developments of various dreamer series frameworks from 2020 to 2023.

3.4. Current developments and uses of world models

World models were first created for fields like gaming and robotics [58], but they have since developed and shown promise in a variety of other fields. They can be successfully applied to scientific problems like electricity management cloud system [59], drug discovery, materials optimization, and surface catalyst reconstruction because of their fundamental characteristics of latent space representation, dynamics prediction, and latent imagination [33]. These models are especially useful in situations where direct environmental interaction is risky or costly.

Novel architectures and strategies are presented in recent works to address issues like task generalization, uncertainty, long-term planning, and partial observability:

IRIS Framework: To achieve state-of-the-art sample efficiency in the Atari 100k benchmark, IRIS (Imagination with auto-Regression over an Inner Speech), a reinforcement learning agent, employs a world model made up of an autoregressive Transformer and a discrete autoencoder [60]. IRIS shows that agents can effectively learn behaviors by simulating millions of trajectories with accuracy.

TrajWorld: In order to generalize across heterogeneous environments, TrajWorld [61] introduces temporal attention mechanisms and interleaved variates. Although it is very adaptable, its intricate architecture necessitates more training resources.

TransDreamer: TransDreamer enhances training stability and dynamics prediction by incorporating transformers into the MBRL framework [62] via a Transformer State-Space Model [63]. It surpasses the original Dreamer on challenging visual benchmarks and performs exceptionally well in tasks requiring long-range memory by sharing the world model with a transformer-based policy network. This demonstrates how transformers can improve world models, even in the face of real-world deployment challenges.

SafeDreamer: By incorporating safety constraints during policy learning, a critical component for applications with safety-critical considerations, SafeDreamer [64] expands upon the Dreamer framework. It may adopt conservative behaviors, which could limit exploratory performance, while guaranteeing safer exploration.

STORM: Transformer architectures and stochastic latent variables are combined in STORM [65] to improve model robustness and uncertainty, particularly for visual tasks with noisy or insufficient observations. Carefully adjusting stochastic components to strike a balance between prediction accuracy and uncertainty is essential to its operation.

Theoretical developments have given world model design significant underpinnings beyond applied frameworks. In order to learn compressed spatiotemporal representations that facilitate reinforcement learning, Ha and Schmidhuber [31] invented the use of generative recurrent neural networks, which combine VAEs with mixture density RNNs (MDN-RNN). While Qin [66] used diffusion transformers for video simulation, Kipf [67] introduced graph neural networks for structured, object-based modeling. Zhang [68] showed how language-based guidance can be incorporated into world model dynamics using encoder-decoder Transformers. Furthermore, hierarchical reinforcement learning and unsupervised learning frameworks have been investigated to enhance high-level control and long-term memory [69].

These models are grounded in theoretical frameworks such as Bayesian networks for interpretable stochastic process modeling [70], multi-timescale state-space models and hidden-parameter models to capture short-term and long-term dependencies [71], and predictive processing architectures that

incorporate sensorimotor contingencies [72]. Furthermore, it has been suggested that probabilistic programs and language models can produce structured, interpretable latent representations [73].

Each of these world model variations has distinct qualities that broaden the application and resilience of model-based reinforcement learning, creating new avenues for catalyst surface reconstruction and associated scientific issues.

4. Conceptual outlook and comparative perspectives

Conventional simulation-based and reinforcement learning frameworks face significant challenges due to the highly dynamic, stochastic, and multi-objective restructuring behaviors of catalyst surfaces like AgPd under operating conditions, as covered in Section 2.5. Latent world models offer an uncertainty-aware scalable solution that learns internal representations of surface dynamics to capture the coupled effects of defect formations, adsorbate-induced transitions, and surface energy variations on similar timescales. In complex, high-dimensional environments, these models allow agents to envision future paths, lessen their dependency on costly DFT analyses, and iteratively improve control policies.

While earlier methods like PlaNet [74] and DreamerV1 [54] provided fundamental contributions to model-based reinforcement learning. However, as Table 1 illustrates, these frameworks struggle to scale for high-dimensional, long-horizon tasks. Discrete latent representations, continual learning, and training stability were all enhanced by later models such as DreamerV2 and Continual Dreamer [56]. However, computational demands and limitations in handling complex visual inputs continue to limit their adaptability to complex, safety-sensitive scientific applications.

DreamerV3 [57] is one of the frameworks that has the most promise for applications like catalyst surface reconstruction. One of its main advantages is its capacity to control various action spaces. DreamerV3 is positioned as a particularly promising candidate when combined with enhanced regularization techniques, data-efficient, and scalable training strategies that facilitate effective learning from complex simulations. For precise modeling of atomic-scale surface interactions and configurations, DreamerV3 provides better scalability, increased stability in high-dimensional latent spaces, and better handling of intricate, dynamic visual observations as compared to DreamerV2 [55] and Continual Dreamer [75].

A strong and trustworthy world model is also necessary because catalyst surface optimization is a safety-critical process, where erroneous simulations may indicate physically unstable structures, energetically impractical surfaces, or potentially dangerous material configurations. These issues can be effectively addressed by DreamerV3, which supports simulation safety constraints (such as avoiding dangerous or unrealistic configurations) during policy learning and optimization. This is due to its ability to make long-horizon, consistent predictions within a latent imagination framework.

Based on these findings, this review presents a conceptual framework for catalyst surface optimization utilizing DreamerV3's benefits. Through a multi-objective reward balancing strategy, the proposed method would simultaneously optimize critical catalyst properties like surface energy, adsorption strength, and mechanical stability. To effectively manage trade-offs, it may use Pareto-optimal policy search frameworks or sensitivity-optimized scalarization techniques. Therefore, to manage trade-offs between catalyst performance criteria like surface energy, adsorption strength, and mechanical stability, a strong Dreamer-based catalyst optimization framework should include multi-objective reward balancing strategies. In computational catalyst discovery, this conceptual outlook emphasizes the

potential of advanced latent world models, especially DreamerV3, as fundamental instruments for developing scalable, data-efficient, and safety-conscious reinforcement learning frameworks.

Table 1. A detailed comparative analysis of various world model types.

World Model	World Model Component	Latent Space	Strengths	Limitations
PlaNet [74]	VAE + RNN	Continuous (Gaussian)	Modest and active for basic tasks.	Limited scalability, difficulties with long horizons, and high-dimensional inputs.
DreamerV1 [54]	Recurrent State Space Model	Continuous (Gaussian)	More sample effective than model-free RL.	Can undergo from uncertainty during training.
DreamerV2 [55]	Recurrent State Space Model	Discrete	Better training stability and performance.	Can still be difficulties during training.
DreamerV3 [57]	Recurrent State Space Model	Discrete	Robustness and generality; solves complex tasks like Minecraft in a single configuration across all tasks.	Significant computational cost.
Continual Dreamer [56]	Recurrent State Space Model	Discrete	Allows for constant learning without forgetting.	Inherits DreamerV2 challenges.
IRIS [60]	Discrete Autoencoder + Transformer	Discrete (Visual Tokens)	Cutting-edge sample effectiveness.	High cost of computational.
TrajWorld [61]	Attention-based Model	Implicit; uses proprioceptive data as low-dimensional vectors directly.	Adapted to diverse settings, Allows for cross-environmental dynamics.	High computational cost and inadequate scalability.
TransDreamer [62]	RSSM + Transformer	Continuous/Discrete	Improved long-term planning that accounts for dependencies.	Promising computational cost, tuning sensitivity.
SafeDreamer [64]	RSSM + Safety Constraints	Continuous	Risk-aware planning and safe policy learning.	Restricted scalability, additional overhead constraint.
STORM [65]	Variational Model + Selective Replay	Continuous	Constant learning prevents forgetting.	Weaker in stochastic/visual tasks, heavy replay cost.

Future studies will present the thorough development, application, and empirical verification of this suggested methodology. The growing potential of sophisticated latent world models, especially DreamerV3, as fundamental instruments for creating scalable, secure, and effective reinforcement learning frameworks in computational materials science is highlighted by this early conceptual outlook.

5. Challenges and future directions

Despite notable advancements, the integration of RL and latent world models into catalyst surface reconstruction continues to encounter several persistent obstacles. The following discussion highlights these challenges and recent strategies proposed to address them.

5.1. Ongoing challenges and mitigation approaches

Sample Inefficiency: RL agents typically require between 10,000 and 100,000 environment interactions to achieve policy convergence, which imposes considerable computational demands especially in atomic-level simulations.

Mitigation: Recent efforts to improve sample efficiency include the adoption of experience replay, prioritized sampling, and the use of latent imagination rollouts. These techniques enable agents to simulate long-horizon trajectories within a latent space, thereby reducing the need for costly direct interactions. Furthermore, methods such as uncertainty-driven exploration with ensemble critics facilitate more targeted learning on informative state-action pairs [76].

Sim-to-Real Gaps: Discrepancies between simulated and real-world conditions can lead to unstable or inaccurate policies—for instance, models may overestimate the stability of Pd-SA under oxidizing environments.

Mitigation: Researchers are addressing these gaps by incorporating physics-based constraints (e.g., SE(3)-equivariance) [77] to preserve key symmetries in simulated configurations. Additionally, emerging neuro-symbolic frameworks, which combine latent neural models with DFT-informed symbolic rules, show significant promise for enhancing model robustness and generalizability.

Reward Sparsity: In the context of atomic-scale surface reconstructions, rare but critical events such as defect healing and surface segregation result in sparse rewards, which can impede stable RL training.

Mitigation: Reward shaping strategies have been proposed to address this, such as designing intermediate surrogate rewards based on latent-state similarity, KL-divergence thresholds [75], or energy gradients, thereby providing smoother and more informative learning signals.

Transferability Limitations: Policies trained on a specific catalyst system (e.g., AgPd) often fail to generalize to other systems (e.g., PtCo) without substantial retraining.

Mitigation: To enhance transferability, recent approaches advocate for pretraining foundation world models on diverse catalyst datasets, followed by few-shot fine-tuning for new systems. In parallel, multi-task RL [78] and meta-learning frameworks [79] are actively being explored to facilitate cross-system generalization [80] within nanoalloy and surface catalysis applications.

In summary, although these challenges remain significant, ongoing research continues to advance both the fundamental understanding and practical capabilities of RL and latent world models in catalysis.

5.2. Future research directions

Some pretty compelling new research avenues are beginning to explore:

Neuro-Symbolic World Models: There's growing interest in combining neural latent dynamics with symbolic, physics-based constraints [81], think DFT reaction rules or adsorption site preferences. The idea is to bring together the flexibility of neural networks with the rigor of physical principles, improving not just interpretability but also how well these models match up with real-world results.

Uncertainty Aware World Model: Future catalyst world model frameworks ought to incorporate uncertainty-aware mechanisms [82] like ensemble-based uncertainty estimation, KL divergence anomaly detection, and Bayesian latent dynamics. These methods have been effectively used in molecular dynamics and robotics, and they may increase the accuracy of separating significant atomic transitions in catalyst environments from random noise.

Interpretable Latent Representations: Enhancing the interpretability of latent world models in scientific applications is another significant new direction. Although methods such as PCA and t-SNE provide initial understanding of latent state spaces, they are unable to establish a meaningful connection between latent representations and well-known catalyst descriptors like adsorption energy, coordination number, or d-band center. To find surface areas or atomic configurations affecting model predictions, future catalyst world models should investigate incorporating attention-based latent visualization mechanisms [83]. Furthermore, the credibility and interpretability of the model could be improved by using symbolic regression frameworks (like PySR) [84] to extract clear, physics-consistent relationships between latent variables and catalyst properties.

Multi-Objective Reward Balancing: In catalyst surface optimization, balancing trade-offs between several performance criteria is still a constant challenge. It is challenging for reinforcement learning frameworks to optimize all goals at once because surface energy, adsorption strength, and mechanical stability frequently interact in intricate ways. Already used in robotics and molecular systems, methods like Pareto-optimal policy search [85] and Inverse Reinforcement Learning (IRL) [86] hold promise for generating reward structures that are balanced and do not place undue emphasis on any one goal. To ensure physically meaningful, trade-off-sensitive policy learning in dynamic catalytic environments, future catalyst world models could incorporate these strategies, guided by expert-curated configurations or benchmarks validated by DFT.

Multi-Scale Modeling Integration: Another key direction involves integrating atomic-scale world models with coarser-grained or mesoscale molecular dynamics simulations. This approach enables researchers to capture both the fine atomistic details and the larger-scale reconstruction phenomena that happen in dynamic catalyst environments [87], providing a more holistic understanding of these complex systems.

Automated Experimentation and Closed-Loop RL: There's also momentum building around connecting reinforcement learning-driven catalyst models with automated synthesis and characterization platforms, like robotic TEM-STM setups [88]. By enabling real-time, iterative policy refinement through closed-loop experimental feedback, this strategy promises to accelerate model improvement and experimental discovery.

Foundation Models for Catalyst Systems: Finally, the development of large, pretrained world models trained on diverse catalyst datasets is gaining traction. These foundation models can leverage few-shot or transfer learning [89] to adapt quickly to new catalyst systems [90] deep, dramatically increasing data efficiency and model transferability.

In summary, the field is undergoing fast change, and these new avenues present intriguing chances to further catalyst discovery using clever, understandable, and effective reinforcement learning frameworks.

6. Conclusion

Catalyst surfaces are remarkably dynamic as atomic arrangements, oxidation states, and active sites shift continuously in response to changing reaction conditions. Conventional static catalyst design fails to capture this inherent reactivity. Accurately modeling and predicting such dynamic behavior demands approaches that fully embrace this complexity. RL, especially when combined with advanced latent world models like DreamerV3 or STORM, has gained traction in other disciplines for its strength in

tackling long-term, sequential decisions within intricate environments. Adopting similar strategies in catalysis research presents a promising path forward.

Despite their limited exploration within catalysis, these architectures exhibit considerable potential for simulating atomic-level surface changes, particularly when integrated with physics-informed constraints and adaptive policy optimization. This positions them as power tools for catalyst design. For instance, in systems such as AgPd nanoalloys, where Pd surface segregation significantly influences formate oxidation activity, reinforcement learning-driven latent world models could facilitate targeted exploration of complex reconstruction pathways. By utilizing latent space reasoning and possibly incorporating SE(3)-equivariant frameworks, these models offer a useful way to lessen reliance on laborious first-principles computations while maintaining accuracy with fundamental physical concepts.

Integrating recent developments in AI, reinforcement learning, and operando catalysis, this review articulates an ambitious vision for the future of catalyst design. Rather than relying solely on retrospective analysis, world models are envisioned as active participants interpreting and shaping dynamic reaction interfaces in real time. This approach aims to transform catalyst development into a predictive and adaptive process, fundamentally shifting the field toward more proactive and innovative discovery. To guarantee physically significant and reliable catalyst control systems, future frameworks should also tackle the difficulties of uncertainty quantification, multi-objective reward balancing, and latent space interpretability.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant no. 51874243, 51271148, and 50971100), the Research Fund of the State Key Laboratory of Solidification Processing (NPU), China (grant no. 2020-TS-02). The authors thank colleagues at Northwestern Polytechnical University for scholarly discussions that enriched this work.

Authors' contribution

Aisha Samreen: conceptualization, investigation, methodology, writing original draft, visualization; Muhammad Azim: formal analysis, writing review & editing; Fuyi Chen: conceptualization, supervision, writing review & editing. All authors have read and agreed to the published version of the manuscript.

Conflicts of interests

The authors declare no conflict of interest.

References

[1] Feng J, Wang X, Pan H. *In-situ* reconstruction of catalyst in electrocatalysis. *Adv. Mater*. 2024, 36(50):2411688.

[2] Wang S, Zhang Y, Liu J, Chen F, Tan T, *et al*. Structural evolution of metal single-atoms and clusters in catalysis: which are the active sites under operative conditions? *Chem. Sci*. 2025, 16:6203–6218.

[3] Kolluru A, Tran K, Grambow C, Zitnick CL, Ulissi ZW. Open challenges in developing generalizable large-scale machine-learning models for catalyst discovery. *ACS Catal*. 2022, 12(14):8572–8581.

[4] Negreiros FR, Kuntová Z, Barcaro G, Rossi G, Ferrando R, *et al*. Structures of gas-phase Ag-Pd nanoclusters: a computational study. *J. Chem. Phys*. 2010, 132(23):234703.

[5] Guo L, Zhang X, Chen M, Li B, Liu W, *et al*. Structural evolutions under surface oxidation of AgPd alloy: from orientation, composition and strain effects to catalytic application. *Appl. Surf. Sci.* 2024, 648:159026.

[6] Hu T, Zhao J, Jia W, Wang Y, Jia J. Oxidative calcination of PdCo with unexpected electrocatalytic performance for ethylene glycol oxidation. *J. Fuel Chem. Technol.* 2021, 49(6):835–843.

[7] Over H, Kim YD, Seitsonen AP, Wendt S, Lundgren E, *et al*. Atomic-scale structure and catalytic reactivity of the RuO2(110) surface. *Science* 2000, 287(5457):1474–1476.

[8] Buzková Arvajová A, Boutikos P, Pečinka R, Kočí P. Global kinetic model of NO oxidation on Pd/γ-Al2O3 catalyst including PdOx formation and reduction by CO and C3H6. *Appl. Catal. B Environ.* 2020, 260:118141.

[9] Duchesne PN, Chen G, Zheng N, Gao C, Zhang P, *et al*. Golden single-atomic-site platinum electrocatalysts. *Nat. Mater.* 2018, 17(11):1033–1039.

[10] Zhang N, Chen F, Guo L. Catalytic activity of palladium-doped silver dilute nanoalloys for formate oxidation from a theoretical perspective. *Phys. Chem. Chem. Phys.* 2019, 21(40):22598–22610.

[11] Jaeger S, Fulle S, Turk S. Mol2vec: unsupervised machine learning approach with chemical intuition. *J. Chem. Inf. Model.* 2018, 58:27–35.

[12] Yoon J, Kim Y, Kwon H, Lee J, Moon S, *et al*. Deep reinforcement learning for predicting kinetic pathways to surface reconstruction in a ternary alloy. *Mach. Learn. Sci. Technol.* 2021, 2(4):045028.

[13] Silver D, Huang A, Maddison CJ, Guez A, Sifre L, *et al*. Mastering the game of Go with deep neural networks and tree search. *Nature* 2016, 529(7587):484–489.

[14] OpenAI. OpenAI Five. Available: https://blog.openai.com/openai-five/ (accessed on 25 June 2018).

[15] Ouyang L, Wu J, Jiang X, Almeida D, Wainwright C, *et al*. Training language models to follow instructions with human feedback. *Adv. Neural Inf. Process. Syst.* 2022, 35:27730–27744.

[16] Le H, Wang Y, Gotmare AD, Savarese S, Hoi SCH. CodeRL: mastering code generation through pretrained models and deep reinforcement learning. *Adv. Neural Inf. Process. Syst.* 2022, 35:21314–21328.

[17] Goel M, Raghunathan S, Laghuvarapu S, Priyakumar UD. MoleGuLAR: molecule generation using reinforcement learning with alternating rewards. *J. Chem. Inf. Model.* 2021, 61:5815–5826.

[18] Ahmed N, Farooq MU, Chen F. Materials discovery through reinforcement learning: a comprehensive review. *AI Mater.* 2025, 1(2):1–2.

[19] Usman M, Chen F. Generation and optimization of gold nanoclusters via reinforcement learning. *Eur. Phys. J. D.* 2025, 79(5):58.

[20] Sridharan B, Sinha A, Ehara M, Priyakumar UD, Bardhan J, *et al*. Deep reinforcement learning in chemistry: a review. *J. Comput. Chem.* 2024, 45(1):1–13.

[21] You J, Ying R, Ren X, Hamilton W, Leskovec J. Graph convolutional policy network for goal-directed molecular graph generation. In *Advances in Neural Information Processing Systems 31 (NeurIPS 2018)*, Montréal, Canada, December 2–8, 2018, pp. 1–12.

[22] T Long, Y Zhang, H Zhang. Generative deep learning for the inverse design of materials. *arXiv* 2024, arXiv:2409.19124.

[23] Ng WL, Goh GL, Goh GD, Ten JSJ, Yeong WY. Progress and opportunities for machine learning in materials and processes of additive manufacturing. *Adv. Mater.* 2024, 36(34):1–15.

[24] Mubeen MA, Chen F. Deep reinforcement learning unveils ternary nanocluster configurations: a case study on $Ag_6Pd_5Cu_4$. *J. Appl. Phys.* 2025, 137(22):225101.

[25] Usman M, Chen F. Deep reinforcement learning for structural optimization and potential energy landscape of 13-atom gold nanoclusters for application in nanomaterial discovery. *ACS Appl. Nano Mater.* 2025, 8(26):13418–13428.

[26] Zhou Z, Kearnes S, Li L, Zare RN, Riley P. Optimization of molecules via deep reinforcement learning. *Sci. Rep.* 2019, 9(1):10752.

[27] Meldgaard SA, Mortensen HL, Jørgensen MS, Hammer B. Structure prediction of surface reconstructions by deep reinforcement learning. *J. Phys. Condens. Matter.* 2020, 32(40):404002.

[28] Sutton RS. Dyna, an integrated architecture for learning, planning, and reacting. *ACM SIGART Bull.* 1991, 2(4):160–163.

[29] Higgins I, Matthey L, Pal A, Burgess C, Glorot X, *et al*. DARLA: improving zero-shot transfer in reinforcement learning. *arXiv* 2017, arXiv:1707.08475.

[30] Bousmalis K, Irpan A, Wohlhart P, Bai Y, Kelcey M, *et al*. Using simulation and domain adaptation to improve efficiency of deep robotic grasping. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, Brisbane, Australia, May 21–25, 2018, pp. 4243–4250.

[31] Granger CWJ. World models. In *Forecasting in Business and Economics*, 1st ed. New York: Academic Press, 1980. pp. 201–209.

[32] Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, *et al*. Mastering the game of Go without human knowledge. *Nature* 2017, 550(7676):354–359.

[33] Driess D, Schubert I, Florence P, Li Y, Toussaint M. Reinforcement learning with neural radiance fields. *Adv. Neural Inf. Process. Syst.* 2022, 35:16931–16945.

[34] Lillicrap TP, Hunt JJ, Pritzel A, Heess N, Erez T, *et al*. Continuous control with deep reinforcement learning. *arXiv* 2019, arXiv:1509.02971.

[35] Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, *et al*. Human-level control through deep reinforcement learning. *Nature* 2015, 518(7540):529–533.

[36] Schrittwieser J, Antonoglou I, Hubert T, Simonyan K, Sifre L, *et al*. Mastering Atari, Go, chess and shogi by planning with a learned model. *Nature* 2020, 588(7839):604–609.

[37] Jaderberg M, Mnih V, Czarnecki WM, Schaul T, Leibo JZ, *et al*. Reinforcement learning with unsupervised auxiliary tasks. *arXiv* 2016, arXiv:1611.05397.

[38] Anand A, Racah E, Ozair S, Bengio Y, Côté MA, *et al*. Unsupervised state representation learning in Atari. *arXiv* 2020, arXiv:1906.08226.

[39] Schulman J, Wolski F, Dhariwal P, Radford A, Klimov O. Proximal policy optimization algorithms. *arXiv* 2017, arXiv:1707.06347.

[40] Andrychowicz M, Baker B, Chociej M, Józefowicz R, McGrew B, *et al*. What matters in on-policy reinforcement learning? A large-scale empirical study. *arXiv* 2020, arXiv:2006.05990.

[41] Buckman J, Hafner D, Tucker G, Brevdo E, Lee H. Sample-efficient reinforcement learning with stochastic ensemble value expansion. *arXiv* 2019, arXiv:1807.01675.

[42] Chua K, Calandra R, McAllister R, Levine S. Deep reinforcement learning in a handful of trials using probabilistic dynamics models. *arXiv* 2018, arXiv:1805.12114.
[43] Zhang M, Vikram S, Smith L, Abbeel P, Johnson MJ, *et al*. SOLAR: deep structured representations for model-based reinforcement learning. *arXiv* 2019, arXiv:1808.09105.
[44] Henaff M, LeCun Y, Canziani A. Model-predictive policy learning with uncertainty regularization for driving in dense traffic. *arXiv* 2019, arXiv:1901.02705.
[45] Srinivas A, Jabri A, Abbeel P, Levine S, Finn C. Universal planning networks. *arXiv* 2018, arXiv:1804.00645.
[46] Finn C, Levine S. Deep visual foresight for planning robot motion. *arXiv* 2017, arXiv:1610.00696.
[47] Plaat A. *Deep Reinforcement Learning*, 1st ed. Singapore: Springer, 2022.
[48] Watter M, Springenberg JT, Boedecker J, Riedmiller M. Embed to control: a locally linear latent dynamics model for control from raw images. *arXiv* 2015, arXiv:1506.07365.
[49] Oh J, Singh S, Lee H. Value prediction network. *arXiv* 2017, arXiv:1707.03497.
[50] Gregor K, Rezende DJ, Besse F, Wu Y, Merzic H, *et al*. Shaping belief states with generative environment models for RL. *arXiv* 2019, arXiv:1906.09237.
[51] Krishnan RG, Shalit U, Sontag D. Deep Kalman filters. *arXiv* 2015, arXiv:1511.05121.
[52] Doerr A, Daniel C, Schiegg M, van der Smagt P, Toussaint M, *et al*. Probabilistic recurrent state-space models. *arXiv* 2018, arXiv:1801.10395.
[53] Buesing L, Weber T, Racanière S, Eslami SMA, Rezende DJ, *et al*. Learning and querying fast generative models for reinforcement learning. *arXiv* 2018, arXiv:1802.03006.
[54] Hafner D, Lillicrap T, Ba J, Norouzi M. Dream to control: learning behaviors by latent imagination. *arXiv* 2020, arXiv:1912.01603.
[55] Hafner D, Lillicrap T, Norouzi M, Ba J. Mastering Atari with discrete world models. *arXiv* 2022, arXiv:2010.02193.
[56] Kessler S, Ostaszewski M, Bortkiewicz M, Zarski M, Wolczyk M, *et al*. The effectiveness of world models for continual reinforcement learning. *arXiv* 2023, arXiv:2211.15944.
[57] Hafner D, Pasukonis J, Ba J, Lillicrap T. Mastering diverse domains through world models. *arXiv* 2023, arXiv:2301.04104.
[58] Piergiovanni A, Wu A, Ryoo MS. Learning real-world robot policies by dreaming. *arXiv* 2019, arXiv:1805.07813.
[59] Mehta Y, Trivedi R, Alazab M, Gadekallu TR, Baig ZA, *et al*. Renewable electricity management cloud system for smart communities using advanced machine learning. *Energies* 2025, 18(6):1418.
[60] Micheli V, Alonso E, Fleuret F. Transformers are sample-efficient world models. *arXiv* 2023, arXiv:2209.00588.
[61] Yin S, Luo Y, Zhao Z, Lu X, Zhang Q, *et al*. Trajectory world models for heterogeneous environments. *arXiv* 2025, arXiv:2502.01366.
[62] Valencia D, Jimenez A, Castaneda J, Tavera R, Ruiz-del-Solar J, *et al*. Comparison of model-based and model-free reinforcement learning for real-world dexterous robotic manipulation tasks. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, London, United Kingdom, May 29–June 2, 2023, pp. 871–878.
[63] Chen C, Wu Y, Yoon J, Ahn S. Reinforcement learning with transformer world models. *arXiv* 2022, arXiv:2206.04176.

[64] Huang W, Ji J, Xia C, Zhang B, Yang Y. SafeDreamer: safe reinforcement learning with world model. *arXiv* 2024, arXiv:2307.07176

[65] Zhang W, Wang G, Sun J, Yuan Y, Huang G. STORM: efficient stochastic transformer-based world models for reinforcement learning. *arXiv* 2023, arXiv:2310.09615.

[66] Qin Y, Luo X, Liang Y, Zhu L, Chen H, *et al*. WorldSimBench: towards video generation models as world simulators. *arXiv* 2024, arXiv:2410.18072.

[67] Kipf T, van der Pol E, Welling M. Contrastive learning of structured world models. *arXiv* 2020, arXiv:1911.12247.

[68] Zhang A, Nguyen K, Tuyls J, Lin A, Narasimhan K. Language-guided world models: a model-based approach to AI control. *arXiv* 2024, arXiv:2402.01695.

[69] Vitku J, Cermak M, Habala O, Simko M. ToyArchitecture: unsupervised learning of interpretable models of the environment. *PLoS One* 2020, 15(5):e0230432.

[70] Da Costa L. Toward universal and interpretable world models for open-ended learning agents. *arXiv* 2024, arXiv:2409.18676.

[71] Vaisakh V, Kumar S. Learning world models with hierarchical temporal abstractions: a probabilistic perspective. *arXiv* 2024, arXiv:2404.16078.

[72] Hemion NJ. Discovering latent states for model learning: applying sensorimotor contingencies theory and predictive processing to model context. *arXiv* 2016, arXiv:1608.00359.

[73] Wong L, Zaremba W, Lake BM. From word models to world models: translating from natural language to the probabilistic language of thought. *arXiv* 2023, arXiv:2306.12672.

[74] Hafner D, Lillicrap T, Fischer I, Villegas R, Ha D, *et al*. Learning latent dynamics for planning from pixels. In *Proceedings of the 36th International Conference on Machine Learning*, Long Beach, USA, June 9–15, 2019, pp. 4528–4547.

[75] Kurian JF, Allali M. Detecting drifts in data streams using Kullback-Leibler (KL) divergence measure for data engineering applications. *J. Data Inf. Manage.* 2024, 6(3):207–216.

[76] Gong Z, Kumar A, Varakantham P. Offline safe reinforcement learning using trajectory classification. *arXiv* 2024, arXiv:2412.15429.

[77] Fuchs FB, Worrall DE, Fischer V, Welling M. SE(3)-transformers: 3D roto-translation equivariant attention networks. *Adv. Neural Inf. Process. Syst.* 2020. 33:1970–1981.

[78] Varghese NV, Mahmoud QH. A survey of multi-task deep reinforcement learning. *Electronics* 2020, 9(9):1363

[79] Akkaya I, Andrychowicz M, Chociej M, Litwin M, McGrew B, *et al*. Solving Rubik's cube with a robot hand. *arXiv* 2019, arXiv:1910.07113.

[80] Merchant A, Batzner S, Schoenholz SS, Aykol M, Cheon G, *et al*. Scaling deep learning for materials discovery. *Nature* 2023, 624(7990):80–85.

[81] Bougzime O, Cruz C, André J, Zhou K, Qi H, *et al*. Neuro-symbolic artificial intelligence in accelerated design for 4D printing: status, challenges, and perspectives. *Mater. Des.* 2025, 252:113737.

[82] Seo J, Nakamura K, Bajcsy A. Uncertainty-aware latent safety filters for avoiding out-of-distribution failures. *arXiv* 2025. arXiv:2505.00779.

[83] Wapnick S, Manderson T, Meger D, Dudek G. Trajectory-constrained deep latent visual attention for improved local planning in presence of heterogeneous terrain. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Prague, Czech Republic, September 27–October 1, 2021, pp. 460–467.

[84] Sorour SS, Saleh CA, Shazly M. Integrating machine learning and symbolic regression for predicting damage initiation in hybrid FRP bolted connections. *Sci. Rep.* 2025, 15(1):1–19.

[85] Shu T, Shang K, Gong C, Nan Y, Ishibuchi H. Learning Pareto set for multi-objective continuous robot control. *arXiv* 2024, arXiv:2406.18924.

[86] Liu Y, Wang Y, Zhou Q, Lin J, Chen X, *et al*. Defining problem from solutions: inverse reinforcement learning (IRL) and its applications for next-generation networking. *arXiv* 2024, arXiv:2404.01583.

[87] Wang L, Meng Q, Xiao M, Liu C, Xing W, *et al*. Insights into the dynamic surface reconstruction of electrocatalysts in oxygen evolution reaction. *Renewables* 2024, 2(5):272–296.

[88] Tsurusawa H, Kishimoto K, Miura A, Shibata N, Ikuhara Y. Robotic fabrication of high-quality lamellae for aberration-corrected transmission electron microscopy. *Sci. Rep.* 2021, 11(1):1–12.

[89] Zhu H, Dong Z, Topollai K, Choromanska A. AD-L-JEPA: self-supervised spatial world models with joint embedding predictive architecture for autonomous driving with LiDAR data. *arXiv* 2025, arXiv:2501.04969.

[90] Suvarna M, Pérez-Ramírez J. Embracing data science in catalysis research. *Nat. Catal.* 2024, 7(6):624–635.