AI for Materials
https://www.sciltp.com/journals/aimat

Review

Driving Catalytic Innovation: The Development and Application of Machine Learning Force Fields

Xiaoyi Hu [1], Guanjie Wang [2,*], Cuilian Wen [1] and Baisheng Sa [1,*]

[1] Materials Genome Institute, State Key Laboratory of Green and Efficient Development of Phosphorus Resources, College of Materials Science and Engineering, Fuzhou University, Fuzhou 350108, China
[2] School of Advanced Materials Innovation, University of Science and Technology Beijing, Beijing 100083, China
* Correspondence: gjwang@ustb.edu.cn (G.W.); bssa@fzu.edu.cn (B.S.)

How To Cite: Hu, X.; Wang, G.; Wen, C.; et al. Driving Catalytic Innovation: The Development and Application of Machine Learning Force Fields. *AI for Materials* **2026**, *1*(1), 5.

Received: 22 October 2025
Revised: 25 December 2025
Accepted: 26 December 2025
Published: 7 January 2026

Abstract: Catalytic reactions lie at the heart of the modern chemical industry, where their efficiency being determined by microscopic interactions occurring on the surface of the catalyst. A comprehensive understanding of these mechanisms in atomic-scale is imperative for the rational design and process optimization of chemical reactions. However, computational simulations have long grappled with inherent limitations: high-accuracy first-principles methods demand exorbitant computational resources, whereas classical force fields often fail to capture the nuanced dynamics of chemical reactions. The emergence of machine learning force fields (MLFFs) has led to significant advancements in addressing this bottleneck, offering a promising balance between precision and efficiency. This review provides a systematic overview of the evolution of MLFFs methodologies in practical catalytic applications, and highlights their respective advantages as well as the persisting challenges. By consolidating contemporary research trends and unresolved obstacles, this work provides valuable insights for subsequent researchers by synthesizing existing approaches and challenges, contributing to the refinement of MLFFs computational algorithms and their broader application in complex catalytic systems.

Keywords: machine learning; traditional molecular dynamics; machine learning force field; material design; catalytic reaction

1. Introduction

Catalytic reactions represent pivotal technologies in modern chemical industry, energy conversion, and environmental protection, with their fundamental objective being the efficient and highly selective synthesis of target products [1,2]. Conventional research methodologies predominantly rely on experimental approaches, wherein investigators systematically modulate catalyst composition and architecture while evaluating catalytic performance in reactor systems [3]. Nevertheless, when probing atomic or molecular-scale mechanistic details, experimental techniques frequently encounter substantial limitations [4]. Standard characterization methods often fail to directly capture or identify crucial reaction intermediates due to their elevated reactivity and transient stability, consequently impeding comprehensive understanding of precise reaction pathways [5]. Under realistic operational conditions, catalyst surfaces undergo continuous dynamic reconstruction, making accurate characterization of active site configurations particularly challenging [6]. Therefore, structure-activity correlations are often constrained to theoretical postulation. Furthermore, the considerable experimental expenditures and protracted time requirements severely restrict exhaustive screening of potential catalyst candidates, especially for intricately designed multicomponent or nanostructured catalytic systems [7–10]. Precisely due to these experimental challenges, computational chemistry simulations play an indispensable role in catalyst design [11].

Through first-principles calculations, we can construct atomic-scale models of catalytic active sites, accurately determining the adsorption configurations and reaction energetics of reactants, intermediates, transition states, and products on catalyst surfaces [12,13]. This enables comprehensive mapping of complete catalytic cycles, identification of rate-determining steps, and prediction of catalytic activity and selectivity [14]. Computational modeling not only provides mechanistic interpretations for experimental observations, but also offers strategic guidance for experimental work. By predicting the performance of diverse catalyst structures, it facilitates screening of the most promising candidate materials, thereby reducing research costs and enabling rational design of high-performance catalysts [15,16]. Consequently, the deep integration of computational and experimental approaches has become crucial for advancing catalytic reaction research to new levels.

The accuracy of molecular simulations represents a critical factor in investigating microscopic mechanisms of catalytic reactions, yet computational methods have long confronted the dilemma of balancing precision with efficiency [17,18]. Quantum mechanical approaches such as density functional theory demonstrate superior accuracy in describing chemical bond breaking and formation, but their substantial computational cost strictly limits simulations to systems of hundreds of atoms and picosecond timescales. This limitation significantly impedes the study of slow kinetic processes and solvent effects in catalytic reactions [19,20]. Conventional molecular dynamics simulation method, while computationally efficient, rely on predetermined functional forms that cannot accurately capture crucial transition states or electron redistribution during chemical reactions [21]. This situation creates a fundamental trade-off: either achieving high accuracy at the expense of computational speed, or maintaining efficiency while sacrificing predictive reliability, thereby obstructing comprehensive understanding of intrinsic catalytic reaction mechanisms [22]. The advantage of machine learning lies in its ability to derive patterns directly from vast datasets without requiring predefined theoretical formulas set by humans. The emergence of machine learning has introduced transformative approaches for catalytic reaction studies, as it operates without requiring pre-defined theoretical formulas, instead autonomously extracting patterns from extensive datasets [23–27]. Specifically, researchers can input substantial quantum chemical calculation results, enabling machine learning models to automatically establish correlations between atomic structures and corresponding energies or forces.

To effectively apply this capability to complex dynamic processes, a crucial bridge connecting machine learning models with molecular dynamics simulations becomes essential, with MLFFs serving as this pivotal link [28,29]. MLFFs efficiently and accurately simulate adsorption, diffusion, and reaction kinetics at catalytic active sites by autonomously learning interatomic interactions from quantum mechanical data [30–36]. Furthermore, MLFFs demonstrate remarkable generalization capability, allowing trained models to be transferred to different reaction environments, thereby enabling high-throughput screening of multi-step catalytic mechanisms [37]. Simultaneously, MLFFs maintain quantum-level accuracy while performing realistic simulations under practical catalytic conditions, establishing a robust computational foundation for rational design of high-performance catalysts [38].

MLFFs models are trained on first-principles data, with the aim of simultaneously predicting the total potential energy of an atomic system and the forces acting on each atom. This dual capability enables MLFFs to serve as surrogate potential energy surfaces, thereby supporting large-scale, long-timescale molecular dynamics simulations that can capture dynamic processes [39]. In contrast, machine learning property predictors are designed to directly map structural or compositional features to specific target properties without the need to predict atomic forces. Although property predictors are highly powerful in high-throughput screening, they cannot be used to simulate atomic trajectories or study reaction kinetics [22].

This review systematically examines the emergence context and developmental trajectory of MLFFs, alongside current methodologies with their respective strengths and limitations. It particularly focuses on investigating MLFFs applications in catalytic reaction research, emphasizing their roles in heterogeneous catalysis, reaction pathway exploration, and mechanistic studies. Building upon this foundation, the article further prospects future development trends and potential research directions for MLFFs. The primary objective is to provide valuable references for developing more accurate and efficient molecular simulation methods, thereby promoting widespread MLFFs implementation across catalysis design, materials development, and pharmaceutical research domains.

2. The Development and Architectural Evolution of MLFFs

The rapid development of machine learning is very evident from the data retrieved from the Web of Science. As shown in Figure 1a,b, since 2016, the number of published articles on machine learning has shown a significant upward trend. In terms of disciplinary distribution, the field of materials science has a particularly prominent participation in machine learning research, with related papers accounting for 8.7% of all machine learning papers,

a proportion second only to computer-related disciplines. This trend is particularly evident in the branch of computational materials science, especially in the classic research area of molecular dynamics simulations, where the introduction of machine learning techniques is triggering a paradigm shift. To understand the significance of this transformation, we first need to review the development history of molecular dynamics simulation methods.

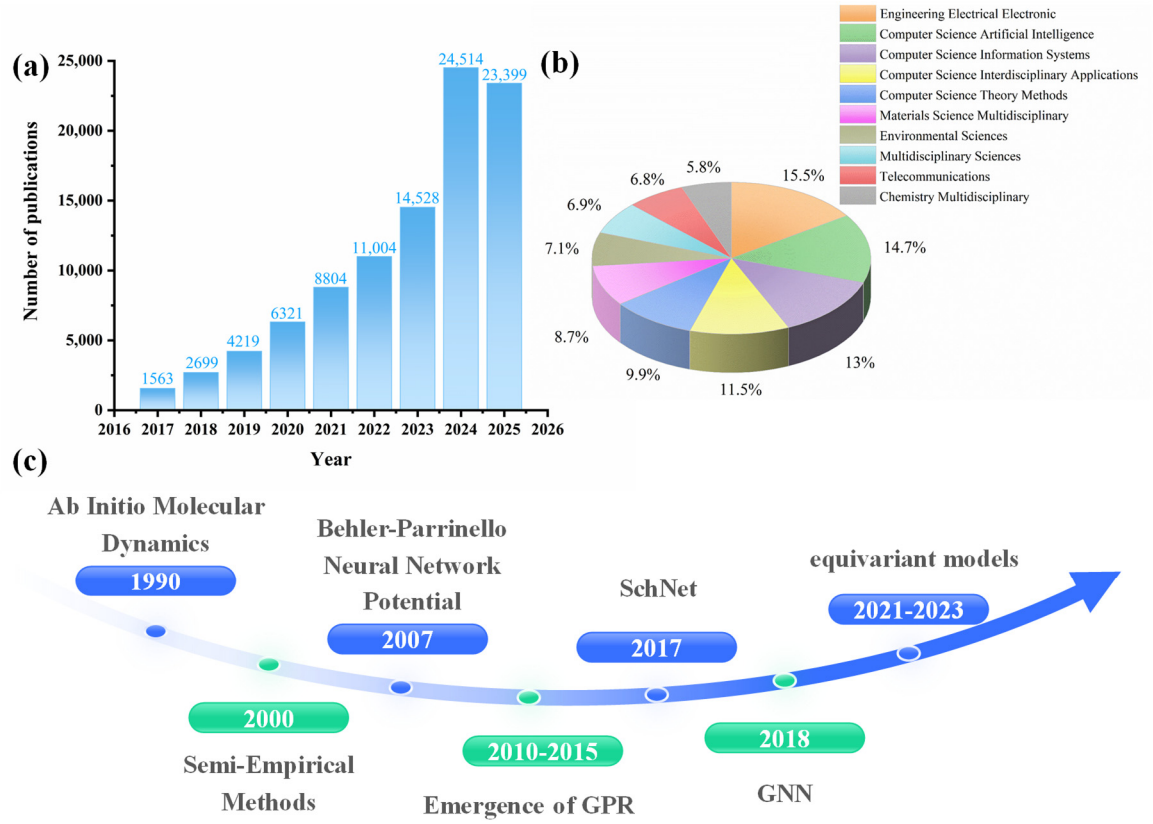

Figure 1. Trends in Machine Learning Development (based on data retrieved from Web of Science). (**a**) The number of machine learning articles published from 2016. (**b**) The proportion of machine learning papers in different fields, including materials science, relative to all machine learning papers. (**c**) The historical development from traditional molecular dynamics simulations to MLFFs.

Molecular Dynamics (MD) simulations emerged during the 1950s–1960s, with classical Newtonian mechanics and simple empirical force fields serving as their fundamental theoretical basis [40]. Early force field models were suitable for biomolecular and simple material systems, where they employed parameterization to fit experimental or quantum chemical data, but remained incapable of describing chemical reactions or electronic structure evolution [41]. As illustrated in Figure 1c, the emergence of first-principles molecular dynamics in the 1990s enabled precise quantum mechanical simulations, though the prohibitive computational cost restricted applications to small systems and short timescales [42–45]. Semi-empirical methods developed in the 2000s achieved better balance between accuracy and efficiency [46–49]. Nevertheless, they still relied on predetermined physical models with limited transferability. In conventional molecular dynamics simulations, classical force fields established a computationally feasible framework by explicitly distinguishing bonded from non-bonded interactions, subsequently building reliable parameter libraries for common molecular systems like proteins and aqueous solutions [50]. However, traditional force fields exhibit inherent accuracy limitations and dependence on fixed functional forms, preventing them from describing chemical bond breaking/formation, electronic polarization, charge transfer, transition states, or excited state phenomena [51]. Their poor generalization capability further confines applicability to specific chemical environments. The simplified treatment of long-range interactions in conventional force fields introduces significant errors in ionic solutions and interfacial systems, while failing to accurately capture complex interactions like π-π stacking and hydrogen-bond networks [52]. Assuming potential energy surfaces as minor perturbations near equilibrium states, traditional force fields prove inadequate for simulating extreme conditions such as high pressure or high temperature environments and non-equilibrium processes [53].

The evolution of MLFFs represents a coordinated advancement across theoretical foundations, core architectures, and training paradigms. In 2007, Behler and Parrinello introduced the conventional neural network potential, which employed atom-centered neural networks to describe local chemical environments, thereby demonstrating the feasibility of using machine learning to approximate high-dimensional potential energy surfaces [39,54]. Concurrently, between 2010 and 2015, kernel methods represented by Gaussian Process Regression (GPR), including implementations like sGDML, established an alternative robust technical pathway [55]. The emergence of deep learning, particularly with growing emphasis on invariance and equivariance in atomic system modeling, catalyzed transformative progress in MLFFs development [56]. MLFFs architectures subsequently underwent distinct evolutionary phases: progressing from early convolutional neural networks like SchNet in 2017, to graph neural networks such as CGCNN in 2018 that better represent chemical structures, and accelerating toward equivariant models including NequIP, MACE, Equiformer, and eSCN during 2021–2023 [57,58]. These architectures, by rigorously embedding physical symmetries, achieved quantum leaps in accuracy, data efficiency, and convergence speed, while spawning specialized foundation models like MACE-OFF and So3krates (So3LR) for targeted applications. Parallel to architectural innovations, training methodologies and data ecosystems advanced synchronously [59]. Active learning strategies substantially reduced dependency on DFT calculations through iterative sampling optimization, thereby lowering training costs. Meanwhile, the establishment of large-scale databases provided both standardized training datasets and crucial benchmarking platforms [60]. These coordinated breakthroughs across architecture, algorithms, and data resources collectively transformed MLFFs from theoretical concepts into practical research tools [61]. The subsequent discussion will systematically trace this developmental trajectory, examining technological evolution and core innovations from traditional neural network potentials to cutting-edge pretrained foundation models.

2.1. Traditional Neural Network Potential (NNP)

NNP establish their core concept by decomposing complex potential energy surfaces into summations of localized atomic energy contributions, representing the foundational architecture for MLFFs. Their theoretical foundation originates from the Behler-Parrinello scheme, which proposed that manually designed symmetry functions should transform the local chemical environment of central atoms into mathematical descriptors, thereby ensuring translational, rotational, and permutational invariance [39]. These descriptors subsequently serve as inputs to feedforward neural networks to approximate each atom's contribution to the total system energy. NNP initially demonstrated that neural networks could learn atomic interactions with quantum-mechanical accuracy while significantly improving computational efficiency [62]. However, their performance critically depends on descriptor quality and completeness, with manual optimization of optimal descriptors presenting substantial challenges [61]. Furthermore, their strict locality assumptions inherently limit the capacity to describe long-range interactions.

2.2. Kernel Methods

Kernel methods represent another significant paradigm in MLFFs, developing concurrently with neural network approaches. The core concept of kernel methods involves predicting energies and forces for new atomic configurations not through explicit parametric functions, but by measuring similarity between new configurations and known structures in quantum mechanical databases [55]. Exemplified by GAP and gradient-domain machine learning, these methods operate on the principle that potential energy surfaces follow Gaussian processes, where predictions emerge as weighted averages of training set outputs, with weights determined by kernel functions [63]. As non-parametric models, they demonstrate remarkable robustness with small datasets and naturally provide uncertainty quantification, thereby offering crucial guidance for active learning strategies [64]. However, their computational cost escalates dramatically with training set size, typically ranging between $O(N^2)$ and $O(N^3)$, which restricts applications with very large datasets and renders their scalability inferior to parametric neural network models [65].

2.3. Invariant Graph Neural Network

The development of invariant graph neural networks, exemplified by SchNet and CGCNN, marked MLFFs's entry into the deep learning era by naturally representing atomic systems as graph structures with atoms as nodes and chemical bonds as edges. This framework enables models to automatically learn vector representations of atomic environments through message-passing mechanisms, thereby reducing dependency on handcrafted descriptors [66]. Their operational principle involves learning node and edge features, then iteratively propagating and updating information across graph neighborhoods, ultimately forming globally invariant graph-level representations to predict total system energy [67]. These architectures inherently incorporate translational,

rotational, and permutational invariance to ensure physical consistency. They demonstrate superior capability in automatically capturing complex interatomic interactions, with generalization performance significantly surpassing traditional NNP [68]. Atomic forces are computed via automatic differentiation, though invariant graph neural networks typically employ scalar fields to learn potential energy surfaces. While computationally efficient, this approach lacks strict equivariance and may introduce theoretical inaccuracies that potentially compromise predictive precision [69].

2.4. Equivariant Graph Neural Network

Equivariant graph neural networks currently represent the cutting edge of MLFFs development, rigorously embedding physical symmetries including rotation, translation, and mirror symmetry directly into their architectural design to ensure coordinated transformation of output physical quantities according to proper geometric rules when input atomic configurations undergo these transformations. The rigorous embedding of such physical symmetries lies at the core of these models. Specifically, E(3) equivariance requires that the model's outputs transform covariant under the full Euclidean group in three-dimensional space, reflecting the complete symmetry of real physical systems [67]. SO(3) equivariance is a more commonly adopted constraint in many implementations, demanding covariance only under pure rotations. For most catalytic systems where chirality is not a primary concern, enforcing either E(3) or SO(3) equivariance is crucial. By directly incorporating these geometric transformation rules into the model's architecture, the model is inherently prevented from learning non-physical relationships that violate these symmetries [67]. This significantly reduces the effective dimensionality of the hypothesis space that the model must explore during training. Consequently, the model can extract correct physical patterns from substantially less training data, leading to remarkable improvements in data efficiency, generalization accuracy, and training convergence speed [61]. Representative implementations such as NequIP, MACE, and Allegro demonstrate this paradigm [70,71]. These equivariant models utilize spherical harmonics and tensor products to construct equivariant features, maintaining transformation properties throughout all message-passing steps, thereby enabling simultaneous prediction of energies and forces without requiring energy differentiation [69]. This fundamental innovation delivers exceptional data efficiency, enhanced accuracy, and accelerated convergence rates. However, the model architectures generally exhibit greater complexity with consequently increased computational overhead, while their implementation and optimization demand higher technical proficiency from researchers [61].

2.5. Transformer Architecture

Transformer-based architectures, exemplified by Equiformer [72], introduce a novel paradigm for MLFFs by employing self-attention mechanisms that enable global, pairwise direct interactions among atoms, thereby transcending the limitations of conventional graph neural networks confined to local neighborhood message passing [73]. Their fundamental principle treats atomic systems as fully connected graphs, where dynamically computed attention weights determine interatomic correlations across all possible pairs, subsequently aggregating information to update atomic representations [74]. These architectures demonstrate particular suitability for modeling phenomena like charge fluctuations and polarization effects that challenge traditional GNN, owing to their inherent capability to capture long-range interactions. However, the computational complexity of self-attention mechanisms scales quadratically with atom count, resulting in substantial computational costs that currently restrict applications to large-scale systems [75].

2.6. Foundation Model

Foundation models represent an emerging trend in MLFFs development, involving large-scale pretraining on extensive and diverse first-principles datasets to establish universal, broadly knowledgeable foundational potential functions, with MACE-OFF and So3krates (So3LR) serving as representative examples [59,76]. Their underlying principle parallels large language models in natural language processing, where learning from comprehensive datasets encompassing elements, crystal structures, and molecular configurations internalizes fundamental physicochemical principles to create a powerful knowledge baseline [77]. For new systems, users can achieve quantum-mechanical accuracy without full retraining, requiring minimal or even no fine-tuning, thereby significantly reducing implementation barriers. Consequently, these models demonstrate remarkable transferability and user accessibility, though their pretraining entails substantial computational costs and their performance in unconventional or extreme chemical environments requires further validation [78]. However, catalytic applications face specific challenges such as accurate prediction of adsorption energies on complex surfaces and precise description of localized charge transfer. Researchers have developed specialized architectures

optimized for catalytic systems to address these challenges, achieving remarkable performance on benchmark tests such as the Open Catalyst Project. For instance, EquiformerV2, through its improved equivariant attention mechanism, has achieved leading performance on the OC20 benchmark, demonstrating the potential of Transformer architectures in handling long-range interactions and complex geometries within catalytic systems [79,80]. Models based on higher-order tensor representations excel at capturing subtle orientation and bonding interactions of surface adsorbates by constructing richer geometric descriptors, and these models consistently rank among the top performers on OC benchmark leaderboards [80]. These advances indicate that sustained improvement in catalytic performance depends not only on data scale but also on model inductive biases specifically optimized for the physics of surface science. Although general benchmarks for inorganic materials, such as Matbench, are not specifically designed for catalysis, their evaluation of properties such as formation energy and band gap also provides valuable complementary insights into the predictive capabilities of MLFFs models regarding catalyst bulk stability and electronic structure-related attributes [81].

3. Software Ecosystem for Developing and Applying MLFFs

3.1. Core Software Libraries for MLFFs Development

Core training framework software forms the foundation of the entire MLFFs workflow, providing researchers with the underlying architecture to construct high-accuracy, customized potential functions from DFT data [82]. These software packages are predominantly developed based on the principles of equivariant graph neural networks, incorporating mathematical tools from group theory such as irreducible representations and spherical harmonics to ensure strict adherence to rotational and translational symmetries in three-dimensional space for both energy and force predictions. By embedding physical laws directly into machine learning models, this approach guarantees fundamental physical consistency [83]. NequIP established itself as the pioneering software in this domain, while MACE and Allegro currently represent the cutting-edge technologies due to their exceptional accuracy and computational efficiency. DeepMD-kit remains one of the most widely adopted packages in practical applications, and SchNet maintains historical significance for its substantial impact on the development of deep learning potentials [67].

Taking the widely adopted DeepMD-kit as an example, Wang et al. [82]. systematically elaborated its design principles and performance metrics in their publication. Figure 2a illustrates the software architecture and workflow, demonstrating integration with TensorFlow for automated training and its implementation as a plugin for simulation packages like LAMMPS, thereby bridging the complete pathway from data processing to molecular simulation. The training procedure incorporates continuous monitoring through test sets to prevent overfitting, ensuring robust model generalization as visualized in Figure 2b. Performance evaluation reveals rapid convergence during water model training, with energy and force errors reduced to minimal levels (2–3%) in the learning curve shown in Figure 2c. Figure 2d presents comparative analysis between DeepMD and first-principles calculations, where nearly identical radial distribution functions confirm that the generated potential functions not only fit training data but also reliably predict and reproduce genuine physical properties.

In the work by Ilyes Batatia et al. [84], the fundamental principles and innovations of MACE were comprehensively introduced. Conventional two-body message-passing networks typically require five to six layers, whereas MACE achieves convergence with merely two message-passing iterations by constructing four-body interactions, substantially reducing computational overhead. Researchers observed that even without incorporating equivariant information, the integration of higher-order features significantly alters the learning curve slope, indicating the model's enhanced capacity to extract more information from limited datasets. Through comparative analysis of energy distributions across three dihedral angle sections, systematically evaluates the performance of BOTNet, NequIP, MACE, and DFT calculations. The results demonstrate that MACE not only generates exceptionally smooth potential energy surfaces but also achieves remarkable alignment with DFT benchmarks, confirming its superior extrapolation capabilities.

Collectively, software frameworks represented by NequIP, DeepMD-kit, MACE, and Allegro have continuously advanced through stricter symmetry constraints and higher-order interaction modeling at the theoretical level. Simultaneously, they have optimized training workflows while enhancing model generalization capability and simulation reliability at the engineering level. These systematic breakthroughs in accuracy, data efficiency, and transferability and generalization ability are progressively narrowing the gap between machine learning potentials and DFT calculations. Consequently, they are establishing a robust foundation for large-scale molecular dynamics simulations of complex systems.

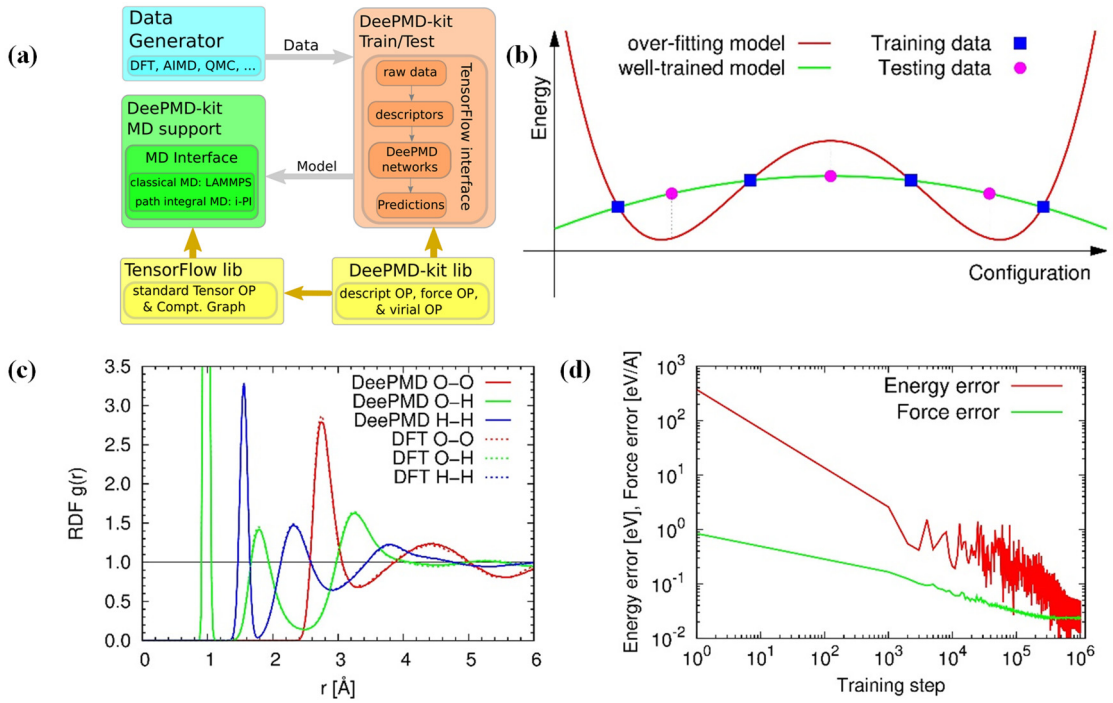

Figure 2. (**a**) Schematic plot of the DeePMD-kit architecture and the workflow. (**b**) Schematic illustration of over-fitting. (**c**) The radial distribution functions of the DeePMD compared with the PBE0+TS DFT water model. (**d**) The learning curves of the liquid water system [82]. Copyright 2018 Elsevier.

3.2. Integrated Platforms for Automated Training and Simulation

Integrated simulation platforms within MLFFs software unify critical components including potential function training, molecular dynamics simulation, and enhanced sampling techniques within a cohesive automated framework [85]. These platforms execute initial simulations while leveraging built-in uncertainty quantification modules to automatically identify new atomic configurations with low prediction reliability. Subsequently, DFT calculations validate these uncertain configurations, incorporating them into training datasets to iteratively refine potential functions. This methodology enables efficient and automated exploration of unknown phase space while ensuring model robustness [86]. The sGDML framework represents a significant implementation of this integrated approach, delivering a comprehensive high-precision environment spanning from accurate force field construction to dynamical simulations, though primarily focused on small molecular systems [87].

In the work by Klaus-Robert Müller et al. [88], the sGDML workflow was systematically presented, as illustrated in the upper portion of Figure 3, demonstrating how user-provided molecular configurations with energy and force labels are transformed into cross-validated, ready-to-use force field models through an automated pipeline. The lower portion of Figure 3 reveals how these lightweight models seamlessly interface with mainstream simulation engines, enabling researchers to perform structural optimization and compute vibrational spectra through Atomic Simulation environment (ASE) integration, while achieving path-integral molecular dynamics simulations via deep interoperability with i-PI [89,90]. This demonstrates how researchers can complete the entire workflow from force field construction to final property prediction within a unified framework, substantially streamlining high-precision computational workflows.

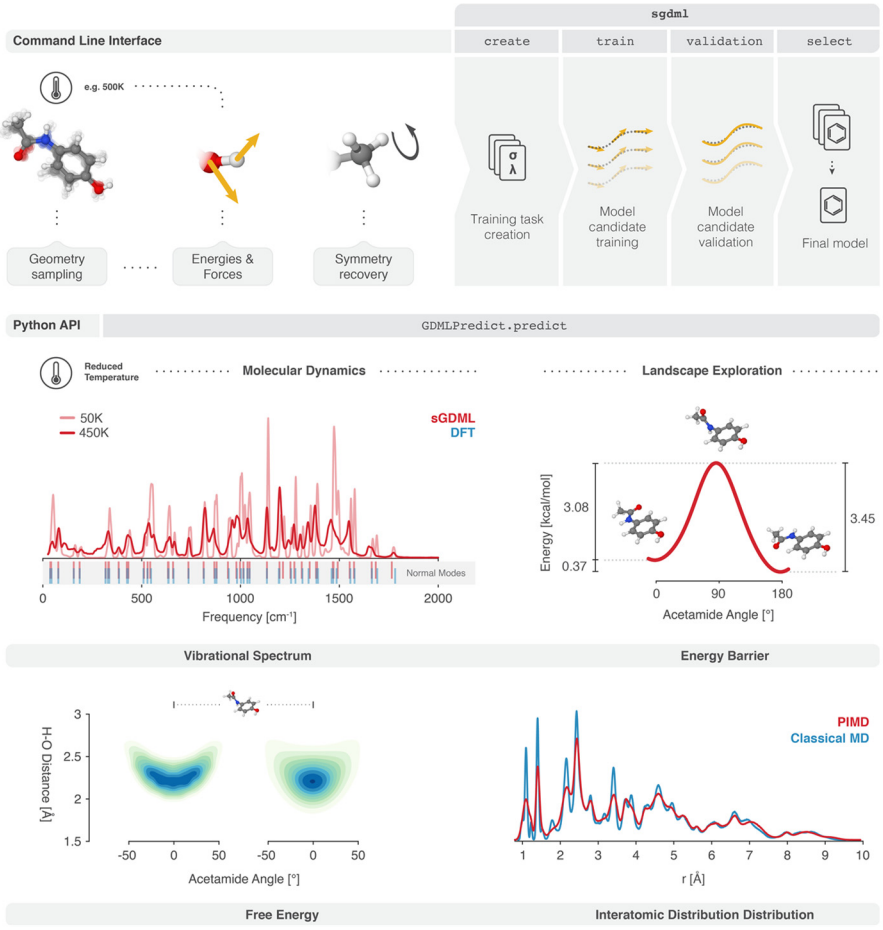

Figure 3. Schematic illustration of the integrated sGDML workflow, from force field construction to property prediction through seamless integration with ASE and i-PI for structural optimization, vibrational spectra calculation, and path-integral molecular dynamics simulations [88]. Copyright 2019 Elsevier.

3.3. Pretrained Model Repositories and Foundation Models

The development of pretrained model repositories aims to lower the entry barrier for MLFFs applications, enabling users to bypass expensive and specialized data generation and model training processes while directly performing high-accuracy molecular dynamics simulations [76]. The underlying principle involves pretraining on extensive datasets spanning diverse chemical spaces, allowing models to learn broad interatomic interaction patterns, thereby requiring only minimal fine-tuning with limited data when adapting to new systems [77]. Current representative implementations include MACE-OFF, which provides universal potential functions covering most elements of the periodic table based on the MACE architecture, alongside the ANI series that specifically focuses on organic molecules and biologically relevant systems.

Researcher Gábor Csányi and his team developed MACE-OFF [76], demonstrating exceptional accuracy and generalization capability. As illustrated in Figure 4a, across diverse test sets including small molecules, dimers, water clusters, and peptides, MACE-OFF achieves energy and force prediction errors well below the chemical accuracy threshold (1 kcal/mol). Remarkably, it maintains outstanding performance even on tripeptide systems completely absent from its training data, confirming its robust and transferable generalization. For the more challenging task of predicting molecular dihedral rotation energy barriers, MACE-OFF predictions show excellent agreement with high-level quantum chemical benchmarks. Its errors are substantially lower than those of traditional empirical force fields, semi-empirical quantum chemical methods, and earlier machine learning potentials like ANI, as quantitatively demonstrated in Figure 4b. The development of pretrained model libraries indicates that the research focus of MLFFs is shifting from tool development to providing readily usable services. By sharing pre-trained parameters, these models help researchers avoid redundant foundational modeling work, enabling them to concentrate more effectively on addressing specific scientific challenges.

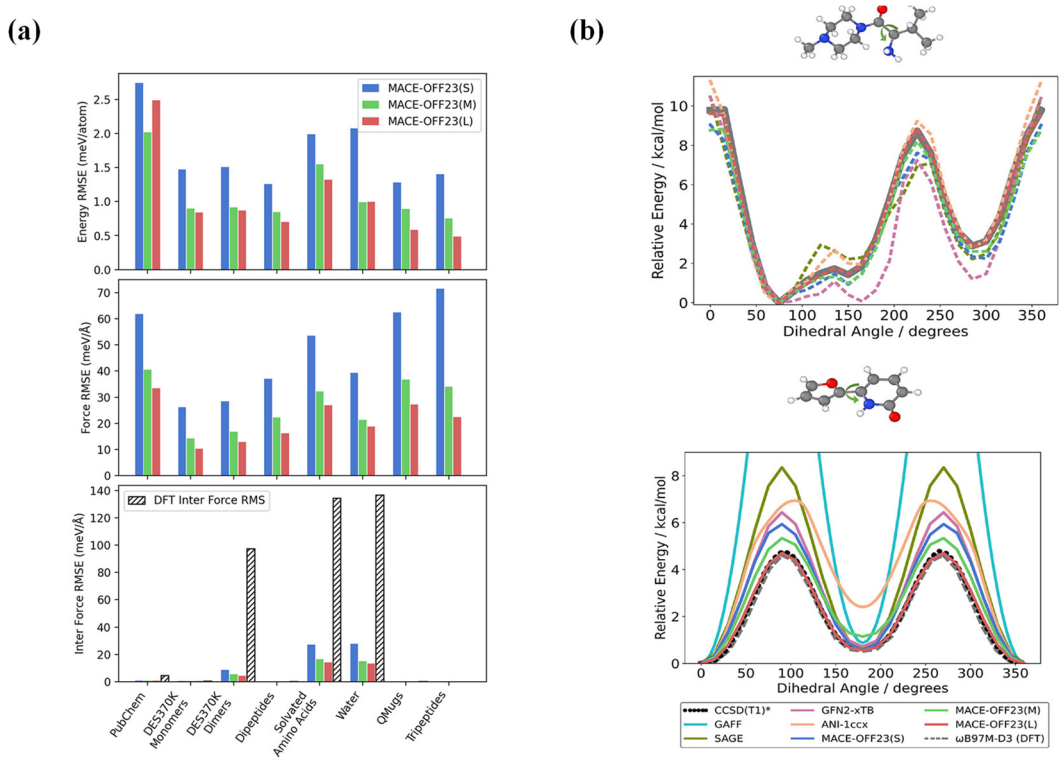

Figure 4. Performance evaluation of MACE-OFF23 models. (**a**) Test set root-mean-square errors (RMSE) for energy and force predictions across diverse molecular systems, with bottom panels specifically comparing intermolecular force errors against DFT reference magnitudes. (**b**) Dihedral angle benchmark scans: torsion drive analysis for the TorsionNet-500 dataset (**top**) and detailed torsion behavior between aromatic rings in a biaryl benchmark (**bottom**) [76]. Copyright 2025 American Chemical Society.

3.4. General-Purpose Auxiliary Tools for Catalysis Modeling

General-purpose auxiliary tools, such as the Atomic Simulation Environment (ASE), provide essential support for MLFFs workflows in catalysis research. They enable the creation and manipulation of surface and interface models, execution of tasks such as transition state searches and reaction path analysis, and integration with MLFFs and first-principles codes [91]. As a complement to ASE, pymatgen is a powerful Python library for materials analysis, offering comprehensive functionalities including crystal structure manipulation, phase diagram construction, composition analysis, and high-throughput computational data management, making it a standard tool in materials informatics for catalysis research [92]. Fairchem serves as the core toolkit of the Open Catalyst Project, providing specialized models, data loaders, preprocessing utilities, and evaluation metrics specifically tailored for catalytic surface modeling [80,93]. These tools help automate and standardize computational workflows, enhancing reproducibility and efficiency in catalytic simulations [90].

In the research by Hjorth Larsen et al. [90], ASE demonstrated its capability to construct complex periodic crystal structures such as Beryl with ease, utilizing space groups and unit cell parameters as illustrated in Figure 5a. The platform further provides comprehensive atomic-scale algorithms to validate MLFFs accuracy in transition state regions, exemplified in Figure 5b where the Nudged Elastic Band method calculates energy barriers for gold atom diffusion on aluminum surfaces. Figure 5c presents ASE's integrated database module and web interface, facilitating efficient management, querying, and visualization of extensive computational datasets, thereby enhancing workflow efficiency in MLFFs research. As shown in Figure 5d, ASE incorporates global optimization tools including genetic algorithms to automatically identify the most stable configurations for surface adsorption systems, collectively demonstrating its robust capabilities as an integrated platform for supporting materials design.

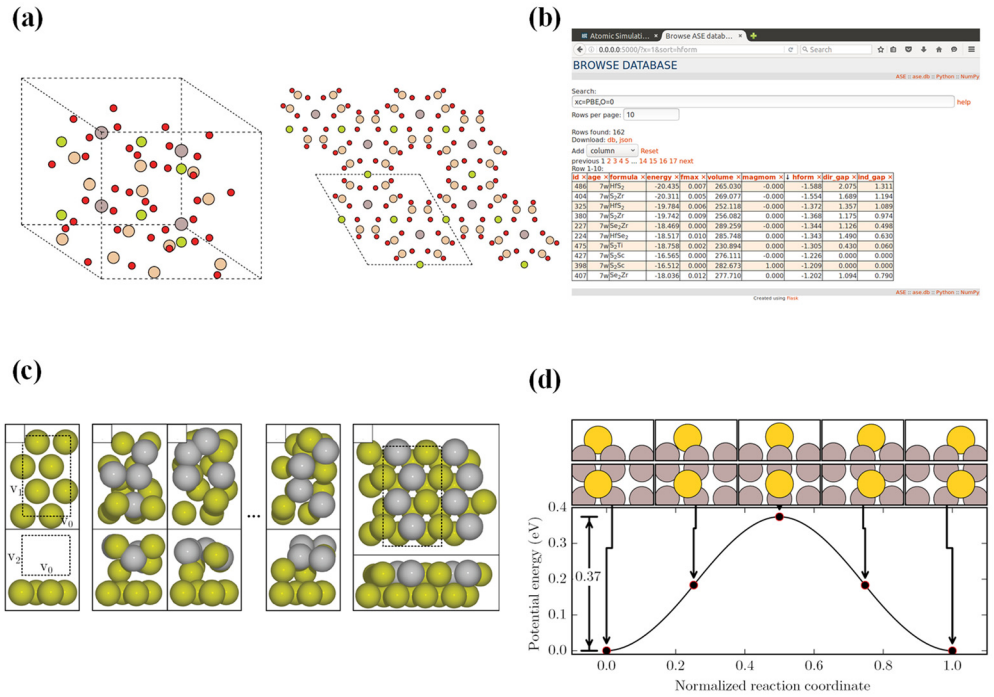

Figure 5. Functional demonstrations of the ASE platform. (**a**) The unit cell of beryl crystal structure, displayed along the [001] direction with two repeated cells. (**b**) Database query interface showing the first 10 rows of results filtered by xc = PBE and O = 0, sorted by formation energy (hform). (**c**) Surface slab construction workflow: predefined simulation box with randomized initial atom positions, three candidate structures generated by the start generator, and the global minimum surface structure optimized using the EMT potential. (**d**) Potential energy profile along the normalized reaction coordinate for the diffusion of a Au atom on the Al(001) surface [90]. Copyright 2017 IOP Publishing Ltd.

4. Practical Applications of MLFFs in Catalysis

The application of MLFFs to catalytic reaction research relies not only on high-precision models but also on data infrastructure, evaluation standards, and reliable workflows oriented toward real-world catalytic problems. In recent years, the establishment of large-scale benchmark datasets, such as the Open Catalyst Project (OC20/OC22), has fundamentally advanced this field [22]. These datasets systematically encompass diverse adsorption configurations on metal, alloy, and oxide surfaces, with the core objective of evaluating model performance in predicting energies and forces for realistic catalytic interfaces—including defects, steps, and co-adsorption phenomena [92]. This benchmark has become a crucial criterion for assessing whether MLFFs can transition from proof-of-principle demonstrations to practical applications, and has fostered the development of automated workflows integrating active learning, uncertainty quantification, and high-throughput screening [94]. Driven by this framework, MLFFs have been successfully applied to the study of numerous key catalytic reactions, enabling simulations of surface defect dynamics, interfacial solvation effects, and diffusion and reaction processes in porous materials under near-realistic conditions. These simulations provide deep insights into active site structures, reactant dynamic behavior, and complete microscopic catalytic cycles, thereby establishing a solid theoretical and simulation foundation for the rational design and optimization of high-performance catalysts [95].

Subsequently, we will examine the current applications of MLFFs in catalytic reaction simulations. Based on this foundation, potential directions for improvement and development space will be proposed.

4.1. Application of MLFFs in HER

The Hydrogen Evolution Reaction (HER) serves as the fundamental half-reaction in electrochemical water splitting, representing a critical pathway for establishing a green hydrogen economy and achieving decarbonization of energy systems [96,97]. The comprehension of its microscopic mechanisms, in conjunction with the development of catalysts exhibiting optimal efficiency, constitutes a persistent challenge that has persisted for decades [98]. Conventional HER research relies on electrochemical measurements and various ex situ characterization techniques, which can evaluate overall catalyst performance but struggle to directly capture

atomic-scale dynamics at the electrode-electrolyte interface during reactions [99]. These limitations include real-time adsorption/desorption behavior of reaction intermediates such as H*, and detailed transport mechanisms of hydronium ions through interfacial water networks [100,101]. Computational simulations, particularly DFT calculations, have played an indispensable role in investigating these microscopic properties, successfully establishing volcano plot relationships between microscopic descriptors like hydrogen adsorption free energy and macroscopic catalytic activity [102–104]. However, the substantial computational cost of DFT restricts simulations primarily to static, idealized surface models at absolute zero temperature [105]. Thereby preventing accurate treatment of finite-temperature effects including atomic thermal vibrations and entropy contributions, explicit solid-liquid interfaces with dynamic solvation effects, catalyst reconstruction processes, and realistic kinetic pathways with energy barriers for chemical reactions [106]. Consequently, MLFFs have emerged and rapidly developed as primary tools for investigating HER dynamic mechanisms [107]. During their initial development phase, research primarily focused on methodological validation and foundation establishment, aiming to demonstrate that MLFFs could achieve DFT comparable accuracy in describing key fundamental systems. These included hydrogen adsorption on metal surfaces and structural/dynamic properties of water clusters, while simultaneously enhancing computational efficiency [108]. Subsequently, MLFFs research has progressed to addressing more complex challenges, entering a phase dedicated to simulating realistic chemical reaction interfaces. Therefore, Abbas et al. [109]. Building upon the structural tunability of single-atom configurations and the high catalytic activity of Mo_3P, systematically explored the configurational space of ternary molybdenum sulfur phosphorus compounds through MLFFs driven moment tensor potential methodology combined with particle swarm optimization and artificial bee colony algorithms, successfully predicting novel Mo_2SP and Mo_3SP material systems. The fundamental advantage of this approach lies in its integration of DFT-level accuracy with MLFFs computational efficiency, as demonstrated in Figure 6a where MLFFs-predicted energies show excellent agreement with DFT benchmarks, ensuring search reliability while the convex hull diagram confirms the method's powerful discovery capability by identifying thermodynamically stable new phases. Furthermore, the trained MLFFs model enabled tasks impractical for direct DFT computation, as shown in Figure 6b, MLFF-driven molecular dynamics simulations validate the thermal stability of the predicted materials at room temperature over nanosecond timescales, which highlights the key capability of MLFFs in enabling large-scale, long-timescale dynamical simulations. This further underscore the indispensable role of MLFFs in assessing dynamic stability and structural evolution under realistic working conditions.

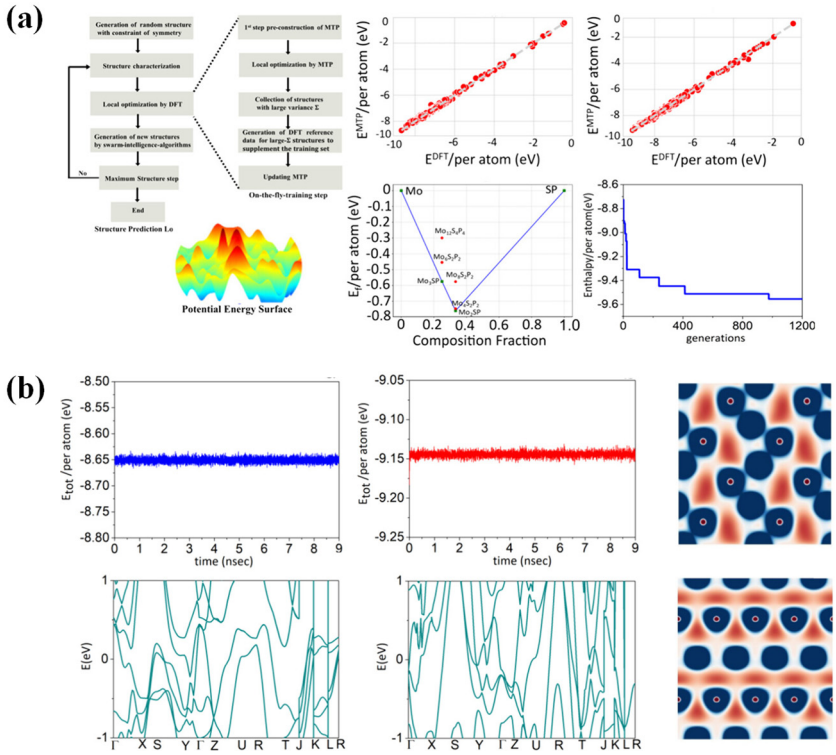

Figure 6. (**a**) Workflow of crystal structure prediction via on-the-fly training of the moment tensor potential (MTP), an MLFFs method, enabling efficient exploration of complex potential energy surfaces. (**b**) MLFFs-enabled multiscale characterization: molecular dynamics simulations using the trained MTP validate kinetic and thermal

stability, while DFT-calculated band structures and electron localization function (ELF) maps reveal the electronic and bonding properties of Mo_2SP and Mo_3SP [109]. Copyright 2024 American Chemical Society.

4.2. Application of MLFFs in CO_2RR

Against the backdrop of addressing global climate change and promoting sustainable development, the efficient catalytic conversion of CO_2 into high-value chemicals represents a crucial pathway toward achieving carbon recycling [110]. In the design of catalysts for the CO_2 reduction reaction, researchers face a fundamental conflict between the vast exploration space and the prohibitive computational cost of DFT calculations [40,111–114]. In the study by Tran and Ulissi [94], a high-throughput computational framework integrating MLFFs with active learning was developed, where local chemical environments of adsorption sites were numerically encoded as feature descriptors to train surrogate models for rapid prediction of the key reaction descriptor CO adsorption energy. The core advantage of this computational framework lies in its closed-loop iterative optimization mechanism, as illustrated in Figure 7a, where the model not only predicts adsorption energies but also evaluates its own predictive uncertainty to intelligently select the most computationally valuable candidate structures for subsequent DFT calculations. These newly generated DFT data points are continuously incorporated into training, progressively enhancing the model's accuracy and generalization capability. This strategy achieved optimal allocation of computational resources, completing 42,785 DFT calculations across a vast space encompassing 31 elements and 1499 intermetallic compounds, while successfully identifying 131 catalytic surfaces with near-ideal adsorption energies across 54 alloys. This research transcends conventional catalyst screening, as the generated large-scale, high-quality dataset enables global analysis of catalytic trends. The bimetallic activity map shown in Figure 7b reveals structure-activity relationships between different elemental pairings and CO_2 Reduction Reaction (CO_2RR) performance, while the distribution of active sites in the reduced-dimensional space presented in Figure 7c identifies key local coordination patterns associated with high activity through structural similarity analysis. This work empirically demonstrates that MLFFs combined with active learning strategies can advance computational catalysis from discrete case studies to systematic material design and pattern discovery, establishing a methodological foundation for efficient exploration of complex catalytic systems.

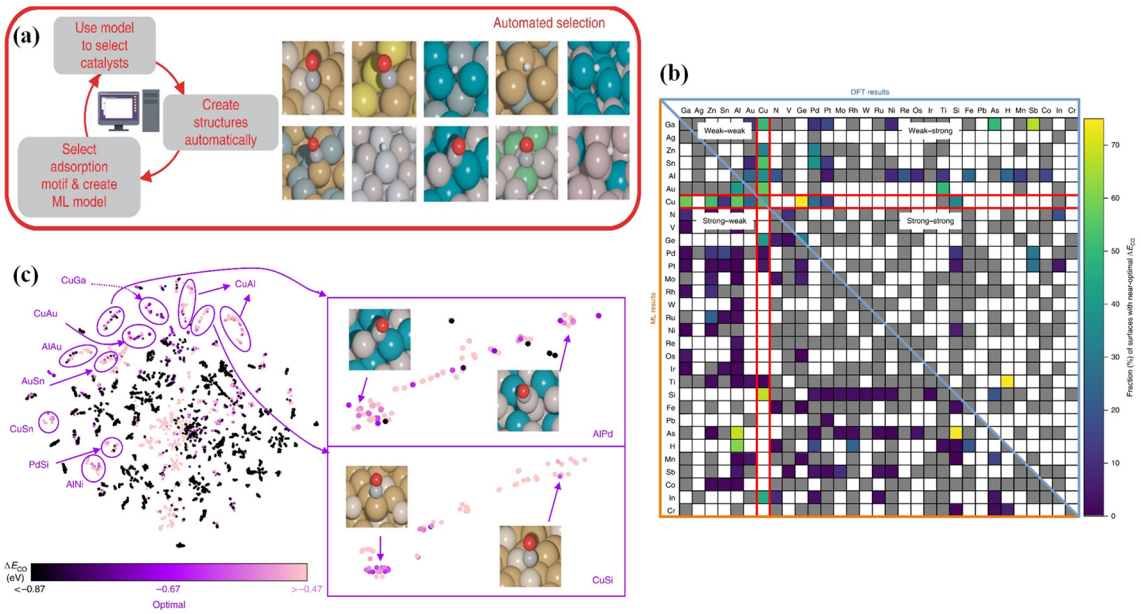

Figure 7. Active learning framework for high-throughput screening of CO_2 reduction electrocatalysts. (**a**) Workflow (highlighted in red) illustrating the systematic and automated machine learning-guided candidate selection strategy using active learning and surrogate-based optimization. (**b**) CO_2 reduction activity map for bimetallic materials, visualizing catalytic performance trends across elemental combinations. (**c**) Active site motif analysis: t-SNE40 visualization of the latent space constructed from all DFT-simulated adsorption sites, revealing structural patterns associated with catalytic activity [94]. Copyright 2018 Springer Nature.

4.3. Application of MLFFs in Oxygen Reduction Reaction Mechanism Research

The oxygen reduction reaction (ORR) serves as the fundamental cathodic process in clean energy technologies including fuel cells and metal-air batteries, yet its sluggish multi-step reaction kinetics significantly constrain energy conversion efficiency [115]. Platinum-based catalysts currently represent the most efficient ORR electrocatalysts, but their high cost and resource scarcity have redirected research focus toward enhancing atomic utilization efficiency [116]. Under practical ORR operating potentials and oxygen-rich environments, catalyst surfaces and bulk structures undergo dynamic reconstruction and oxidation rather than remaining static, making comprehension of this dynamic evolution process crucial for designing high-performance, stable catalysts [117]. Zheng et al. [118]. investigated dynamic active phases of catalysts under ORR conditions by introducing a topology-based persistent homology sampling algorithm, as illustrated in Figure 8a. This algorithm automatically identifies all potential adsorption/embedding sites within material structures by analyzing geometric topological features of atomic aggregates. Based on 100,000 initial PtO_x configurations generated through this algorithm for Pt_{55} clusters, efficient structural optimization and energy calculations were performed via MLFFs, achieving energy accuracy of 0.008 eV/atom and force accuracy of 0.21 eV/Å in the Pt-O system, as demonstrated in Figure 8b,c. Through systematic analysis of optimized configurations, the study successfully constructed a Pourbaix phase diagram for Pt clusters under varying potentials and pH conditions (Figure 8d), revealing that sub-nanometer Pt clusters undergo thermodynamically driven deep oxidation forming highly coordinated Pt-O structural units at typical ORR working potentials. This work demonstrates that combining topology-guided intelligent sampling with high-performance MLFFs enables automated discovery of catalyst active phases in complex environments, providing powerful computational tools for understanding and designing catalysts under realistic working conditions.

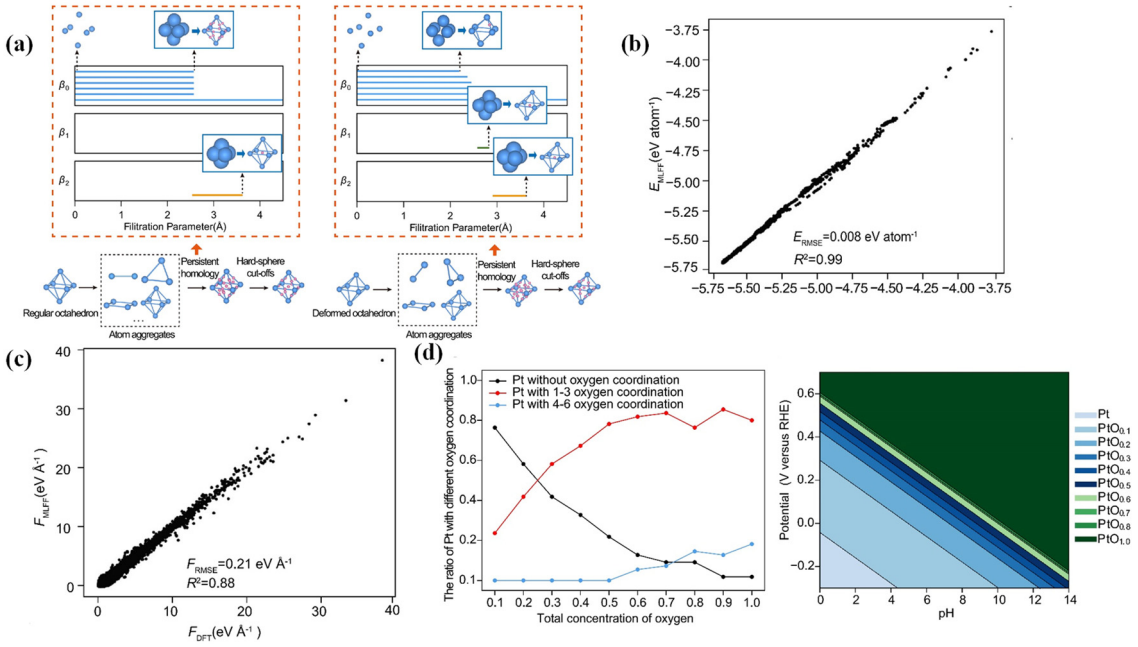

Figure 8. (**a**) Application of MLFFs combined with topology-guided sampling in ORR catalyst exploration. (**a**) Illustration of the persistent homology-based sampling algorithm (PH-SA) for automated identification of potential adsorption/embedding sites in material structures. (**b**,**c**) Comparison between MLFFs predictions and DFT calculations for (**b**) energies and (**c**) forces in the Pt–O system, demonstrating high accuracy (energy error: 0.008 eV/atom; force error: 0.21 eV/Å). (**d**) Analysis of oxygen coordination environments in PtO_x clusters: the proportion of Pt atoms with different oxygen coordination numbers as a function of total oxygen concentration, accompanied by the Pourbaix diagram illustrating thermodynamic phase stability under varying potential and pH conditions [118]. Copyright 2018 Springer Nature.

4.4. Towards Robust Workflows and Catalytic Challenges

The application of MLFFs in catalysis research has extended beyond the pursuit of high precision for individual systems, evolving along a logical trajectory that encompasses standardized evaluation, automated exploration, and realistic simulation. On one hand, benchmark datasets exemplified by the Open Catalyst Project (OC20/OC22) have established a standard for evaluating model performance on realistic surface adsorption configurations. This advancement has driven the development of specialized model architectures, such as

EquiformerV2, tailored for catalytic interfaces, highlighting the crucial role of high-order geometric feature learning in capturing complex surface interactions. On the other hand, active learning workflows integrated with uncertainty quantification have substantially enhanced the efficiency of exploring new materials and catalysts. This is achieved through a closed-loop iterative process comprising initial training, configuration exploration, uncertainty-based filtering, and DFT validation, thereby transforming MLFFs from static predictive tools into dynamic discovery platforms. Nevertheless, fully realizing the core value of MLFFs in catalytic design still faces several practical challenges: scaling simulations to realistic nano catalyst sizes and explicit solvent environments, accurately calculating free energy barriers by integrating enhanced sampling methods, and reliably correlating microscopic descriptors derived from MLFFs—such as adsorption energies and reaction barriers—with macroscopic catalytic performance metrics like activity and selectivity. In summary, MLFFs are transitioning from an advanced potential function technique into a comprehensive computational paradigm that includes data benchmarks, intelligent algorithms, and integrated workflows. The vision of rational catalyst design driven by MLFFs can only be achieved through synergistic progress in model architecture innovation, workflow automation, and the tackling of multiscale challenges.

5. Conclusions and Future Prospects

This review systematically summarizes the developmental trajectory of MLFFs and their significant applications in catalytic reaction studies. Traditional molecular simulation methods have long confronted the challenge of balancing accuracy with computational efficiency, where quantum chemical approaches like DFT provide high precision but require substantial computational resources, while classical molecular dynamics offers efficiency but inadequate description of chemical reaction processes. MLFFs address this gap by employing machine learning to extract interatomic interactions from quantum mechanical data, successfully constructing next-generation potential functions that combine quantum accuracy with classical efficiency. The evolutionary path of MLFFs demonstrates progression from simple descriptor-based models to sophisticated deep neural network architectures, substantially enhancing both accuracy and applicability domains. In catalytic reaction research, MLFFs have demonstrated crucial roles across various systems including HER, CO_2RR and ORR. By simulating adsorption, migration, and transformation processes of reactants on catalyst surfaces, MLFFs enable revelation of dynamic evolution patterns at active sites and elucidate selectivity in reaction pathways, thereby providing atomic-level theoretical guidance for understanding catalytic mechanisms and designing high-performance catalysts.

Despite remarkable progress, MLFFs still confront numerous challenges. The performance of MLFFs is fundamentally constrained by the quality and coverage of training data. Since MLFFs models are typically trained on DFT generated data, inherent systematic errors from DFT calculations inevitably propagate into MLFFs predictions. These include biases in exchange-correlation functional selection and inadequate description of van der Waals interactions, which become particularly pronounced when simulating surface adsorption processes or reaction energy barriers. Significant challenges remain in accurately describing long-range electrostatic and van der Waals interactions, especially in catalytic systems involving charge transfer or polar interfaces [119–124]. Furthermore, MLFFs based on ground-state electronic structure struggle to handle processes involving excited electronic states, such as photocatalysis or plasma-driven catalysis. Future MLFFs development should prioritize enhancing model interpretability, creating novel architectures that balance accuracy with efficiency to support extended spatiotemporal simulations, and advancing seamless integration with multiscale methodologies like continuum models. As the technology matures and application scope expands, MLFFs will undoubtedly play increasingly vital roles in catalysis science and broader materials research, providing powerful theoretical foundations and technical capabilities for addressing critical energy and environmental challenges.

Author Contributions

X.H.: Conceptualization, Investigation, Formal analysis, Data curation, Writing—Original Draft. G.W.: Conceptualization, Methodology, Writing—Review & Editing. C.W.: Writing—Review & Editing. B.S.: Supervision, Project administration, Methodology, Funding acquisition, Software, Writing—Review & Editing. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by the Advanced Materials-National Science and Technology Major Project (No. 2025ZD0618802), the National Natural Science Foundation of China (No. 52571005, 52401001), the Natural

Science Foundation of Fujian Province (No. 2025J02008, 2024J01262), and the Minjiang Scholar Program of Fujian Province.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

No new data were generated in this study.

Conflicts of Interest

The authors declare no conflict of interest.

Use of AI and AI-Assisted Technologies

No AI tools were utilized for this paper.

References

1. Centi, G.; Quadrelli, E.A.; Perathoner, S. Catalysis for CO_2 conversion: A key technology for rapid introduction of renewable energy in the value chain of chemical industries. *Energy Environ. Sci.* **2013**, *6*, 1711–1731. https://doi.org/10.1039/c3ee00056g.
2. Deng, R.J.; You, K.Y.; Zhao, F.F.; et al. Highly selective preparation of valuable dinitronaphthalene from catalytic nitration of 1-nitronaphthalene with NO_2 over HY zeolite. *Can. J. Chem. Eng.* **2018**, *96*, 2586–2592. https://doi.org/10.1002/cjce.23194.
3. Deng, Z.; Guo, Y.; Sun, Z.; et al. Electrocatalytic organic transformation reactions in green chemistry: Exploring nanocrystals and single atom catalysts. *Nano Res.* **2024**, *17*, 9326–9344. https://doi.org/10.1007/s12274-024-6887-8.
4. Yoon, Y.; You, H.M.; Oh, J.; et al. Recent advances in atomic-scale simulations for supported metal catalysts. *Mol. Catal.* **2024**, *554*, 113862. https://doi.org/10.1016/j.mcat.2024.113862.
5. Galwey, A.K. Is the science of thermal analysis kinetics based on solid foundations? A literature appraisal. *Thermochim. Acta* **2004**, *413*, 139–183. https://doi.org/10.1016/j.tca.2003.10.013.
6. Kiss, G.; Pande, V.S.; Houk, K.N. Molecular Dynamics Simulations for the Ranking, Evaluation, and Refinement of Computationally Designed Proteins. In *Methods in Protein Design*; Keating, A.E., Ed.; Academic Press: Cambridge, MA, USA, 2013; Volume 523, pp. 145–170.
7. Chen, H.Y. Recent advances in computational study and design of MOF catalysts for CO_2 conversion. *Front. Energy Res.* **2022**, *10*, 1016406. https://doi.org/10.3389/fenrg.2022.1016406.
8. Chen, X.Y.; Yang, X.Z. Mechanistic Insights and Computational Design of Transition-Metal Catalysts for Hydrogenation and Dehydrogenation Reactions. *Chem. Rec.* **2016**, *16*, 2364–2378. https://doi.org/10.1002/tcr.201600049.
9. Luo, Y.; Maeda, S.; Ohno, K. Water-catalyzed gas-phase reaction of formic acid with hydroxyl radical: A computational investigation. *Chem. Phys. Lett.* **2009**, *469*, 57–61. https://doi.org/10.1016/j.cplett.2008.12.087.
10. Li, X.; Zou, Y.; Zhao, Y.Q.; et al. Density functional theory studies of Pt-based catalysts for CO oxidation. *Sci. China-Technol. Sci.* **2025**, *68*, 1310501. https://doi.org/10.1007/s11431-024-2867-9.
11. Chizallet, C.; Raybaud, P. Density functional theory simulations of complex catalytic materials in reactive environments: Beyond the ideal surface at low coverage. *Catal. Sci. Technol.* **2014**, *4*, 2797–2813. https://doi.org/10.1039/c3cy00965c.
12. Wang, Y.G.; Mei, D.H.; Glezakou, V.A.; et al. Dynamic formation of single-atom catalytic active sites on ceria-supported gold nanoparticles. *Nat. Commun.* **2015**, *6*, 6511. https://doi.org/10.1038/ncomms7511.
13. Fu, D.L.; Guo, W.Y.; Li, M.; et al. Adsorption and reaction mechanisms of SO_2 on Rh(111) surface: A first-principle study. *J. Mol. Struct.* **2014**, *1062*, 68–76. https://doi.org/10.1016/j.molstruc.2014.01.035.
14. Fu, C.; Shi, Y.Y.; Li, K.; et al. CO_2 reduction catalysis of CO doped Cu nanoparticles anchored to graphene: A DFT study. *Comput. Theor. Chem.* **2025**, *1254*, 115489. https://doi.org/10.1016/j.comptc.2025.115489.
15. Dudnik, A.S.; Xia, Y.Z.; Li, Y.H.; et al. Computation-Guided Development of Au-Catalyzed Cycloisomerizations Proceeding via 1,2-Si or 1,2-H Migrations: Regiodivergent Synthesis of Silylfurans. *J. Am. Chem. Soc.* **2010**, *132*, 7645–7655. https://doi.org/10.1021/ja910290c.
16. Cornaton, Y.; Djukic, J.P. Noncovalent Interactions in Organometallic Chemistry: From Cohesion to Reactivity, a New Chapter. *Acc. Chem. Res.* **2021**, *54*, 3828–3840. https://doi.org/10.1021/acs.accounts.1c00393.
17. Matsumoto, K.; Misawa, N.; Kanesato, S.; et al. Atomistic simulation of olefin polymerization reaction by organometallic

catalyst: Significant role of microscopic structural dynamics of (pyridylamido) Hf(IV) complex in catalytic reactivity. *Front. Chem.* **2025**, *13*, 1618025. https://doi.org/10.3389/fchem.2025.1618025.

18. Steinmann, S.N.; Wang, Q.; Seh, Z.W. How machine learning can accelerate electrocatalysis discovery and optimization. *Mater. Horiz.* **2023**, *10*, 393–406. https://doi.org/10.1039/d2mh01279k.
19. Barford, W. Exciton dynamics in conjugated polymer systems. *Front. Phys.* **2022**, *10*. https://doi.org/10.3389/fphy.2022.1004042.
20. Gu, Y.; Wang, L.; Xu, B.Q.; et al. Recent advances in the molecular-level understanding of catalytic hydrogenation and oxidation reactions at metal-aqueous interfaces. *Chin. J. Catal.* **2023**, *54*, 1–55. https://doi.org/10.1016/S1872-2067(23)64550-4.
21. Zhou, Z.H.; Cui, H.Q.; Fan, H.J.; et al. A Reactive Explicit Electron Force Field for Hydrocarbons. *J. Chem. Theory Comput.* **2025**, *21*, 6988–7001. https://doi.org/10.1021/acs.jctc.5c00633.
22. Keith, J.A.; Vassilev-Galindo, V.; Cheng, B.Q.; et al. Combining Machine Learning and Computational Chemistry for Predictive Insights Into Chemical Systems. *Chem. Rev.* **2021**, *121*, 9816–9872. https://doi.org/10.1021/acs.chemrev.1c00107.
23. Nyangiwe, N.N. Applications of density functional theory and machine learning in nanomaterials: A review. *Next Mater.* **2025**, *8*, 100683. https://doi.org/10.1016/j.nxmate.2025.100683.
24. Yoon, U.; Jeong, K.; Kim, S.H. Advancing electrocatalysis through density functional theory: From reaction mechanisms to machine learning integration. *J. CO_2 Util.* **2025**, *101*, 103192. https://doi.org/10.1016/j.jcou.2025.103192.
25. Yu, Q.M.; Ma, N.G.; Leung, C.; et al. AI in single-atom catalysts: A review of design and applications. *J. Mater. Inform.* **2025**, *5*, 9. https://doi.org/10.20517/jmi.2024.78.
26. Deshmukh, G.; Ghanekar, P.; Greeley, J. Deep Learning for Computational Heterogeneous Catalysis: Fundamentals and Applications. *J. Indian Inst. Sci.* **2025**. https://doi.org/10.1007/s41745-025-00484-6.
27. Wayo, D.D.K.; Goliatt, L.; Ganji, M.D. DFT and hybrid classical-quantum machine learning integration for photocatalyst discovery and hydrogen production. *Rev. Chem. Eng.* **2025**, *41*, 741–774. https://doi.org/10.1515/revce-2025-0022.
28. Liu, M.F.; Wang, J.T.; Hu, J.W.; et al. Layer-by-layer phase transformation in Ti_3O_5 revealed by machine-learning molecular dynamics simulations. *Nat. Commun.* **2024**, *15*, 3079. https://doi.org/10.1038/s41467-024-47422-1.
29. Chen, M.Y.; Hu, J.W.; Yu, Y.C.; et al. Advances in Machine Learning Molecular Dynamics to Assist Materials Nucleation and Solidification Research. *Acta Metall. Sin.* **2024**, *60*, 1329–1344. https://doi.org/10.11900/0412.1961.2024.00192.
30. Kubo, M.; Ando, M.; Sakahara, S.; et al. Development of tight-binding, chemical-reaction-dynamics simulator for combinatorial computational chemistry. *Appl. Surf. Sci.* **2004**, *223*, 188–195. https://doi.org/10.1016/S0169-4332(03)00912-7.
31. Park, J.; Kim, H.S.; Lee, W.B.; et al. Trends and Outlook of Computational Chemistry and Microkinetic Modeling for Catalytic Synthesis of Methanol and DME. *Catalysts* **2020**, *10*, 655. https://doi.org/10.3390/catal10060655.
32. Chen, B.W.J.; Zhang, X.L.; Zhang, J. Accelerating explicit solvent models of heterogeneous catalysts with machine learning interatomic potentials. *Chem. Sci.* **2023**, *14*, 8338–8354. https://doi.org/10.1039/d3sc02482b.
33. Lele, A.D.; Shi, Z.Y.; Khetan, S.; et al. Machine-Learned Force Field for Molecular Dynamics Simulations of Nonequilibrium Ammonia Synthesis on Iron Catalysts. *J. Phys. Chem. C* **2025**, *129*, 4937–4949. https://doi.org/10.1021/acs.jpcc.4c07596.
34. Pisal, P.; Krejcí, O.; Rinke, P. Machine learning accelerated descriptor design for catalyst discovery in CO2 to methanol conversion. *NPJ Comput. Mater.* **2025**, *11*, 213. https://doi.org/10.1038/s41524-025-01664-9.
35. Schaaf, L.L.; Fako, E.; De, S.; et al. Accurate energy barriers for catalytic reaction pathways: An automatic training protocol for machine learning force fields. *NPJ Comput. Mater.* **2023**, *9*, 180. https://doi.org/10.1038/s41524-023-01124-2.
36. Wu, S.L.; Zheng, S.S.; Zhang, W.T.; et al. Machine-learning prediction of facet-dependent CO coverage on Cu electrocatalysts. *J. Mater. Inform.* **2025**, *5*, 14. https://doi.org/10.20517/jmi.2024.77.
37. Botu, V.; Batra, R.; Chapman, J.; et al. Machine Learning Force Fields: Construction, Validation, and Outlook. *J. Phys. Chem. C* **2017**, *121*, 511–522. https://doi.org/10.1021/acs.jpcc.6b10908.
38. Zhang, J.K.; Li, Y.P.; Chen, M.; et al. Atomistic simulation of batteries via machine learning force fields: From bulk to interface. *J. Energy Chem.* **2025**, *106*, 911–929. https://doi.org/10.1016/j.jechem.2025.02.051.
39. Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401. https://doi.org/10.1103/PhysRevLett.98.146401.
40. Alder, B.J.; Wainwright, T.E. Studies in Molecular Dynamics. I. General Method. *J. Chem. Phys.* **1959**, *31*, 459–466. https://doi.org/10.1063/1.1730376.
41. Peslherbe, G.H.; Hase, W.L. Semiempirical MNDO, AM1, and PM3 direct dynamics trajectory studies of formaldehyde unimolecular dissociation. *J. Chem. Phys.* **1996**, *104*, 7882–7894. https://doi.org/10.1063/1.471504.
42. Omranpour, A.; Elsner, J.; Lausch, K.N.; et al. Machine Learning Potentials for Heterogeneous Catalysis. *ACS Catal.* **2025**, *15*, 1616–1634. https://doi.org/10.1021/acscatal.4c06717.
43. Payne, M.C.; Teter, M.P.; Allan, D.C.; et al. Iterative minimization techniques for ab initio total-energy calculations: Molecular dynamics and conjugate gradients. *Rev. Mod. Phys.* **1992**, *64*, 1045–1097. https://doi.org/10.1103/RevModPhys.64.1045.
44. Car, R.; Parrinello, M. Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.* **1985**, *55*, 2471–2474. https://doi.org/10.1103/PhysRevLett.55.2471.

45. Remler, D.K.; Madden, P.A. Molecular dynamics without effective potentials via the Car-Parrinello approach. *Mol. Phys.* **1990**, *70*, 921–966. https://doi.org/10.1080/00268979000101451.
46. Thiel, W. Semiempirical quantum–chemical methods. *WIREs Comput. Mol. Sci.* **2014**, *4*, 145–157. https://doi.org/10.1002/wcms.1161.
47. Dewar, M.J.S.; Zoebisch, E.G.; Healy, E.F.; et al. Development and use of quantum mechanical molecular models. 76. AM1: A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909. https://doi.org/10.1021/ja00299a024.
48. Stewart, J.J.P. Optimization of parameters for semiempirical methods I. Method. *J. Comput. Chem.* **1989**, *10*, 209–220. https://doi.org/10.1002/jcc.540100208.
49. Elstner, M.; Porezag, D.; Jungnickel, G.; et al. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B* **1998**, *58*, 7260–7268. https://doi.org/10.1103/PhysRevB.58.7260.
50. Cornell, W.D.; Cieplak, P.; Bayly, C.I.; et al. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197. https://doi.org/10.1021/ja955032e.
51. Chenoweth, K.; van Duin, A.C.T.; Goddard, W.A. ReaxFF Reactive Force Field for Molecular Dynamics Simulations of Hydrocarbon Oxidation. *J. Phys. Chem. A* **2008**, *112*, 1040–1053. https://doi.org/10.1021/jp709896w.
52. Halgren, T.A.; Damm, W. Polarizable force fields. *Curr. Opin. Struct. Biol.* **2001**, *11*, 236–242. https://doi.org/10.1016/S0959-440X(00)00196-2.
53. Ponder, J.W.; Case, D.A. Force Fields for Protein Simulations. In *Advances in Protein Chemistry*; Academic Press: Cambridge, MA, USA, 2003; Volume 66, pp. 27–85.
54. Behler, J. Constructing high-dimensional neural network potentials: A tutorial review. *Int. J. Quantum Chem.* **2015**, *115*, 1032–1050. https://doi.org/10.1002/qua.24890.
55. Bartók, A.P.; Payne, M.C.; Kondor, R.; et al. Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons. *Phys. Rev. Lett.* **2010**, *104*, 136403. https://doi.org/10.1103/PhysRevLett.104.136403.
56. Xia, J.; Yaolong, Z.; Jiang, B. The evolution of machine learning potentials for molecules, reactions and materials. *Chem. Soc. Rev.* **2025**, *54*, 4790–4821. https://doi.org/10.1039/d5cs00104h.
57. Schütt, K.T.; Sauceda, H.E.; Kindermans, P.J.; et al. SchNet—A deep learning architecture for molecules and materials. *J. Chem. Phys.* **2018**, *148*, 241722. https://doi.org/10.1063/1.5019779.
58. Dai, M.; Demirel, M.F.; Liang, Y.; et al. Graph neural networks for an accurate and interpretable prediction of the properties of polycrystalline materials. *NPJ Comput. Mater.* **2021**, *7*, 103. https://doi.org/10.1038/s41524-021-00574-w.
59. Frank, J.T.; Unke, O.; Müller, K.-R. So3krates: Equivariant attention for interactions on arbitrary length-scales in molecular systems. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 29400–29413.
60. Jain, A.; Ong, S.P.; Hautier, G.; et al. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **2013**, *1*, 011002. https://doi.org/10.1063/1.4812323.
61. Unke, O.T.; Chmiela, S.; Sauceda, H.E.; et al. Machine Learning Force Fields. *Chem. Rev.* **2021**, *121*, 10142–10186. https://doi.org/10.1021/acs.chemrev.0c01111.
62. Behler, J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J. Chem. Phys.* **2011**, *134*, 074106. https://doi.org/10.1063/1.3553717.
63. Bartók, A.P.; Kondor, R.; Csányi, G. On representing chemical environments. *Phys. Rev. B* **2013**, *87*, 184115. https://doi.org/10.1103/PhysRevB.87.184115.
64. Deringer, V.L.; Csányi, G. Machine learning based interatomic potential for amorphous carbon. *Phys. Rev. B* **2017**, *95*, 094203. https://doi.org/10.1103/PhysRevB.95.094203.
65. Vandermause, J.; Torrisi, S.B.; Batzner, S.; et al. On-the-fly active learning of interpretable Bayesian force fields for atomistic rare events. *NPJ Comput. Mater.* **2020**, *6*, 20. https://doi.org/10.1038/s41524-020-0283-z.
66. Schütt, K.T.; Kessel, P.; Gastegger, M.; et al. SchNetPack: A Deep Learning Toolbox For Atomistic Systems. *J. Chem. Theory Comput.* **2019**, *15*, 448–455. https://doi.org/10.1021/acs.jctc.8b00908.
67. Thomas, N.; Smidt, T.E.; Kearnes, S.M.; et al. Tensor Field Networks: Rotation- and Translation-Equivariant Neural Networks for 3D Point Clouds. *arXiv* **2018**, arXiv:1802.08219.
68. Musil, F.; Grisafi, A.; Bartók, A.P.; et al. Physics-Inspired Structural Representations for Molecules and Materials. *Chem. Rev.* **2021**, *121*, 9759–9815. https://doi.org/10.1021/acs.chemrev.1c00021.
69. Musaelian, A.; Batzner, S.; Johansson, A.; et al. Learning local equivariant representations for large-scale atomistic dynamics. *Nat. Commun.* **2023**, *14*, 579. https://doi.org/10.1038/s41467-023-36329-y.
70. Batatia, I.; Batzner, S.; Kovács, D.P.; et al. The design space of E(3)-equivariant atom-centred interatomic potentials. *Nat. Mach. Intell.* **2025**, *7*, 56–67. https://doi.org/10.1038/s42256-024-00956-x.
71. Nomura, K.i.; Hattori, S.; Ohmura, S.; et al. Allegro-FM: Toward an Equivariant Foundation Model for Exascale Molecular Dynamics Simulations. *J. Phys. Chem. Lett.* **2025**, *16*, 6637–6644.
72. Liao, Y.; Smidt, T.E.J.A. Equiformer: Equivariant Graph Attention Transformer for 3D Atomistic Graphs. *arXiv* **2022**,

arXiv:2206.11990.
73. Zitnick, C.L.; Das, A.; Kolluru, A.; et al. Spherical Channels for Modeling Atomic Interactions. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 8054–8067.
74. Vaswani, A.; Shazeer, N.; Parmar, N.; et al. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 6000–6010.
75. Schütt, K.T.; Unke, O.T.; Gastegger, M. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In Proceedings of the International Conference on Machine Learning, Virtual, 18–24 July 2021.
76. Kovács, D.P.; Moore, J.H.; Browning, N.J.; et al. MACE-OFF: Short-Range Transferable Machine Learning Force Fields for Organic Molecules. *J. Am. Chem. Soc.* **2025**, *147*, 17598–17611. https://doi.org/10.1021/jacs.4c07099.
77. Batatia, I.; Benner, P.; Chiang, Y.; et al. A foundation model for atomistic materials chemistry. *J. Chem. Phys.* **2025**, *163*, 184110. https://doi.org/10.1063/5.0297006.
78. Chen, C.; Ong, S.P. A universal graph deep learning interatomic potential for the periodic table. *Nat. Comput. Sci.* **2022**, *2*, 718–728. https://doi.org/10.1038/s43588-022-00349-3.
79. Liao, Y.-L.; Wood, B.; Das, A.; et al. EquiformerV2: Improved Equivariant Transformer for Scaling to Higher-Degree Representations. *arXiv* **2023**, arXiv:2306.12059.
80. Tran, R.; Lan, J.; Shuaibi, M.; et al. The Open Catalyst 2022 (OC22) Dataset and Challenges for Oxide Electrocatalysts. *ACS Catal.* **2023**, *13*, 3066–3084. https://doi.org/10.1021/acscatal.2c05426.
81. Dunn, A.; Wang, Q.; Ganose, A.; et al. Benchmarking materials property prediction methods: The Matbench test set and Automatminer reference algorithm. *NPJ Comput. Mater.* **2020**, *6*, 138. https://doi.org/10.1038/s41524-020-00406-3.
82. Wang, H.; Zhang, L.; Han, J.; et al. DeePMD-kit: A deep learning package for many-body potential energy representation and molecular dynamics. *Comput. Phys. Commun.* **2018**, *228*, 178–184. https://doi.org/10.1016/j.cpc.2018.03.016.
83. Gao, X.; Ramezanghorbani, F.; Isayev, O.; et al. TorchANI: A Free and Open Source PyTorch-Based Deep Learning Implementation of the ANI Neural Network Potentials. *J. Chem. Inf. Model* **2020**, *60*, 3408–3415. https://doi.org/10.1021/acs.jcim.0c00451.
84. Batatia, I.; Kovacs, D.; Simm, G.; et al. MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 11423–11436.
85. Vandermause, J.; Xie, Y.; Lim, J.S.; et al. Active learning of reactive Bayesian force fields applied to heterogeneous catalysis dynamics of H/Pt. *Nat. Commun.* **2022**, *13*, 5183. https://doi.org/10.1038/s41467-022-32294-0.
86. Zhang, L.; Lin, D.-Y.; Wang, H.; et al. Active learning of uniformly accurate interatomic potentials for materials simulation. *Phys. Rev. Mater.* **2019**, *3*, 023804. https://doi.org/10.1103/PhysRevMaterials.3.023804.
87. Chmiela, S.; Tkatchenko, A.; Sauceda, H.E.; et al. Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv.* **2017**, *3*, e1603015. https://doi.org/10.1126/sciadv.1603015.
88. Chmiela, S.; Sauceda, H.E.; Poltavsky, I.; et al. sGDML: Constructing accurate and data efficient molecular force fields using machine learning. *Comput. Phys. Commun.* **2019**, *240*, 38–45. https://doi.org/10.1016/j.cpc.2019.02.007.
89. Kapil, V.; Rossi, M.; Marsalek, O.; et al. i-PI 2.0: A universal force engine for advanced molecular simulations. *Comput. Phys. Commun.* **2019**, *236*, 214–223. https://doi.org/10.1016/j.cpc.2018.09.020.
90. Hjorth Larsen, A.; Jørgen Mortensen, J.; Blomqvist, J.; et al. The atomic simulation environment-a Python library for working with atoms. *J. Phys. Condens. Matter Inst. Phys. J.* **2017**, *29*, 273002. https://doi.org/10.1088/1361-648X/aa680e.
91. Deelman, E.; Gannon, D.; Shields, M.; et al. Workflows and e-Science: An overview of workflow system features and capabilities. *Future Gener. Comput. Syst.* **2009**, *25*, 528–540. https://doi.org/10.1016/j.future.2008.06.012.
92. Ong, S.P.; Richards, W.D.; Jain, A.; et al. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **2013**, *68*, 314–319. https://doi.org/10.1016/j.commatsci.2012.10.028.
93. Chanussot, L.; Das, A.; Goyal, S.; et al. Open Catalyst 2020 (OC20) Dataset and Community Challenges. *ACS Catal.* **2021**, *11*, 6059–6072. https://doi.org/10.1021/acscatal.0c04525.
94. 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. https://doi.org/10.1038/s41929-018-0142-1.
95. Choung, S.; Park, W.; Moon, J.; et al. Rise of machine learning potentials in heterogeneous catalysis: Developments, applications, and prospects. *Chem. Eng. J.* **2024**, *494*, 152757. https://doi.org/10.1016/j.cej.2024.152757.
96. Agarwal, T.; Kumar, N.; Ritu; et al. Homogeneous HER electrocatalysis using monothiolate ligand-based {FeS} complexes: A review. *Electrochim. Acta* **2024**, *507*, 144852. https://doi.org/10.1016/j.electacta.2024.144852.
97. Lin, Y.W.; Li, Y.S.; Chang, C.W.; et al. Kinetic Analysis of Oxygen Evolution on Spin-Coated Thin-Film Electrodes via Electrochemical Impedance Spectroscopy. *Coatings* **2023**, *13*, 1957. https://doi.org/10.3390/coatings13111957.
98. He, Y.Z.; Liu, S.S.; Wang, M.F.; et al. Advancing the Electrochemistry of Gas-Involved Reactions through Theoretical Calculations and Simulations from Microscopic to Macroscopic. *Adv. Funct. Mater.* **2022**, *32*, 2208474. https://doi.org/10.1002/adfm.202208474.
99. Chen, M.P.; Smart, T.J.J.; Wang, S.W.; et al. The coupling of experiments with density functional theory in the studies of the

electrochemical hydrogen evolution reaction. *J. Mater. Chem. A* **2020**, *8*, 8783–8812. https://doi.org/10.1039/d0ta02549f.
100. Tu, Y.C.; Deng, J.; Ma, C.; et al. Double-layer hybrid chainmail catalyst for high-performance hydrogen evolution. *Nano Energy* **2020**, *72*, 104700. https://doi.org/10.1016/j.nanoen.2020.104700.
101. Rice, P.S.; Liu, Z.P.; Hu, P. Hydrogen Coupling on Platinum Using Artificial Neural Network Potentials and DFT. *J. Phys. Chem. Lett.* **2021**, *12*, 10637–10645. https://doi.org/10.1021/acs.jpclett.1c02998.
102. Zhao, X.; Xiao, S.J.; Yao, B.M.; et al. DFT-Based Mechanistic Exploration and Application in Photocatalytic Heterojunctions. *J. Chem. Theory Comput.* **2024**, *20*, 9770–9786. https://doi.org/10.1021/acs.jctc.4c01051.
103. Sharma, V.; Roondhe, B.; Saxena, S.; et al. Role of functionalized graphene quantum dots in hydrogen evolution reaction: A density functional theory study. *Int. J. Hydrogen Energy* **2022**, *47*, 41748–41758. https://doi.org/10.1016/j.ijhydene.2022.02.161.
104. Zhang, X.X.; Wang, L.; Xie, Y.; et al. Theoretical insights into performance descriptors and their impact on activity optimization strategies for cobalt-based electrocatalysts. *Coord. Chem. Rev.* **2025**, *533*, 216560. https://doi.org/10.1016/j.ccr.2025.216560.
105. Özönder, S.; Küçükkartal, H.K. Rapid Discovery of Graphene Nanoflakes with Desired Absorption Spectra Using DFT and Bayesian Optimization with Neural Network Kernel. *J. Phys. Chem. A* **2025**, *129*, 4591–4600. https://doi.org/10.1021/acs.jpca.5c00405.
106. Chakraborty, T.; Chattaraj, P.K. Density functional theory for exploration of chemical reactivity: Successes and limitations. *J. Phys. Org. Chem.* **2023**, *36*, e4589. https://doi.org/10.1002/poc.4589.
107. Aidhy, D.S.; Zhang, Y.S.; Wolverton, C. Prediction of a $Ca(BH_4)(NH_2)$ quaternary hydrogen storage compound from first-principles calculations. *Phys. Rev. B* **2011**, *84*, 134103. https://doi.org/10.1103/PhysRevB.84.134103.
108. Morawietz, T.; Singraber, A.; Dellago, C.; et al. How van der Waals interactions determine the unique properties of water. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 8368–8373. https://doi.org/10.1073/pnas.1602375113.
109. Abbas, H.G.; Hahn, J.R. Machine Learning-Predicted Ternary Molybdenum Chalcogenophosphides for High-Efficiency Hydrogen Evolution Catalysis. *J. Phys. Chem. C* **2025**, *129*, 322–331. https://doi.org/10.1021/acs.jpcc.4c06879.
110. Hu, X.; Li, Q.; Li, W.; et al. Spin-polarized binuclear transition metal doping on g-C_3N_4 for photocatalytic CO_2 reduction to C2 products: A DFT study. *Mol. Catal.* **2025**, *572*, 114798. https://doi.org/10.1016/j.mcat.2024.114798.
111. Liu, R.; Li, X.; Lam, K.S. Combinatorial chemistry in drug discovery. *Curr. Opin. Chem. Biol.* **2017**, *38*, 117–126. https://doi.org/10.1016/j.cbpa.2017.03.017.
112. Karthikeyan, A.; Priyakumar, U.D. Artificial intelligence: Machine learning for chemical sciences. *J. Chem. Sci.* **2022**, *134*, 2. https://doi.org/10.1007/s12039-021-01995-2.
113. Jorner, K. Putting Chemical Knowledge to Work in Machine Learning for Reactivity. *Chimia* **2023**, *77*, 22–30. https://doi.org/10.2533/chimia.2023.22.
114. Selvaratnam, B.; Koodali, R.T. Machine learning in experimental materials chemistry. *Catal. Today* **2021**, *371*, 77–84. https://doi.org/10.1016/j.cattod.2020.07.074.
115. Debe, M.K. Electrocatalyst approaches and challenges for automotive fuel cells. *Nature* **2012**, *486*, 43–51. https://doi.org/10.1038/nature11115.
116. Xue, D.; Yuan, Y.; Yu, Y.; et al. Spin occupancy regulation of the Pt d-orbital for a robust low-Pt catalyst towards oxygen reduction. *Nat. Commun.* **2024**, *15*, 5990. https://doi.org/10.1038/s41467-024-50332-x.
117. Jeong, C.; Lee, J.; Jo, H.; et al. Atomic-scale 3D structural dynamics and functional degradation of Pt alloy nanocatalysts during the oxygen reduction reaction. *Nat. Commun.* **2025**, *16*, 8026. https://doi.org/10.1038/s41467-025-63448-5.
118. Zheng, S.; Zhang, X.-M.; Liu, H.-S.; et al. Active phase discovery in heterogeneous catalysis via topology-guided sampling and machine learning. *Nat. Commun.* **2025**, *16*, 2542. https://doi.org/10.1038/s41467-025-57824-4.
119. Smith, J.S.; Nebgen, B.; Lubbers, N.; et al. Less is more: Sampling chemical space with active learning. *J. Chem. Phys.* **2018**, *148*, 241733. https://doi.org/10.1063/1.5023802.
120. Korolev, V.V.; Mitrofanov, A.; Marchenko, E.I.; et al. Transferable and Extensible Machine Learning-Derived Atomic Charges for Modeling Hybrid Nanoporous Materials. *Chem. Mater.* **2020**, *32*, 7822–7831. https://doi.org/10.1021/acs.chemmater.0c02468.
121. Chmiela, S.; Vassilev-Galindo, V.; Unke, O.T.; et al. Accurate global machine learning force fields for molecules with hundreds of atoms. *Sci. Adv.* **2023**, *9*, eadf0873. https://doi.org/10.1126/sciadv.adf0873.
122. Miller, R.E.; Tadmor, E.B. A unified framework and performance benchmark of fourteen multiscale atomistic/continuum coupling methods. *Model. Simul. Mater. Sci. Eng.* **2009**, *17*, 053001. https://doi.org/10.1088/0965-0393/17/5/053001.
123. Grisafi, A.; Ceriotti, M. Incorporating long-range physics in atomic-scale machine learning. *J. Chem. Phys.* **2019**, *151*, 204105. https://doi.org/10.1063/1.5128375.
124. Welborn, M.; Cheng, L.; Miller, T.F., III. Transferability in Machine Learning for Electronic Structure via the Molecular Orbital Basis. *J. Chem. Theory Comput.* **2018**, *14*, 4772–4779. https://doi.org/10.1021/acs.jctc.8b00636.