Physics-informed machine learning for electrocatalyst development

Hao Wu[1], Mingxuan Chen[2], Hao Cheng[1], Tong Yang[1], Minggang Zeng[3], and Ming Yang[1,4,5,*]

[1] Department of Applied Physics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong SAR, China

2 School of Mathematics, The University of Edinburgh, Scotland, United Kingdom

[3] Institute of High-Performance Computing, A*STAR, 1 Fusionopolis Way, Singapore 138632, Singapore

[4] Research Centre on Data Sciences & Artificial Intelligence, The Hong Kong Polytechnic University, Hung Hom, Hong Kong SAR, China

[5] Research Centre for Nanoscience and Nanotechnology, The Hong Kong Polytechnic University, Hung Hom, Hong Kong SAR, China

To whom correspondence should be addressed: kevin.m.yang@polyu.edu.hk (M. Y.)

Abstract

Identifying exceptional electrocatalysts from the vast materials space remains a formidable challenge. Machine learning (ML) has emerged as a powerful tool to address this challenge, offering high efficiency while maintaining good accuracy in predictions. From this perspective, we provide a brief overview of recent advancements in machine learning for electrocatalyst discoveries. We emphasize the applications of physics-informed machine learning (PIML) models and explainable artificial intelligence (XAI) to electrocatalyst development, through which valuable physical/chemical insights can be distilled. Additionally, we delve into the challenges faced by PIML approaches, explore future directions, and discuss potential breakthroughs that could revolutionize the field of electrocatalyst development.

KEYWORDS: electrocatalysts, machine learning, domain knowledge-informed machine learning, explainable artificial intelligence

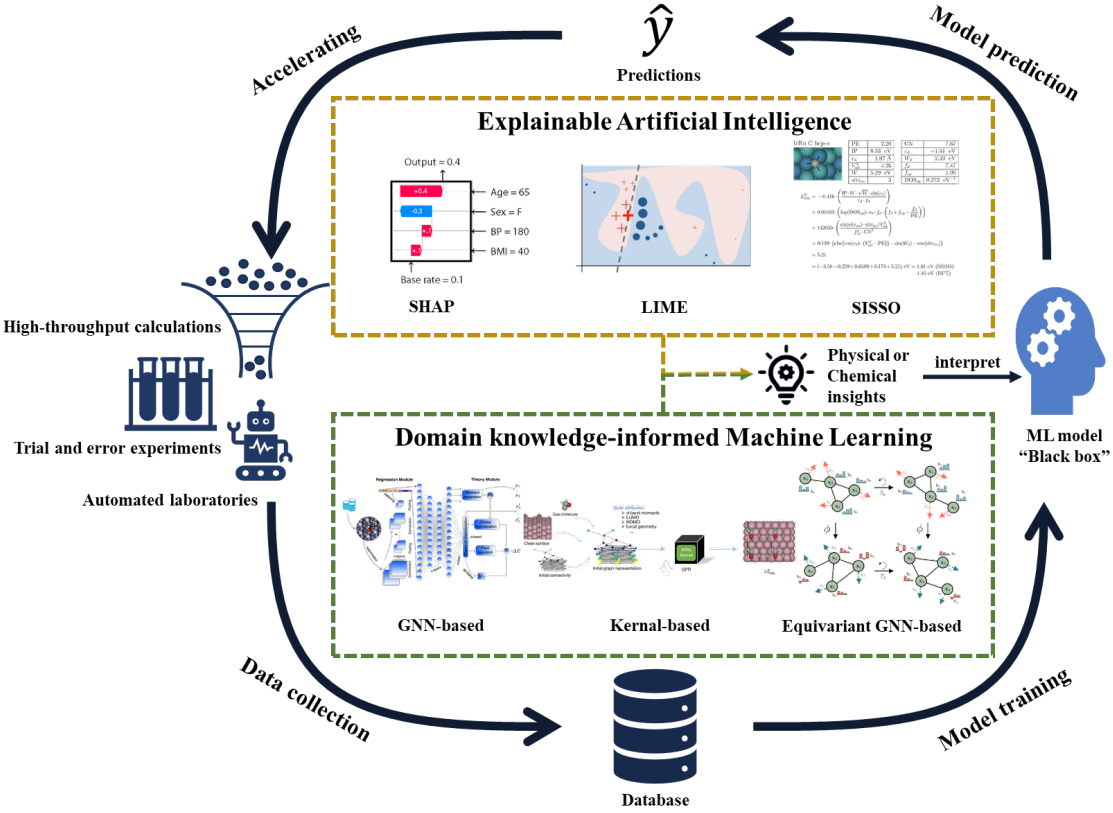

Figure 1. Schematic diagram of the ML accelerated electrocatalyst development using XAI and PIML.

1. Introduction

Excessive fossil fuel combustion since the industrial revolution has led to serious environmental problems such as global warming, extreme weathers and environment pollution. In response, there is a concerted effort to develop green energy solutions to reduce our dependence on fossil fuels. Electrochemical reactions, including but not limited to hydrogen evolution reaction (HER), oxygen evolution reaction (OER), oxygen reduction reaction (ORR), carbon dioxide reduction reaction (CO2RR), and nitrogen reduction reaction (NRR), hold significant promise for storing and utilizing essential intermittent renewable energies in the near future[1, 2]. Electrocatalysts are essential in increasing the reaction rate of the electrochemical reactions, thereby improving the conversion and utilization efficiency of renewable energies. However, the commonly used electrocatalysts are usually based on precious metals such as platinum, which are expensive and scarce. In addition, the existing electrocatalysts also suffer issues such as limited efficiency, durability, and scalability[3]. Therefore, discovering electrocatalysts that can solve above issues is fundamental to the development of high-performance electrochemical energy conversion/storage devices such as fuel cells, batteries, and supercapacitors[4].

Over the past few decades, electrocatalyst discovery has been accelerated by the rapid advancement of theoretical approaches and experimental capabilities[5]. For instance, DFT calculations have been widely used to predict various criticle properties of electrocatalysts, such as formation energies, adsorption energy and *d*-band center[6]. Based on thermodynamic and kinetic calculations, the free energy diagram and micro-kinetic reaction can be constructed, which are essential to identifying catalysts with improved stability/durability, activity and selectivity[7, 8]. Furthermore, high-throughput calculations have been applied to navigate vast materials and configurations. Semi-automated experiments have also been developed to accelerate the development of electrocatalysts. These high-throughput experimental and computational methods only reduce the development time and cost, but also provide large dataset for ML models. However, these high-throughput approaches still require high cost and their efficiency awaits further improvement. In this regard, ML would be a promising alternative for electrocatalyst discovery[9-12]. This is because ML can effectively explore vast material spaces, with the efficiency far beyond both density functional theory (DFT) calculations and traditional trial-and-error experiments. By leveraging large computational or experimental datasets, ML have been used to predict the aforementioned properties of electrocatalysts (Figure 1)[13-18], in which noticeable processes have been made. Despite the success, most models require large and complex neural network, together with their black-box nature, leading to high computing cost and lack physical/chemical interpretability. As a result, intensive efforts have been spared to incorporate physical/chemical insight into ML models, which is known as DIML models and XAI methods (Figure 1). It is noted that remarkable progresses have been made in this area over the past years, necessitating a comprehensive review to summarize their findings.

In this review, we summarize recent advancement of ML for electrocatalysts and focus on the DIML models. We begin with recent progress in machine learning for electrocatalysts, providing an overview of fundamental concepts in this field. We then discuss the necessity and benefits of integrating DIML models, highlighting their recent applications and contributions to advancing electrocatalysis development. Finally, we provide a brief discussion about challenges and perspectives on future directions and potential breakthroughs in DIML for electrocatalysts.

2. Progress of ML for electrocatalysts

~~ML, a branch of artificial intelligence, aims to study the hidden relations from data and generalize it to new data. Based on prediction types, ML tasks can be categorized as classification, regression, clustering etc. Additionally, these tasks can also be classified into supervised or unsupervised learning, depending on whether the training data are labeled.~~ Over the past decade, the high-throughput first-principles calculations have been intensively deployed and many large and open-source materials datasets have therefore been built, including Open Catalyst 2020 (OC20)[19], Open Catalyst 2022 (OC22)[20], Materials Project (MP)[21], the Open Quantum Materials Database (OQMD)[22], Two-Dimensional (2D) Materials Encyclopedia (2DMatPedia)[23] etc. These high-quality datasets are ideal start points to train high-fidelity ML models. Geometric matrices including Coulomb matrix[24, 25], the Smooth Overlap of Atomic Positions (SOAP)[26-29] play an important role in capturing global/local geometrics for most material systems. Moreover, descriptors such as electronegativity, d-band features, covalent radius and fingerprints[30] are vital to electrochemical reactions. These geometric matrices and descriptors are the starting point to construct the ML models.

This section will introduce three domains: machine learning, deep learning, and natural language processing. For each domain, we will start with the foundational theories of the models and then explore their recent applications in electrocatalysis.

2.1. Machine learning

Linear regression (LR) is the simplest ML model assuming a linear relationship between the output ($\hat{y}$) and the input descriptors ($X_1, X_2 \cdots X_n$), which can be written as:

$$\hat{y} = w_0 + w_1 X_1 + \cdots + w_n X_n + \epsilon \quad (1)$$

where $\hat{y}$ is the predicted value of the model, w_0 is the bias, w_n is the regression coefficient of the *n-th* independent variable ($n \in (1,2, \ldots n)$), and ϵ is the error for this model.

LR has been used for simple electrocatalyst applications such as predicting *OH and *O adsorption energy of high-entropy alloy (HEA) IrPdPtRhRu, which suggests a new HEA composition with better performance than pure Pt(111)[31]. In addition, LR has been deployed to discover the active sites in nonplatinum porphyrin-based electrocatalysts for ORR, ~~characterized by X-ray photoelectron spectroscopy (XPS) and rotating disk electrode[32]~~.

Support Vector Machine (SVM), ~~introduced in 1995,~~ is another widely used machine learning model[33]. It was initially designed to find a hyperplane to separate samples from different groups, which is deemed to be a solution to a classification task. By employing kernel functions, SVM can map samples from a low dimension to a higher one, allowing it to handle nonlinear relationships. Moreover, support vector regression (SVR) is capable of handling regression tasks by introducing tolerance margin ϵ into the SVM. Both SVM and SVR have wide applications in the electrochemical domain[34-38]. For example, Mohsen et al. used SVR to predict the Gibbs free energies of various reaction intermediates on single-atom catalysts (SACs) supported by graphene and porphyrin. Based on the trained SVR model, they reported that the most crucial factors in this system are the number of pyridinic nitrogen, the number of *d* electron, and the number of valence electron of the reaction intermediate[37].

Random forest (RF), regarded as one of the most successful ensemble models of ML, employs the Bagging algorithm. It uses decision trees as its base learner while introducing the random features selection method in the training process. RF also has widespread usage for predicting the properties of electrocatalysts[39-46]. For example, four ML models, including SVR, RF, Xtreme Gradient Boosting Regression (XGBR), and Artificial Neural Network (ANN), were tested to predict the Gibbs free energy change of hydrogen adsorption on double-atom catalysts, and RF outperformed the other three ML methods owing to its ensemble nature[42]. Another useful ensemble model is Gradient Boosting Regression (GBR), which also shows improved performance in various cases such as predicting free energy of N_2 electroreduction reaction free energy and selecting MXenes as hydrogen evolution reaction catalysts[40, 41, 45, 47, 48].

2.2. Deep learning

Apart from the aforementioned ML models, deep learning (DL) has attracted increasing interest for catalyst material development. DL includes convolutional neural networks (CNNs) and graph neural networks (GNNs). CNNs can deal with data in Euclidean space, such as images, videos, etc, and show promising performance for electrocatalyst materials. For example, Yang et al. employed a CNN model to predict the CO2RR adsorption energies on 2D SACs with 2D electronic density of states (DOS) and achieved an averaged mean absolute error (MAE) of 0.06 eV across various adsorbates such as CO2, COOH, CO, and CHO, CH2O, OCH3, O, OH, and H. Combining the CNN model with volcano plot into the analysis of catalysis performance, the framework is useful for designing SACs as the electrocatalyst for CO2RR[50]. On the other hand, GNNs are designed to solve problems of non-Euclidean data including social networks, knowledge graphs, and molecules/materials. Given the node, edge and global attributes, the information is transformed using "message passing" algorithms, which can be written as:

$$m_v^{t+1} = \sum_{w \in N(v)} M_t(h_v^t, h_w^t, e_{vw}) \quad (2)$$

$$h_v^{t+1} = U_t(h_v^t, m_v^{t+1}), \quad (3)$$

where m is the message, h and e are the embeddings of nodes and edges, respectively, M_t denotes the message update function, and U_t denotes the vertex update functions. Afterward, the readout layer calculates node, edge and global embeddings. Finally, the results can be predicted by adding fully connected layers to the embeddings corresponding to the task [51]. GNNs usually perform better than conventional ML models without requirement of dedicated feature designs from experts. Specifically, simple and accessible features such as electronegativity, covalent radius, group number, etc.[52] can be directly used as the input of GNNs. Recent applications in machine learning interatomic potentials (MLIPs) have illustrated that GNNs have better generalization and higher accuracy than other models[53-57]. Batatia et al. proposed a new equivariant message passing graph neural network (MPNN) model called MACE[58, 59], one of the most accurate MLIPs, as evidenced by F1 score of 0.669 eV/atom, coefficient of determination (R^2) of 0.697 eV/atom, and MAE of 0.057 eV/atom[60]. MACE can solve complex scenarios for heterogeneous catalysis like Pourbaix diagrams, linear scaling relationship (LSR), CO (electro-)oxidation on Cu, and Carborane rearrangement of heterogeneous catalysis[61].

2.3. Natural language processing (NLP)

The scientific field has been dramatically influenced since the advent of ChatGPT[62]. Scientists have been increasingly using ChatGPT to gain insights and deepen their understanding of specific areas. Large language models (LLMs) are currently the most advanced architectures used in NLP and have received much attention. LLMs often build on transformer architecture designed to effectively process sequential data such as text[63, 64]. The critical component of transformer architecture is the attention mechanism, which focuses on relevant parts of the dataset to enhance performance. However, LLMs can sometimes produce misleading results in professional contexts due to the immense diversity of large datasets[65]. Fortunately, transfer learning algorithms can be utilized to adapt the universal LLMs to specialized LLMs[66]. Beltagy et al. developed the pretrained BERT model, SciBERT, which can automatically extract scientific knowledge from existing papers[67]. Although the application of LLMs in electrocatalysis is still in the early stage, several efforts have been made to show their promising applications such as CataLM[68] and InCrEDible-MaT-GO [69].

3. DIML for **electrocatalysts**

ML has achieved great success in catalysis research. However, these predictions made by ML models are based on previous experiences rather than understanding the mechanism behind them. While those ML models demonstrate high accuracy in interpolation, they fall short in inspiring chemists to creatively discover unseen electrocatalysts, as they are not effective for extrapolation. One solution to this limitation is to develop DIML for electrocatalysts with better generalization and interpretability, and the models will naturally give physical/chemical explanations for their predictions. The other solution is the Explainable AI (XAI) approach. Although the ultimate goal of these two

methods is the same, there are several differences between DIML and XAI. DIML aims to integrate physical/chemical laws and principles into AI models to build an interpretable and robust model, while XAI aims to explain the model's behavior[70]. Importantly, DIML and XAI are complementary: DIML can leverage XAI techniques to better illustrate how physical/chemical principles affect model outputs; XAI can increase its credibility with the help of DIML. This section will briefly introduce these method and their applications in the electrochemical realm.

3.1. DIML models

The most used approach is to add physical constraints to machine learning models. This can be achieved by modifying loss functions, adding more physics/chemistry-based features, and adapting model architecture. Based on the architecture of the models, they can be grouped as GNN-based, kernel-based, or equivariant GNN-based models.

3.1.1. GNN-based models

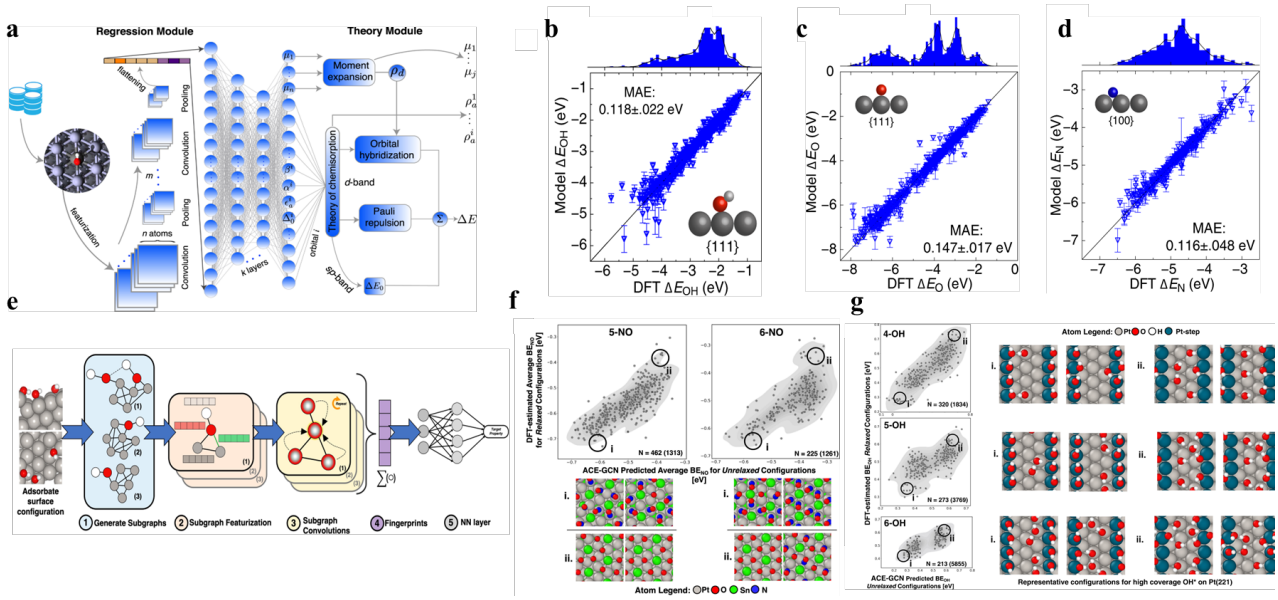

Figure 2. (a) The model architecture of TinNet. The information transitions from the graph representation of an adsorption system to the theory module to calculate adsorption energy ΔE, $\rho_a^1 \dots \rho_a^i$ indicates the projected density of states onto the adsorbate frontier orbital(s) and $\mu_1 \dots \mu_j$ indicates d-band moments. **(b-d)** DFT-calculated vs. TinNet-predicted **(b)** *OH adsorption energies of {111}-transition metal surfaces **(c)** *O adsorption energies at the atop the site of {111}-terminated alloy surfaces **(d)** *N adsorption energies at the hollow site of {100}-terminated alloy surfaces. **(e)** The model architecture of ACE-GCN. **(f, g)** Predictions of conformational stability of unrelaxed **(f)** NO* and **(g)** OH* using the ACEGCN model generated by SurfGraph.

TinNet is one of the most famous models integrating the *d*-band theory into the GNN model to predict the adsorption energies of *OH, *O, and *N on transition-metal surfaces[71]. Two sequential parts are included for the TinNet model, i.e., the regression module and the theory module. The

regression module is similar to the approach used in the Crystal Graph Convolutional Neural Network (CGCNN), where the flattened readout features of the adsorbate-substrate system can be integrated with the *d*-band theory to make the final prediction of adsorption energies based on minimization of the loss function between the predicted properties and physical features in the output layer[52]. By considering the metal *sp*-state, *d*-states (including Pauli repulsion and orbital hybridization), and the projected density of states onto the adsorbate orbitals, the *d*-band theory for chemical bonding at the metal surface was well established (Figure 2a). Comparing the model with the fully connected neural network (FCNN) and CGCNN, TinNet showed comparable predictions but with physical insight, highlighting the importance of the frontier molecular orbital theory and electronic structure methods. Moreover, TinNet demonstrates great generalization capability because it is applicable to various adsorbates and facets: *OH adsorption energies of (111)-transition metal surfaces (Figure 2b), *O at the atop the site of the (111)-terminated bimetallic alloy surfaces (Figure 2c) and *N at the hollow site of (100)-terminated ternary alloy surfaces (Figure 2d). Ghanekar et al. proposed the Adsorbate Chemical Environment-based Graph Convolution Neural Network (ACE-GCN)[72]. ACE-GCN is a screening workflow considering diverse atomistic configurations, which was successfully verified by the cases of NO* adsorption on a Pt_3Sn (111) alloy surface (Figure 3f) and OH* adsorbed on a stepped Pt (221) facet (Figure 3g). Both cases are very complicated in electrocatalysis: one involves strong binding of adsorbates on low-symmetry alloyed surfaces, while the other pertains to directionally dependent adsorption on defective surface structures. The chemical insights originated from ACE-GCN and replaced the full graph with subgraphs, which represent the local environment of the structure and chemistry. These subgraphs not only lead to more accurate predictions of adsorption energy but also a faster training process for considering fewer atoms (Figure 2e).

Physics-informed atomistic line graph neural network (ALIGNN) is another attempt to include physics insight into graph neural network[73, 74]. For example, a recent e study suggested that the adsorption performance for hydrogen on MXenes was based on the H-H bond length and the distance between hydrogen and transition metal. Two physics-informed ALIGNNs were built to predict these descriptors by introducing bond order loss relative to the original ALIGNN. As a result, 23,857 MXene compounds were screened, and the most promising candidate, ScYC, was successfully validated by DFT calculations.

The attention mechanism has recently been applied to GNNs to improve the model's transparency. Zhang et al. proposed the atomic graph attention (AGAT) network to overcome the unaffordable cost of discovering high-entropy electrocatalysts (HEECs) with enormous local environment and phase space by experiments and DFT calculations[75]. Figure 3a shows the model architecture of AGAT. The attention score was calculated on the edges of the graph to decide how much information can be passed from the source atoms, which naturally represents the importance of the nodes. The model was trained and tested for oxygen reduction reaction (ORR) on the RuRhPdIrPt and NiCoFePdPt surfaces, and two

candidates were recommended ($Ni_{0.13}Co_{0.13}Fe_{0.13}Pd_{0.10}Pt_{0.50}$ and $Ni_{0.10}Co_{0.10}Fe_{0.10}Pd_{0.30}Pt_{0.40}$). The attention score was calculated on the edges of the graph and naturally represents the importance of the nodes. The study revealed that an atomic distance of 4.0 Å is a key factor for attention score, highlighting the significance of the local environment (Figure 3b). Also, the AGAT model proves that HEECs can circumvent scaling relations due to their diverse local environments.

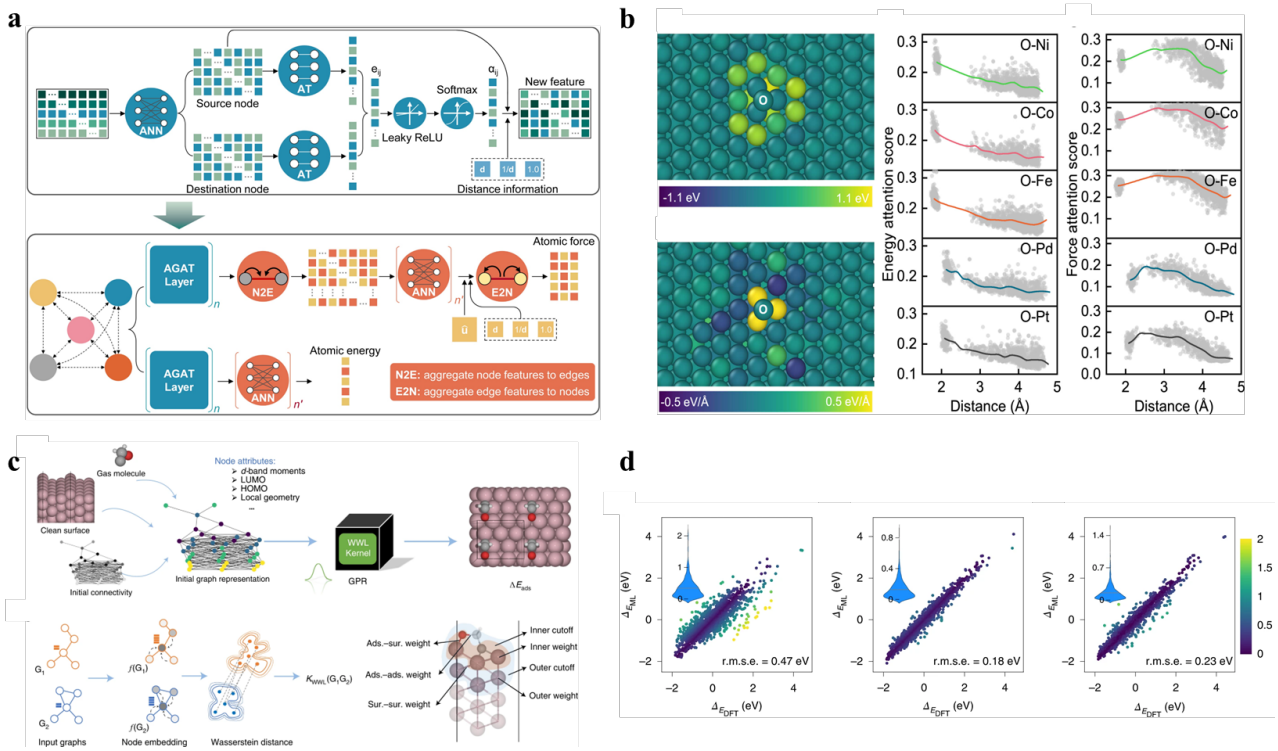

Figure 3. (a) The model architecture of AGAT. The upper border denotes the AGAT layer, and the bottom border denotes the AGAT model. **(b)** The interpretability of the AGAT model. The attention scores of the energy and forces models compared with the energy and forces variations. **(c)** The model architecture of WWL-GPR. The adsorption enthalpy for the relaxed structure is predicted by representing the initial structure as a graph. Node attributes are calculated based on the gas-phase molecule and the pristine surface. The similarity between graphs is assessed using the WWL graph kernel, and this information is then used in a GPR model. **(d)** ML-predicted vs. DFT-calculated adsorption enthalpies in 5-fold cross-validation using RBF-GPR, WWL-GPR, and XGBoost for the simple adsorbates database, respectively.

3.1.2. Kernal-based models

Xu et al. proposed a data-efficient model called the Wasserstein Weisfeiler-Lehman graph kernel and Gaussian Process Regression (WWL-GPR) to predict the binding motifs and adsorption enthalpies of various adsorbates on transition metals (TMs) and their alloys[76, 77]. It was trained on the TMs dataset made up of copper (Cu), rhodium (Rh), palladium (Pd), and cobalt (Co) and achieved a root mean square error (RMSE) of 0.2 eV on the test set of in-domain bimetallic alloys. Figure 3d shows that

WWL-GPR outperforms RBF-GPR and XGBoost for 41 complex adsorbates in various adsorption motifs on surfaces of Cu, Co, Pd and Rh. The high accuracy of WWL-GPR is comparable to DFT calculations, owing to its node attributes through the combination of geometrical information from graph representation and easily accessible physics-informed attributes (e.g. *d*-band moments and HOMO/LUMO energy levels). The node embeddings are then generated to further calculate the Wasserstein distance between their distributions. Based on the Wasserstein distance-based graph similarities, the GPR can be used to make predictions of adsorption enthalpies (Figure 3c).

3.1.3. Equivariant GNN-based models

The GNNs, as mentioned earlier, are usually E(3)-invariant models considering the unchanged properties under the Euclidean group of transformations, which includes translations, rotations, and reflections in three-dimensional space. However, the interactions between molecules and materials are far beyond this. Unlike invariance, equivariance means the output transforms in the same way as the input, which can be written as:

$$\phi\left(T_g(\boldsymbol{x})\right) = S_g\big(\phi(\boldsymbol{x})\big), \tag{3}$$

where $\phi(\cdot)$ is the nonlinear function, $\boldsymbol{x}$ denotes the input vector, T_g is a translation on the input vector and S_g is an equivalent translation on the output set[78]. The formula turns to describe invariance when $S_g = \boldsymbol{I}$, indicating that invariance is just a special case of equivariance. This means that SE(3)-equivariant GNNs usually have more complex geometric constraints than invariant models[79]. Consequently, the equivariant GNN consistently outperforms the invariant GNN in predicting forces, although the difference in energy prediction is insignificant.

OC20 and OC22 are two well-known open-source electrochemical reaction databases[19, 20]. Three primary tasks were proposed, namely Structure to Energy and Forces (S2EF), Initial Structure to Relaxed, Energy (IS2RE), and Initial Structure to Relaxed Structure (IS2RS). Many equivariant models, including SCN[80], eSCN[81], and EquiformerV2[82], have been tested, where EquiformerV2 model is the state-of-the-art model in most cases.

3.2. XAI approaches

Generalized Additive Model (GAM)[83] and Sure Independence Screening and Sparsifying Operator (SISSO) [84] are considered interpretable machine learning models, which are often referred to as "glass box" models due to their inherent simplicity and transparency. GAM is built by constructing additive nonlinear functions of each feature, while SISSO is built on the combination of given features and mathematical operators (e.g., +, -, ×, ÷, log, exp). Both models can give physical or chemical insight into electrochemical reactions, including finding structure descriptors, identifying feature importance, etc.[49, 85-88].

The other type of XAI is called the post-hoc explanation methods, which are usually model-agnostic and can extract the physical or chemical insight after the training. Typically, there are two

approaches to realize the post-hoc explanation. One is visualization, such as local interpretable model-agnostic explanation (LIME)[83]; the other is to calculate the feature importance, e.g., Shapley Additive explanations (SHAP). Because the post-host interpretation approach has the advantage of applying to all models and can be used for both local and global interpretation, it has become one of the top choices of XAI[39, 47, 69, 89-93]. For example, Roy et al. utilized LIME, permutation feature importance (PFI), and accumulated local effects (ALE) for local/global interpretation of the black-box model that established the scaling relationship between CO_2RR intermediates and HEAs surfaces[46]. Moreover, Zhang et al. discovered that SACs in the pyrrole-type coordination could exhibit superior catalytic activity towards NRR when the *d*-orbitals are exactly half occupied and the difference in covalent radius is approximately 140 pm. Their model was based on GBR and employed SHAP explanation methods to interpret the results[41].

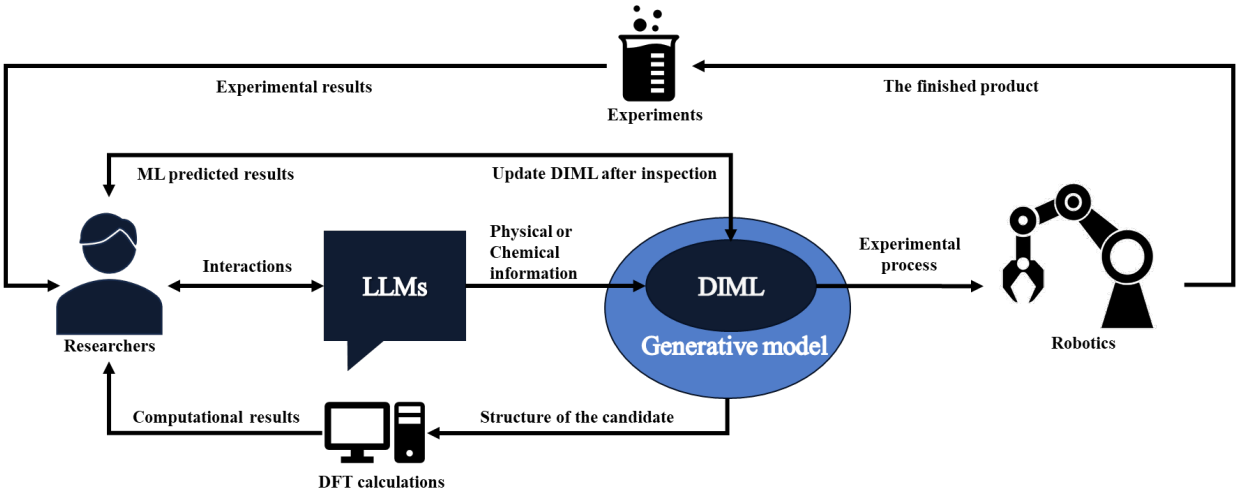

Figure 4. Schematic diagram of the possible future application of DIMLs in an A-lab

4. Discussion and perspectives

DIML shows great potential in developing high-performance electrocatalysts at reduced cost and time. However, it still has some limitations. First, integrating physical/chemical insights will lead to extra burdens for training models, especially when it comes to large datasets. In other words, more complex physical/chemical constraints can result in higher interpretability but lower efficiency, and vice versa[94, 95]. As a result, the trade-off between interpretability and efficiency needs to be navigated. Second, prior knowledge from experts, such as physics/chemistry-based descriptors, is still essential for DIML models. Furthermore, even though some DIML models, like equivariant GNNs, are highly integrated with physical constraints, the complexity of these models makes it difficult to understand the physics/chemistry behind the models. The combination of XAI and equivariant GNNs may become one of the solutions to this issue in the future. Last but not least, there is a lack of benchmarks to determine the interpretability performance of a DIML model.

We believe the limitations mentioned above can gradually be solved, and DIML will become an indispensable part of materials science. Szymanski et al. established an automated laboratory (A-lab) and successfully identified 41 novel materials from 58 candidates over 17 days[96]. This ground-breaking achievement encourages the community to develop more effective A-labs in the future, and DIML will play an important role. DIML-based generation models will likely directly generate physically and chemically meaningful candidates in the future, giving experts physical/chemical insight and reducing the observation error because DIMLs can naturally become the encoder of generative models. Moreover, integrating LLMs with DIMLs will enhance the interactivity of models with humans. In addition, combining these models with robotics will significantly accelerate the synthesization and characterization of new materials with less human interventions demanded. Figure 4 shows possible applications of DIMLs in an A-lab. Researchers can propose their requirements with the help of LLMs. LLMs will generate physic information and send it to DIML for the experimental process predictions. After that, the predicted experimental process will guide the robots to synthesize the predicted new materials. Meanwhile, the DIML-based generative models can generate the structure of candidates. The final product and its structural information will be sent to the researchers for inspection. The data of new materials will be added to the database for training next-generation DIML. This active learning strategy will become a new paradigm for materials science. As DIML keeps progressing, its explosive growth in and contribution to the scientific area will be expected in the near future.

Acknowledgements

This work was supported by the Hong Kong Polytechnic University (project number: P0042711, P0042711and P0048122) and Guangdong Natural Science Foundation (project number: 2024A1515010031).

Reference

(1) Yang, M.; Qin, B.; Si, C.; Sun, X.; Li, B. Electrochemical reaction catalyzed by carbon dots from computational investigations: functional group, dopant, and defect. *Journal of Materials Chemistry A* **2024**.

(2) Ding, K.; Yang, T.; Leung, M. T.; Yang, K.; Cheng, H.; Zeng, M.; Li, B.; Yang, M. Recent advances in the data-driven development of emerging electrocatalysts. *Current Opinion in Electrochemistry* **2023**, 101404.

(3) Banoth, P.; Kandula, C.; Kollu, P. Introduction to electrocatalysts. In *Noble Metal-Free Electrocatalysts: New Trends in Electrocatalysts for Energy Applications. Volume 2*, ACS Publications, 2022; pp 1-37.

(4) Santos, D. M.; Šljukić, B. Advanced materials for electrochemical energy conversion and storage devices. Materials: 2021; Vol. 14, p 7711.

(5) Chen, L.; Zhang, X.; Chen, A.; Yao, S.; Hu, X.; Zhou, Z. Targeted design of advanced electrocatalysts by machine learning. *Chinese Journal of Catalysis* **2022**, *43* (1), 11-32.

(6) Liao, X.; Lu, R.; Xia, L.; Liu, Q.; Wang, H.; Zhao, K.; Wang, Z.; Zhao, Y. Density functional theory for electrocatalysis. *Energy & Environmental Materials* **2022**, *5* (1), 157-185.

(7) Chen, Z. W.; Li, J.; Ou, P.; Huang, J. E.; Wen, Z.; Chen, L.; Yao, X.; Cai, G.; Yang, C. C.; Singh, C. V. Unusual Sabatier principle on high entropy alloy catalysts for hydrogen evolution reactions. *Nature Communications* **2024**, *15* (1), 359.

(8) Peng, J.; Schwalbe-Koda, D.; Akkiraju, K.; Xie, T.; Giordano, L.; Yu, Y.; Eom, C. J.; Lunger, J. R.; Zheng, D. J.; Rao, R. R. Human-and machine-centred designs of molecules and materials for sustainability and decarbonization. *Nature Reviews Materials* **2022**, *7* (12), 991-1009.

(9) Steinmann, S. N.; Wang, Q.; Seh, Z. W. How machine learning can accelerate electrocatalysis discovery and optimization. *Materials Horizons* **2023**, *10* (2), 393-406.

(10) Liu, C.-Y.; Senftle, T. P. Finding physical insights in catalysis with machine learning. *Current Opinion in Chemical Engineering* **2022**, *37*, 100832.

(11) Xin, H.; Mou, T.; Pillai, H. S.; Wang, S.-H.; Huang, Y. Interpretable Machine Learning for Catalytic Materials Design toward Sustainability. *Accounts of Materials Research* **2023**, *5* (1), 22-34.

(12) Kayode, G. O.; Montemore, M. M. Latent Variable Machine Learning Framework for Catalysis: General Models, Transfer Learning, and Interpretability. *JACS Au* **2023**, *4* (1), 80-91.

(13) Rangarajan, S. Artificial intelligence in catalysis. In *Artificial Intelligence in Manufacturing*, Elsevier, 2024; pp 167-204.

(14) Zhang, Y.; Peck, T. C.; Reddy, G. K.; Banerjee, D.; Jia, H.; Roberts, C. A.; Ling, C. Descriptor-free design of multicomponent catalysts. *ACS Catalysis* **2022**, *12* (17), 10562-10571.

(15) Jager, M. O.; Ranawat, Y. S.; Canova, F. F.; Morooka, E. V.; Foster, A. S. Efficient machine-learning-aided screening of hydrogen adsorption on bimetallic nanoclusters. *ACS combinatorial science* **2020**, *22* (12), 768-781.

(16) Fan, X.; Chen, L.; Huang, D.; Tian, Y.; Zhang, X.; Jiao, M.; Zhou, Z. From Single Metals to High-Entropy Alloys: How Machine Learning Accelerates the Development of Metal Electrocatalysts. *Advanced Functional Materials* **2024**, 2401887.

(17) Park, Y.; Hwang, C.-K.; Bang, K.; Hong, D.; Nam, H.; Kwon, S.; Yeo, B. C.; Go, D.; An, J.; Ju, B.-K. Machine learning filters out efficient electrocatalysts in the massive ternary alloy space for fuel cells. *Applied Catalysis B: Environmental* **2023**, *339*, 123128.

(18) Esterhuizen, J. A.; Goldsmith, B. R.; Linic, S. Interpretable machine learning for knowledge generation in heterogeneous catalysis. *Nature catalysis* **2022**, *5* (3), 175-184.
(19) Chanussot, L.; Das, A.; Goyal, S.; Lavril, T.; Shuaibi, M.; Riviere, M.; Tran, K.; Heras-Domingo, J.; Ho, C.; Hu, W. Open catalyst 2020 (OC20) dataset and community challenges. *Acs Catalysis* **2021**, *11* (10), 6059-6072.
(20) Tran, R.; Lan, J.; Shuaibi, M.; Wood, B. M.; Goyal, S.; Das, A.; Heras-Domingo, J.; Kolluru, A.; Rizvi, A.; Shoghi, N. The Open Catalyst 2022 (OC22) dataset and challenges for oxide electrocatalysts. *ACS Catalysis* **2023**, *13* (5), 3066-3084.
(21) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL materials* **2013**, *1* (1).
(22) Saal, J. E.; Kirklin, S.; Aykol, M.; Meredig, B.; Wolverton, C. Materials design and discovery with high-throughput density functional theory: the open quantum materials database (OQMD). *Jom* **2013**, *65*, 1501-1509.
(23) Zhou, J.; Shen, L.; Costa, M. D.; Persson, K. A.; Ong, S. P.; Huck, P.; Lu, Y.; Ma, X.; Chen, Y.; Tang, H. 2DMatPedia, an open computational database of two-dimensional materials from top-down and bottom-up approaches. *Scientific data* **2019**, *6* (1), 86.
(24) Montavon, G.; Hansen, K.; Fazli, S.; Rupp, M.; Biegler, F.; Ziehe, A.; Tkatchenko, A.; Lilienfeld, A.; Müller, K.-R. Learning invariant representations of molecules for atomization energy prediction. *Advances in neural information processing systems* **2012**, *25*.
(25) Rupp, M.; Tkatchenko, A.; Müller, K.-R.; Von Lilienfeld, O. A. Fast and accurate modeling of molecular atomization energies with machine learning. *Physical review letters* **2012**, *108* (5), 058301.
(26) Bartók, A. P.; Kondor, R.; Csányi, G. On representing chemical environments. *Physical Review B—Condensed Matter and Materials Physics* **2013**, *87* (18), 184115.
(27) Willatt, M. J.; Musil, F.; Ceriotti, M. Feature optimization for atomistic machine learning yields a data-driven construction of the periodic table of the elements. *Physical Chemistry Chemical Physics* **2018**, *20* (47), 29661-29668.
(28) Jäger, M. O.; Morooka, E. V.; Federici Canova, F.; Himanen, L.; Foster, A. S. Machine learning hydrogen adsorption on nanoclusters through structural descriptors. *npj Computational Materials* **2018**, *4* (1), 37.
(29) De, S.; Bartók, A. P.; Csányi, G.; Ceriotti, M. Comparing molecules and solids across structural and alchemical space. *Physical Chemistry Chemical Physics* **2016**, *18* (20), 13754-13769.
(30) Mai, H.; Le, T. C.; Chen, D.; Winkler, D. A.; Caruso, R. A. Machine learning for electrocatalyst and photocatalyst design and discovery. *Chemical Reviews* **2022**, *122* (16), 13478-13515.
(31) Batchelor, T. A.; Pedersen, J. K.; Winther, S. H.; Castelli, I. E.; Jacobsen, K. W.; Rossmeisl, J. High-entropy alloys as a discovery platform for electrocatalysis. *Joule* **2019**, *3* (3), 834-845.
(32) Artyushkova, K.; Pylypenko, S.; Olson, T. S.; Fulghum, J. E.; Atanassov, P. Predictive modeling of electrocatalyst structure based on structure-to-property correlations of X-ray photoelectron spectroscopic and electrochemical measurements. *Langmuir* **2008**, *24* (16), 9082-9088.
(33) Hearst, M. A.; Dumais, S. T.; Osuna, E.; Platt, J.; Scholkopf, B. Support vector machines. *IEEE Intelligent Systems and their applications* **1998**, *13* (4), 18-28.
(34) Sun, H.; Li, Y.; Gao, L.; Chang, M.; Jin, X.; Li, B.; Xu, Q.; Liu, W.; Zhou, M.; Sun, X. High throughput screening of single atomic catalysts with optimized local structures for the electrochemical oxygen reduction by machine learning. *Journal of Energy Chemistry* **2023**, *81*, 349-357.

(35) Arjmandi, M.; Fattahi, M.; Motevassel, M.; Rezaveisi, H. Evaluating algorithms of decision tree, support vector machine and regression for anode side catalyst data in proton exchange membrane water electrolysis. *Scientific Reports* **2023**, *13* (1), 20309.
(36) Hossain, S. S.; Ali, S. S.; Rushd, S.; Ayodele, B. V.; Cheng, C. K. Interaction effect of process parameters and Pd‐electrocatalyst in formic acid electro‐oxidation for fuel cell applications: Implementing supervised machine learning algorithms. *International Journal of Energy Research* **2022**, *46* (15), 21583-21597.
(37) Tamtaji, M.; Chen, S.; Hu, Z.; Goddard III, W. A.; Chen, G. A Surrogate Machine Learning Model for the Design of Single-Atom Catalyst on Carbon and Porphyrin Supports towards Electrochemistry. *The Journal of Physical Chemistry C* **2023**, *127* (21), 9992-10000.
(38) Anbari, E.; Adib, H.; Iranshahi, D. Experimental investigation and development of a SVM model for hydrogenation reaction of carbon monoxide in presence of Co–Mo/Al2O3 catalyst. *Chemical Engineering Journal* **2015**, *276*, 213-221.
(39) Sun, J.; Chen, A.; Guan, J.; Han, Y.; Liu, Y.; Niu, X.; He, M.; Shi, L.; Wang, J.; Zhang, X. Interpretable machine learning‐assisted high‐throughput screening for understanding NRR electrocatalyst performance modulation between active center and C‐N coordination. *Energy & Environmental Materials* **2024**, *7* (5), e12693.
(40) Tan, S.; Wang, R.; Song, G.; Qi, S.; Zhang, K.; Zhao, Z.; Yin, Q. Machine learning and Shapley Additive Explanation-based interpretable prediction of the electrocatalytic performance of N-doped carbon materials. *Fuel* **2024**, *355*, 129469.
(41) Zhang, Y.; Wang, Y.; Ma, N.; Fan, J. Directly predicting N2 electroreduction reaction free energy using interpretable machine learning with non-DFT calculated features. *Journal of Energy Chemistry* **2024**, *97*, 139-148. DOI: https://doi.org/10.1016/j.jechem.2024.05.042.
(42) Wei, C.; Shi, D.; Yang, Z.; Xue, Z.; Liu, S.; Li, R.; Mu, T. Data-driven design of double-atom catalysts with high H 2 evolution activity/CO 2 reduction selectivity based on simple features. *Journal of Materials Chemistry A* **2023**, *11* (34), 18168-18178.
(43) Ying, Y.; Fan, K.; Luo, X.; Qiao, J.; Huang, H. Unravelling the origin of bifunctional OER/ORR activity for single-atom catalysts supported on C 2 N by DFT and machine learning. *Journal of Materials Chemistry A* **2021**, *9* (31), 16860-16867.
(44) Lin, S.; Xu, H.; Wang, Y.; Zeng, X. C.; Chen, Z. Directly predicting limiting potentials from easily obtainable physical properties of graphene-supported single-atom electrocatalysts by machine learning. *Journal of Materials Chemistry A* **2020**, *8* (11), 5663-5670.
(45) Lu, S.; Song, P.; Jia, Z.; Gao, Z.; Wang, Z.; Peng, T.; Bai, X.; Jiang, Q.; Cui, H.; Tian, W. Symbolic transform optimized convolutional neural network model for high-performance prediction and analysis of MXenes hydrogen evolution reaction catalysts. *International Journal of Hydrogen Energy* **2024**, *85*, 200-209.
(46) Roy, D.; Charan Mandal, S.; Das, A.; Pathak, B. Unravelling CO2 Reduction Reaction Intermediates on High Entropy Alloy Catalysts: An Interpretable Machine Learning Approach to Establish Scaling Relations. *Chemistry–A European Journal* **2024**, *30* (6), e202302679.

(47) Wang, Y.; Zhang, Y.; Ma, N.; Zhao, J.; Xiong, Y.; Luo, S.; Fan, J. Machine learning accelerated catalysts design for CO reduction: An interpretability and transferability analysis. *Journal of Materials Science & Technology* **2025**, *213*, 14-23.
(48) Jia, X.; Li, H. Machine learning enabled exploration of multicomponent metal oxides for catalyzing oxygen reduction in alkaline media. *Journal of Materials Chemistry A* **2024**.
(49) Liu, X.; Peng, H.-J. Toward Next-Generation Heterogeneous Catalysts: Empowering Surface Reactivity Prediction with Machine Learning. *Engineering* **2024**.
(50) Yang, H.; Zhao, J.; Wang, Q.; Liu, B.; Luo, W.; Sun, Z.; Liao, T. Convolutional Neural Networks and Volcano Plots: Screening and Prediction of Two-Dimensional Single-Atom Catalysts. *arXiv preprint arXiv:2402.03876* **2024**.
(51) Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E. Neural message passing for quantum chemistry. In *International conference on machine learning*, 2017; PMLR: pp 1263-1272.
(52) Xie, T.; Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Physical review letters* **2018**, *120* (14), 145301.
(53) Park, Y.; Kim, J.; Hwang, S.; Han, S. Scalable Parallel Algorithm for Graph Neural Network Interatomic Potentials in Molecular Dynamics Simulations. *Journal of Chemical Theory and Computation* **2024**.
(54) Merchant, A.; Batzner, S.; Schoenholz, S. S.; Aykol, M.; Cheon, G.; Cubuk, E. D. Scaling deep learning for materials discovery. *Nature* **2023**, *624* (7990), 80-85.
(55) Deng, B.; Zhong, P.; Jun, K.; Riebesell, J.; Han, K.; Bartel, C. J.; Ceder, G. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nature Machine Intelligence* **2023**, *5* (9), 1031-1041.
(56) Chen, C.; Ong, S. P. A universal graph deep learning interatomic potential for the periodic table. *Nature Computational Science* **2022**, *2* (11), 718-728.
(57) Tang, D.; Ketkaew, R.; Luber, S. Machine Learning Interatomic Potentials for Catalysis. *Chemistry–A European Journal* **2024**, e202401148.
(58) Batatia, I.; Kovacs, D. P.; Simm, G.; Ortner, C.; Csányi, G. MACE: Higher order equivariant message passing neural networks for fast and accurate force fields. *Advances in Neural Information Processing Systems* **2022**, *35*, 11423-11436.
(59) Batatia, I.; Batzner, S.; Kovács, D. P.; Musaelian, A.; Simm, G. N.; Drautz, R.; Ortner, C.; Kozinsky, B.; Csányi, G. The design space of e (3)-equivariant atom-centered interatomic potentials. *arXiv preprint arXiv:2205.06643* **2022**.
(60) Riebesell, J.; Goodall, R. E.; Jain, A.; Benner, P.; Persson, K. A.; Lee, A. A. Matbench Discovery--An evaluation framework for machine learning crystal stability prediction. *arXiv preprint arXiv:2308.14920* **2023**.
(61) Batatia, I.; Benner, P.; Chiang, Y.; Elena, A. M.; Kovács, D. P.; Riebesell, J.; Advincula, X. R.; Asta, M.; Baldwin, W. J.; Bernstein, N. A foundation model for atomistic materials chemistry. *arXiv preprint arXiv:2401.00096* **2023**.
(62) Achiam, J.; Adler, S.; Agarwal, S.; Ahmad, L.; Akkaya, I.; Aleman, F. L.; Almeida, D.; Altenschmidt, J.; Altman, S.; Anadkat, S. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774* **2023**.
(63) Yao, Y.; Duan, J.; Xu, K.; Cai, Y.; Sun, Z.; Zhang, Y. A survey on large language model (llm) security and privacy: The good, the bad, and the ugly. *High-Confidence Computing* **2024**, 100211.

(64) Chang, Y.; Wang, X.; Wang, J.; Wu, Y.; Yang, L.; Zhu, K.; Chen, H.; Yi, X.; Wang, C.; Wang, Y. A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology* **2024**, *15* (3), 1-45.
(65) Augenstein, I.; Baldwin, T.; Cha, M.; Chakraborty, T.; Ciampaglia, G. L.; Corney, D.; DiResta, R.; Ferrara, E.; Hale, S.; Halevy, A. Factuality challenges in the era of large language models and opportunities for fact-checking. *Nature Machine Intelligence* **2024**, 1-12.
(66) Patil, R.; Gudivada, V. A review of current trends, techniques, and challenges in large language models (llms). *Applied Sciences* **2024**, *14* (5), 2074.
(67) Beltagy, I.; Lo, K.; Cohan, A. SciBERT: A pretrained language model for scientific text. *arXiv preprint arXiv:1903.10676* **2019**.
(68) Wang, L.; Chen, X.; Du, Y.; Zhou, Y.; Gao, Y.; Cui, W. CataLM: Empowering Catalyst Design Through Large Language Models. *arXiv preprint arXiv:2405.17440* **2024**.
(69) Ding, R.; Wang, X. B.; Tan, A. D.; Li, J.; Liu, J. G. Unlocking New Insights for Electrocatalyst Design: A Unique Data Science Workflow Leveraging Internet-Sourced Big Data. *Acs Catalysis* **2023**. DOI: 10.1021/acscatal.3c01914.
(70) Minh, D.; Wang, H. X.; Li, Y. F.; Nguyen, T. N. Explainable artificial intelligence: a comprehensive review. *Artificial Intelligence Review* **2022**, 1-66.
(71) Wang, S.-H.; Pillai, H. S.; Wang, S.; Achenie, L. E.; Xin, H. Infusing theory into deep learning for interpretable reactivity prediction. *Nature communications* **2021**, *12* (1), 5288.
(72) Ghanekar, P. G.; Deshpande, S.; Greeley, J. Adsorbate chemical environment-based machine learning framework for heterogeneous catalysis. *Nature Communications* **2022**, *13* (1), 5788.
(73) Cheng, J.; Li, T.; Wang, Y.; Ati, A. H.; Sun, Q. High-throughput screening of MXenes for hydrogen storage via graph neural network. *Applied Surface Science* **2023**, *641*, 158560.
(74) Choudhary, K.; DeCost, B. Atomistic line graph neural network for improved materials property predictions. *npj Computational Materials* **2021**, *7* (1), 185.
(75) Zhang, J.; Wang, C.; Huang, S.; Xiang, X.; Xiong, Y.; Xu, B.; Ma, S.; Fu, H.; Kai, J.; Kang, X. Design high-entropy electrocatalyst via interpretable deep graph attention learning. *Joule* **2023**, *7* (8), 1832-1851.
(76) Xu, W.; Reuter, K.; Andersen, M. Predicting binding motifs of complex adsorbates using machine learning with a physics-inspired graph representation. *Nature Computational Science* **2022**, *2* (7), 443-450.
(77) Togninalli, M.; Ghisu, E.; Llinares-López, F.; Rieck, B.; Borgwardt, K. Wasserstein weisfeiler-lehman graph kernels. *Advances in neural information processing systems* **2019**, *32*.
(78) Satorras, V. G.; Hoogeboom, E.; Welling, M. E (n) equivariant graph neural networks. In *International conference on machine learning*, 2021; PMLR: pp 9323-9332.
(79) Zhang, X.; Wang, L.; Helwig, J.; Luo, Y.; Fu, C.; Xie, Y.; Liu, M.; Lin, Y.; Xu, Z.; Yan, K. Artificial intelligence for science in quantum, atomistic, and continuum systems. *arXiv preprint arXiv:2307.08423* **2023**.
(80) Zitnick, L.; Das, A.; Kolluru, A.; Lan, J.; Shuaibi, M.; Sriram, A.; Ulissi, Z.; Wood, B. Spherical channels for modeling atomic interactions. *Advances in Neural Information Processing Systems* **2022**, *35*, 8054-8067.
(81) Passaro, S.; Zitnick, C. L. Reducing SO (3) convolutions to SO (2) for efficient equivariant GNNs. In *International Conference on Machine Learning*, 2023; PMLR: pp 27420-27438.

(82) Liao, Y.-L.; Wood, B.; Das, A.; Smidt, T. Equiformerv2: Improved equivariant transformer for scaling to higher-degree representations. *arXiv preprint arXiv:2306.12059* **2023**.
(83) Hastie, T. J. Generalized additive models. In *Statistical models in S*, Routledge, 2017; pp 249-307.
(84) Ouyang, R.; Curtarolo, S.; Ahmetcik, E.; Scheffler, M.; Ghiringhelli, L. M. SISSO: A compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates. *Physical Review Materials* **2018**, *2* (8), 083802.
(85) Lin, X.; Wang, Y.; Chang, X.; Zhen, S.; Zhao, Z. J.; Gong, J. High‐Throughput Screening of Electrocatalysts for Nitrogen Reduction Reactions Accelerated by Interpretable Intrinsic Descriptor. *Angewandte Chemie* **2023**, *135* (19), e202300122.
(86) Ding, Z.; Pang, Y.; Ma, A.; Liu, Z.; Wang, Z.; Fan, G.; Xu, H. Single-atom catalysts based on two-dimensional metalloporphyrin monolayers for electrochemical nitrate reduction to ammonia by first-principles calculations and interpretable machine learning. *International Journal of Hydrogen Energy* **2024**, *80*, 586-598.
(87) Shu, W.; Li, J.; Liu, J.-X.; Zhu, C.; Wang, T.; Feng, L.; Ouyang, R.; Li, W.-X. Structure Sensitivity of Metal Catalysts Revealed by Interpretable Machine Learning and First-Principles Calculations. *Journal of the American Chemical Society* **2024**, *146* (12), 8737-8745.
(88) Su, Y.; Wang, X.; Ye, Y.; Xie, Y.; Xu, Y.; Jiang, Y.; Wang, C. Automation and machine learning augmented by large language models in a catalysis study. *Chemical Science* **2024**, *15* (31), 12200-12233.
(89) Omidvar, N.; Wang, S. H.; Huang, Y.; Pillai, H. S.; Athawale, A.; Wang, S.; Achenie, L. E.; Xin, H. Explainable AI for optimizing oxygen reduction on Pt monolayer core–shell catalysts. *Electrochemical Science Advances* **2024**, e202300028.
(90) Li, Y.; Zhang, X.; Li, T.; Chen, Y.; Liu, Y.; Feng, L. Accelerating materials discovery for electrocatalytic water oxidation via center-environment deep learning in spinel oxides. *Journal of Materials Chemistry A* **2024**, *12* (30), 19362-19377.
(91) Roh, J.; Park, H.; Kwon, H.; Joo, C.; Moon, I.; Cho, H.; Ro, I.; Kim, J. Interpretable machine learning framework for catalyst performance prediction and validation with dry reforming of methane. *Applied Catalysis B: Environmental* **2024**, *343*, 123454.
(92) Ding, R.; Chen, Y.; Chen, P.; Wang, R.; Wang, J.; Ding, Y.; Yin, W.; Liu, Y.; Li, J.; Liu, J. Machine learning-guided discovery of underlying decisive factors and new mechanisms for the design of nonprecious metal electrocatalysts. *ACS Catalysis* **2021**, *11* (15), 9798-9808.
(93) Pillai, H. S.; Li, Y.; Wang, S.-H.; Omidvar, N.; Mu, Q.; Achenie, L. E.; Abild-Pedersen, F.; Yang, J.; Wu, G.; Xin, H. Interpretable design of Ir-free trimetallic electrocatalysts for ammonia oxidation with graph neural networks. *Nature communications* **2023**, *14* (1), 792.
(94) Pablo-García, S.; Morandi, S.; Vargas-Hernández, R. A.; Jorner, K.; Ivković, Ž.; López, N.; Aspuru-Guzik, A. Fast evaluation of the adsorption energy of organic molecules on metals via graph neural networks. *Nature Computational Science* **2023**, *3* (5), 433-442.
(95) Schütt, K.; Unke, O.; Gastegger, M. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *International Conference on Machine Learning*, 2021; PMLR: pp 9377-9388.
(96) Szymanski, N. J.; Rendy, B.; Fei, Y.; Kumar, R. E.; He, T.; Milsted, D.; McDermott, M. J.; Gallant, M.; Cubuk, E. D.; Merchant, A. An autonomous laboratory for the accelerated synthesis of novel materials. *Nature* **2023**, *624* (7990), 86-91.