

# Active learning across intermetallics to guide discovery of electrocatalysts for CO<sub>2</sub> reduction and H<sub>2</sub> evolution

Kevin Tran & \*Zachary W. Ulissi  
Department of Chemical Engineering,  
Carnegie Mellon University,  
Pittsburgh, PA 15217  
zulissi@andrew.cmu.edu

## Abstract

Electrochemical reduction of CO<sub>2</sub> or H<sub>2</sub> evolution from water can be used to store renewable energy that is produced intermittently. Scale up of these reactions requires the discovery of effective electrocatalysts, but the electrocatalyst search space is too large to explore exhaustively. Here we present a theoretical, fully automated screening method that uses a combination of machine learning and optimization to guide density functional theory calculations, which are then used to predict electrocatalyst performance. We demonstrate feasibility of this method by screening various alloys of 31 different elements, thereby performing a screening that encompasses 50% of the d-block elements and 33% of the p-block elements. This method has thus far identified 130 candidate surfaces across 54 alloys for CO<sub>2</sub> reduction and 258 surfaces across 102 alloys for H<sub>2</sub> evolution. We use qualitative analyses to prioritize the top candidates for experimental validation.

Global energy demands have increased over time and are likely to continue increasing.<sup>1,2</sup> Meeting these demands using only fossil fuels may not be possible because doing so may negatively impacts the world's environments, climate, and biodiversity.<sup>3–6</sup> Alternative methods for energy production and storage include solar fuels, which are synthetic fuels created with photovoltaic energy or with photoelectrochemical cells.<sup>7,8</sup> Examples include H<sub>2</sub> created by the electrochemical reduction of water, which can be combusted or used in hydrogen fuel cells, and synthetic hydrocarbons created by reducing CO<sub>2</sub>. Solar fuel production is currently inhibited by a lack of sufficiently active, efficient, selective, stable, and low-cost catalysts.<sup>9</sup>

High-performing catalysts can be discovered using *ab initio* methods such as Density Functional Theory (DFT) to predict catalyst properties. A common approach is to use DFT to predict thermodynamic energy descriptors that correlate with detailed microkinetic model results or experimental measurements of catalyst activity and selectivity. For the CO<sub>2</sub> reduction reaction (CO<sub>2</sub>RR), the CO adsorption energy is a common descriptor for predicting activity towards hydrocarbon production.<sup>10</sup> This method of using adsorption energies to predict performance is general and has been applied to many reaction systems,<sup>11</sup> including the Hydrogen Evolution Reaction (HER).<sup>12</sup> Of course, single descriptors cannot completely describe the electrocatalytic performance of an intermetallic, which requires analyses of surface stability, high coverage thermodynamics, and electrochemical kinetics. These analyses take considerable resources, and full theoretical studies cannot keep pace with accelerating experimental studies of intermetallics. Thus, a screening method for prioritizing experimental and theoretical studies is valuable.

Screening large search spaces is nontrivial. Some methods address this by focusing on limited search spaces, such as bimetallic A<sub>3</sub>B crystals<sup>13</sup> or simple cubic lattice intermetallics.<sup>14</sup> Other approaches include creating databases of electronic structure calculations.<sup>15,16</sup> Such wide-scoped datasets are necessary to perform robust screening studies, but creating the results in these databases has required collaboration between dozens of DFT experts. A consequence of this approach is that researchers spend a substantial portion of their time configuring, managing, and waiting for DFT calculations, which are inherently time consuming. The personnel cost of configuring and managing DFT calculations has been addressed by creating computer-science-based solutions. Generalized frameworks exist that are able to enumerate surfaces and adsorption sites on arbitrary intermetallic bulk structures.<sup>17,18</sup> Software exists that manages computationally

intensive calculations across multiple computing clusters.<sup>19</sup> These solutions are vital for performing high-throughput material screenings.

The computational cost of theoretical materials screenings has been partially addressed by the development of machine learning (ML) methods to accelerate DFT calculations. ML regressions on DFT data can yield formation energy predictions at a fraction of the computational cost<sup>20,21</sup> or accelerate screenings within individual bimetallics.<sup>22</sup> Neural networks can accelerate nudged elastic band studies of reaction kinetics<sup>23</sup> or enable the study of disorder on intermetallic configurations.<sup>24</sup> One of the shortfalls of many of these approaches is that they require experts to manually prioritize new materials to screen and study. Many also rely on DFT-computed surface features such as the d-band center, or predict activity only at the surface level without distinguishing between surface sites.<sup>14</sup>

There are some methods that have been used to accelerate computational screenings that may not have been applied to electrocatalyst discovery yet. Surrogate-based optimization<sup>25</sup> is a method where a surrogate model is built to replace a more computationally expensive model, and then an objective function is optimized on the cheaper surrogate model. Another accelerative method is active machine learning—also known as optimal design of experiments.<sup>26</sup> With this method, a surrogate model is created from a given dataset, and then the model is used to select which data should be obtained next. The selected data is added to the original dataset and then used to create an updated surrogate model. The process is repeated iteratively such that the surrogate model is improved continuously. This method of iterative surrogate model screening has already been used in other fields such as discovery of light emitting diodes,<sup>27</sup> drug discovery,<sup>28</sup> or molecular property prediction.<sup>29</sup>

We created a workflow that borrows ideas from both surrogate-based optimization and active machine learning. We used this workflow to screen a search space of 1,499 intermetallics for potentially selective catalysts for CO<sub>2</sub>RR and HER. The workflow identified 54 intermetallics that have surfaces with near-optimal descriptors for CO<sub>2</sub>RR and 102 intermetallics for HER, including both already-discovered and undiscovered catalysts. We then use a qualitative, heuristic method for down-selecting these compounds further, resulting in a shortened list of 10 intermetallics for CO<sub>2</sub>RR and 14 intermetallics for HER. Viewing the intermetallic design space holistically also allows trends and design rules to be identified that are difficult to see from small case studies. This workflow can be extended easily to other reaction chemistries for which ideal thermodynamic descriptors are known.

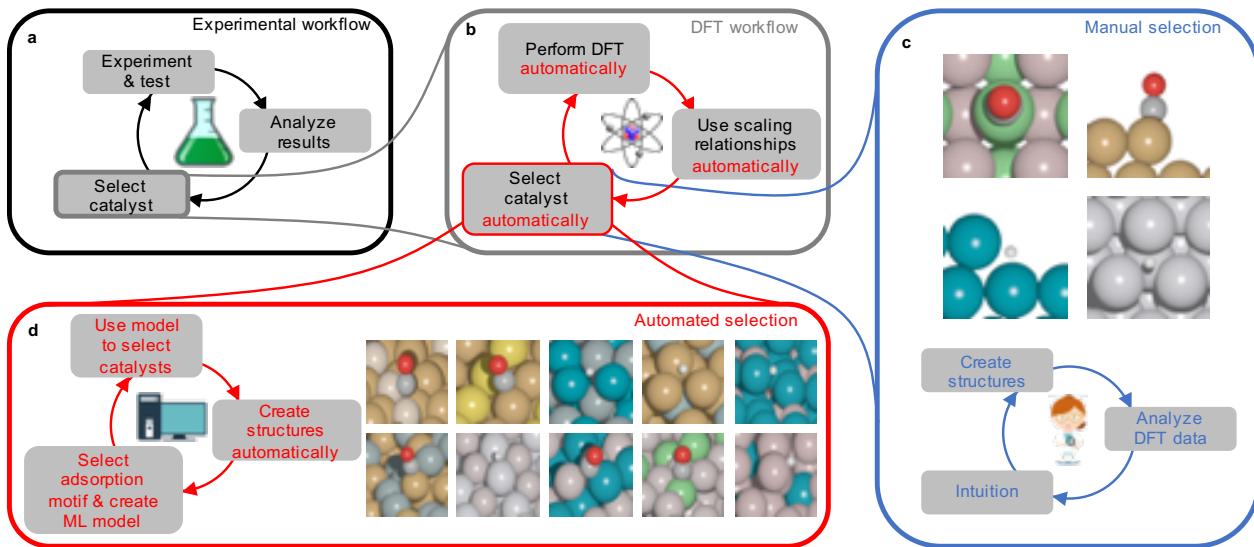
## Results

### Framework Construction

Our workflow uses machine learning models to search an arbitrarily large design space of intermetallic crystals and surfaces for near-optimal activity (Figure 1). Surfaces are searched for ideal CO and H adsorption energies, which are indicative of catalyst performance for CO<sub>2</sub> reduction<sup>10</sup> and H<sub>2</sub> evolution,<sup>12</sup> respectively. The workflow verifies the adsorption energies of these sites by performing DFT calculations automatically. DFT results are stored in a database, which is used to retrain the machine learning models. This yields a closed feedback loop of ML screening, DFT verification, and ML retraining that produces a database of DFT results that grows continuously, systematically, and without the need for user interaction. It is important to note that this workflow does not use ML to accelerate calculations of user-supplied systems. Instead, it uses ML to guide full-accuracy DFT screenings. Thus we use DFT to perform a surrogate-based optimization, and we use an active learning feedback loop as an optimization guide.

### Enumeration of search space

To generate a search space of adsorption sites, we obtained 1,499 different intermetallic combinations from the Materials Project<sup>30</sup> across 31 different elements (Supplementary Figure 1). We enumerated the surfaces on each intermetallic using pymatgen<sup>17</sup> and then used Delaunay triangulation<sup>18</sup> to enumerate all adsorption sites on each surface; see the Methods section below for additional details. This search space encompassed 50% of the d-block elements and 33% of the p-block elements. Every possible adsorption site on each surface was considered for potential activity by our ML models.



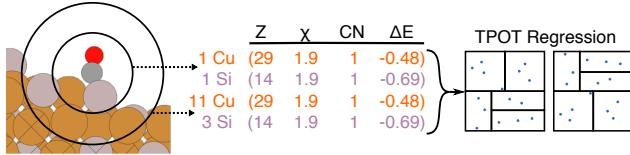
**Figure 1 | Workflow for automating theoretical materials discovery.** **a**, The experimental workflow for finding catalysts is accelerated by **b**, a DFT workflow for screening the catalysts *ab initio*. **c**, Conventional workflows (blue) require scientific intuition to select candidates for DFT screenings, while **d**, our workflow (red) uses machine learning to select candidates systematically and automatically. Red text outlines the framework that this study used to perform active machine learning/surrogate-based optimization.

## Active learning optimization of catalyst descriptors

We enabled the machine learning of catalyst descriptors by developing a fingerprinting method to represent an intermetallic adsorption site numerically (Figure 2). For each site, the types of elements coordinated with the adsorbate were tabulated. Each element type was described with a vector of four numbers: the atomic number of the element ( $Z$ ), the Pauling electronegativity of the element ( $\chi$ ), the number of atoms of the element coordinated with the adsorbate (CN) as determined by Voronoi tessellation performed by pymatgen,<sup>17</sup> and the median adsorption energy between the adsorbate and the pure element ( $\Delta\tilde{E}$ ).  $\Delta\tilde{E}$  values were calculated from our own database of adsorption energies, and  $\chi$  values were obtained from the Mendeleev database.<sup>31</sup> We repeated this vector creation process on the second shell of atoms that are bonded to the coordination atoms. One issue with this method is that it yields a variable number of features. We addressed this issue using a method found in literature.<sup>32</sup> See the Methods section below for additional details. Note also that the illustration in Figure 2 this is a simplification. The real fingerprint vector has 4 items per element and 4 elements per shell, which yields a total of 32 items per fingerprint vector.

These fingerprints were chosen using a combination of intuition, trial-and-error, and success in other surrogate modeling studies. To account for bulk steric effects, atomic radii have been used as features.<sup>32</sup> Atomic radii may change depending on the local environment though, so elemental periods and groups may be appropriate substitutes for atomic radius. Initial, heuristic investigations showed negligible difference in performance between using period/group and atomic number though, so we used the atomic number because of its relatively small dimensionality. To account of electronic affinity effects, Pauling electronegativity has been shown to be a successful feature.<sup>14</sup> To account for both sterics and environmental electronic effects, the coordination number has been shown to be a successful feature.<sup>33</sup> To improve predictive capability, crude estimates of properties have been shown to be successful.<sup>34</sup> In this setting, the crude estimate of adsorption energy on a specific site is  $\Delta\tilde{E}$ .

An automated machine learning package, TPOT,<sup>35</sup> was then used to select a machine learning regression method to predict adsorption energies from the site fingerprints. TPOT's recommended modeling pipeline changed regularly due to the stochastic nature of TPOT and the constantly changing training dataset. To aid prediction we used a preprocessing pipeline to shift and scale each feature across all data points so that the averages and variances for each feature were zero and one, respectively. We also performed a principal component analysis on the fingerprints to orthogonalize the feature space. Supplementary Note 1 outlines



**Figure 2 | Fingerprint of coordination site.** Adsorption sites are reduced to numerical representations, or fingerprints, and these fingerprints are used as model features by TPOT<sup>35</sup> to predict  $\Delta E_{CO}$ .  $Z$  is the atomic number of an element;  $x$  is the Pauling electronegativity;  $CN$  is the number of atoms of an element within a shell; and  $(\tilde{\Delta E})$  is the median adsorption energy between the adsorbate and the pure element.

the TPOT settings used, and the Supplementary Methods section outlines other regression techniques and feature representations we tested during development.

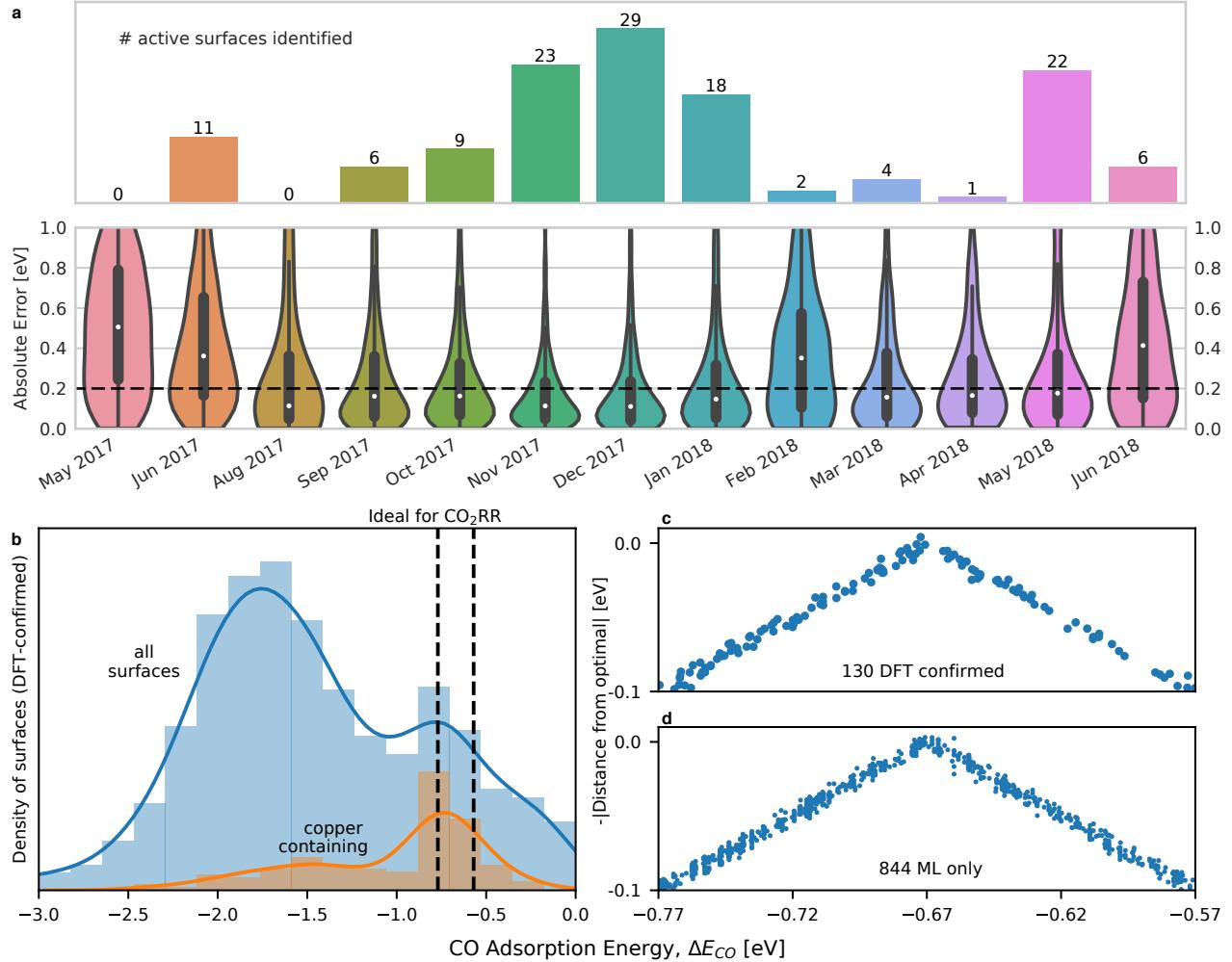
The regression methods illustrated in Figure 2 were combined with all available DFT data to train and update surrogate models daily. These models were used to aid in selecting adsorption sites for DFT calculation. First, the trained models were used to estimate  $\Delta E_{CO}$  and  $\Delta E_H$  for all the adsorption sites that we enumerated. These estimates were pooled with the explicit DFT results stored in the database, and ML-estimated adsorption energies were removed if the DFT result existed for that exact site/surface/adsorbate, leaving only one prediction/estimate per adsorption site. Then we defined the strongest binding energy on each surface as the low-coverage adsorption energy of that surface. The surfaces whose low-coverage adsorption energies were predicted to be near-optimal (as established in the Methods section) were selected for DFT calculation with a Gaussian probability defined by the distance of the predicted site from the optimal values with a standard deviation of 0.2 eV.

The targeting of specific adsorption energies with near-optimal values allows us to exploit the knowledge that the surrogate models have learned thus far, while the addition of Gaussian noise is a heuristic method that allows us to explore search spaces that the surrogate model normally would not suggest. Approximately 80% of the calculations were dedicated to this descriptor optimization goal. The remaining 20% of our resources were dedicated to simulating all of the sites on surfaces whose low-coverage adsorption energies were closest to the top of the volcano—i.e., nearest to a  $\Delta E_{CO}$  of -0.67 eV and a  $\Delta E_H$  of -0.27 eV for CO<sub>2</sub>RR and HER, respectively. This mitigated the chances of finding a false minimum adsorption energy on a surface. The regression and surrogate model prediction was performed once per day, and the Gaussian selection of DFT calculations was performed four times per day. In total, 42,785 DFT calculations of adsorption energies were completed by this study at a rate of approximately 200–300 calculations per day. See the Methods section below for details regarding these DFT calculations.

## Performance of active learning optimization

Due to the iterative nature of the surrogate modeling, we calculated prediction errors via evaluation on a rolling forecasting origin.<sup>36</sup> Specifically: We retrospectively trained a surrogate model on the first 200 data points that we obtained and then calculated the prediction errors between the next 200 data points and this first model’s predictions of these points. We then trained a second model on the first 400 points and then calculated the prediction errors between the next 200 data points and this second model’s predictions of these points. We performed this iteratively until we obtained a single prediction error for every data point, excluding the first 200 points. All prediction errors are plotted against time in Figure 3a and Supplementary Figure 2 for  $\Delta E_{CO}$  and  $\Delta E_H$ , respectively, along with a record of the number of near-optimal surfaces identified over time. The root-mean-squared-error (RMSE), mean absolute error (MAE), and median absolute deviation (MAD) across all of the time-dependent  $\Delta E_{CO}$  predictions are 0.46, 0.29, and 0.17 eV, respectively. The RMSE, MAE, and MAD of  $\Delta E_H$  predictions are 0.41, 0.24, and 0.16 eV, respectively. Note that we chose 200 as the step size because our framework was able to perform at least 200 calculations per day, and so a step size of 200 served as a proxy for surrogate model updates.

The profile of the prediction errors over time provide us with practical insights into our framework. Between November 2017 and January 2018, fellow users of our computing clusters reduced their usage during the holiday season. This allowed our automated framework to effectively consume their unused capacity, thereby increasing our calculation throughput. This temporary increase in throughput is the likely cause



**Figure 3 | Identification of surfaces with near-optimal  $\Delta E_{CO}$  values for CO<sub>2</sub>RR.** **a**, Both the number of near-optimal surfaces identified and violin plots of the absolute error in predicting  $\Delta E_{CO}$  as a function of time. The outer shells of the violins bound all data; narrow vertical lines bound 95% of the data; thick vertical lines bound 50% of the data; and white dots represent medians. Apparently missing months are not shown because no data were collected during those months. **b**, The normalized distribution of the low-coverage, DFT-calculated CO adsorption energies of all of the DFT-analyzed surfaces in this study. The sub-distribution for copper is also illustrated in orange. Dashed lines indicate the  $\pm 0.1$  eV range around the optimal  $\Delta E_{CO}$  value of -0.67 eV. **c**, Surfaces whose low-coverage CO adsorption energies have been calculated and verified with DFT. **d**, Surfaces whose low-coverage CO adsorption energies have been calculated only by the machine learning models.

of the relatively high rate of surface identification, and the improvement in prediction errors at this time may have been caused by an improved sampling of the search space. Then in February 2018, we expanded the number of elements in our search space from ca. 20 elements to the 31 elements that we are searching currently. We seeded this new search space by manually queueing calculations with the new elements. This expansion in search space and subsequent seeding may have caused the decline in both predictive performance and identification rate. In May 2018, we refined our zero-point energy, entropic, and solvation correction calculations to what is now shown in the Supplementary Methods section. This refinement may have caused the increase in the number of surfaces identified during that month. Thus the trends in the prediction errors and identification rates are confounded with both the methods we used and various managerial events, such as changes in throughput capacity, changes in search space, or changes in optimization targets.

Performance metrics and plots that are typically used to judge static surrogate models are shown in the Supplementary Note 2, such as train/test errors, parity plots, and learning curves. Notably, the errors calculated via rolling forecasting origin are generally larger than the errors calculated from the classical train/test split and learning curve methods. This is because classical train/test splitting methods allow the training sets to share the same sampling space as the test sets. Our workflow often searches unexplored sampling spaces though, which are more difficult to predict. Thus the rolling forecasting origin method of evaluating error is more representative of our use scenario because it restricts the models' from seeing data that they would not normally see in practice. We hypothesize that the rolling forecasting origin errors and the train/test errors would converge if we had sampled the search space sufficiently.

## Discovering potential intermetallic catalysts for CO<sub>2</sub> reduction

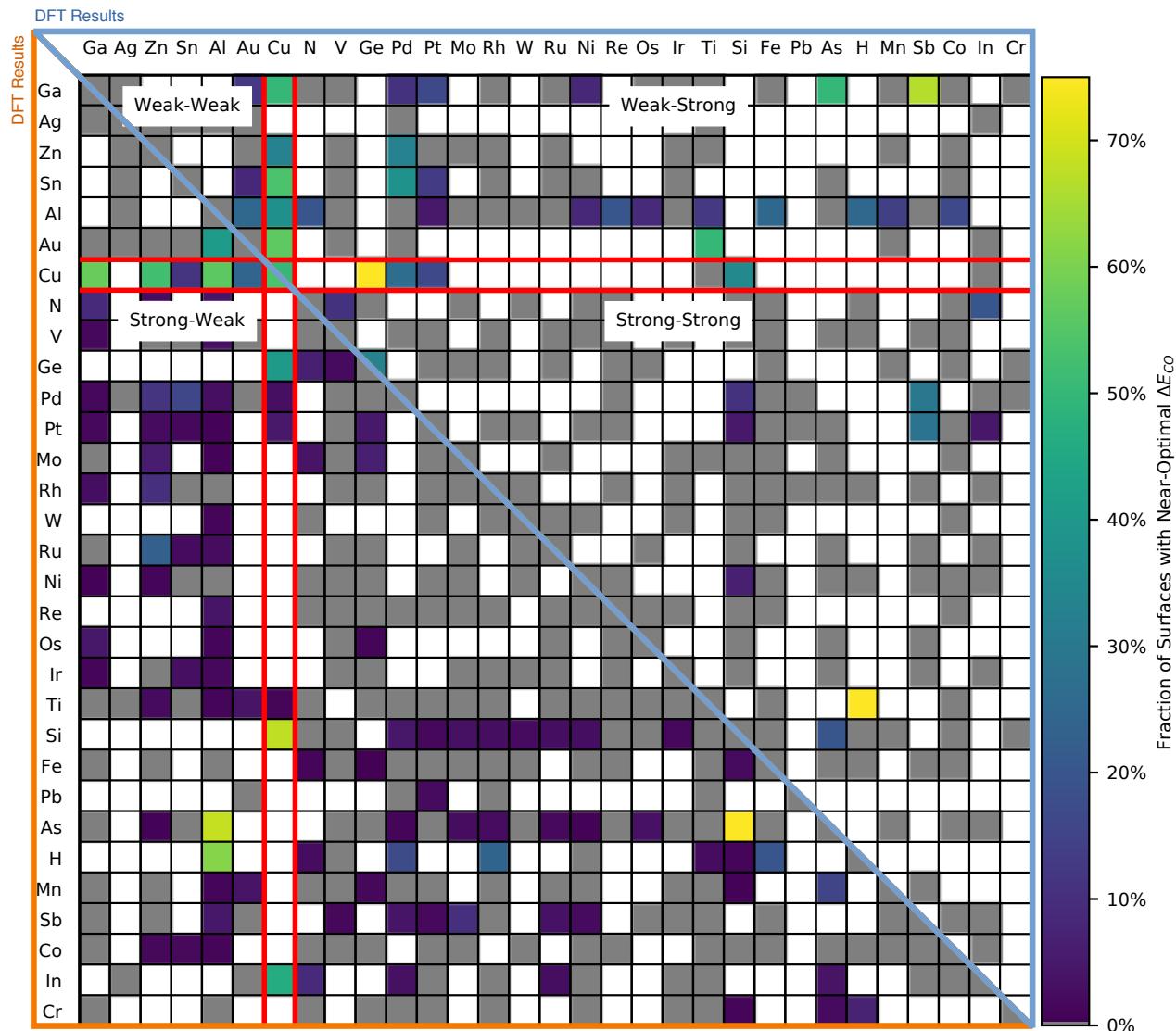
This framework discovered 130 different intermetallic facets with near-optimal  $\Delta E_{CO}$  as confirmed by DFT (Figure 3, Supplementary Table 1). These surfaces correspond to 54 different intermetallic combinations and are recommended for experimental verification of activity. Some of these intermetallics have already been investigated. For example, Cu/Sn blends have been shown to reduce CO<sub>2</sub> to either CO or formate at high Faradaic efficiencies.<sup>37</sup> Ni/Ga intermetallics have been shown to be active for CO<sub>2</sub> reduction,<sup>22,38</sup> Pd/Au bimetallics have been shown to be active for CO<sub>2</sub> reduction to C1-C5 products,<sup>39</sup> and ML results in this study suggest that single Pd atoms surrounded by Au atoms may be the most likely bimetallic active site with  $\Delta E_{CO}$  ca. -0.8 eV, in contrast to the hypothesized Pd-rich Pd/Au site. Cu/Al bimetallics, which have not been previously studied, also show promising experimental results in current ongoing work whose results will published in due course.

## Discovering trends in CO<sub>2</sub> reduction

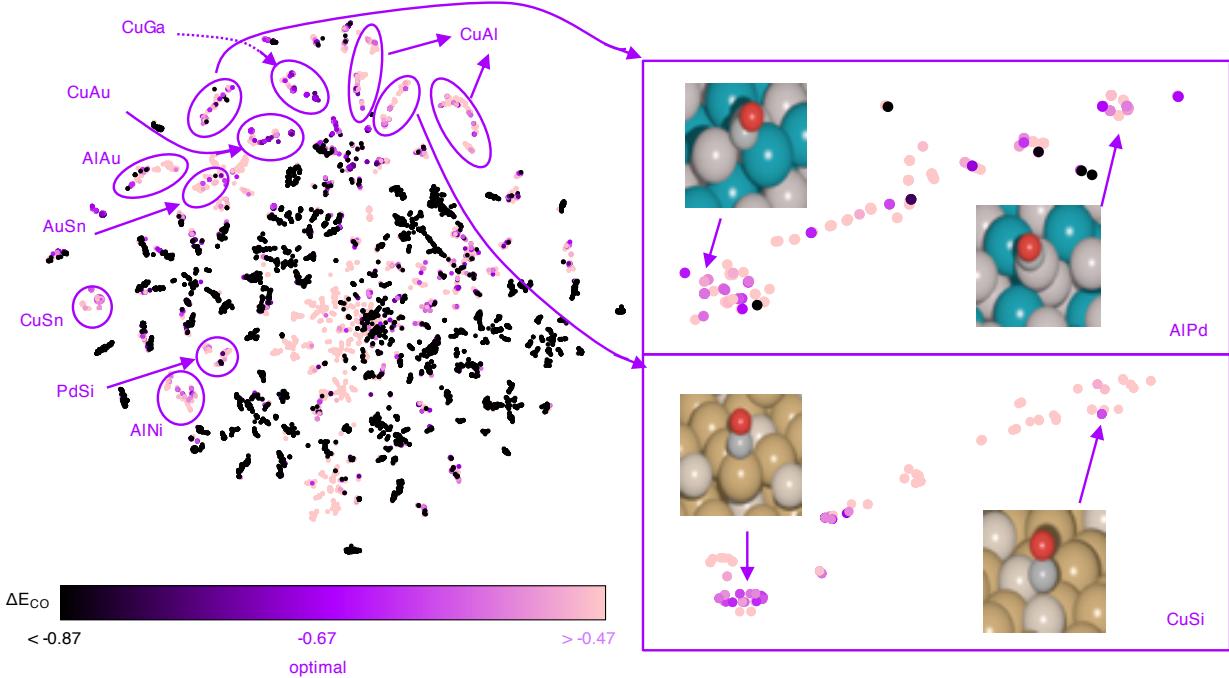
In addition to simply finding potential active surfaces, the data that this study generated can be used to gain insight into the chemistry of CO adsorption. Figure 4 illustrates the fraction of enumerated surfaces that have near-optimal values for  $\Delta E_{CO}$  for various bimetallic combinations. Elements in Figure 4 are rank ordered by the average pure-surface  $\Delta E_{CO}$  values as calculated by this study with DFT. Forming bimetallics with two elements that both have stronger binding than Ge generally led to inactive materials, but silicon broke this trend, suggesting that it has destabilizing properties when alloyed into an intermetallic.

Other trends can also be found when analyzing different elemental pairings. A number of strong-weak elemental pairings yield possibly active surfaces, including combinations of strong-binding elements like Pd, Pt, Ni, or Os with weak-binding elements like Al, Sn, Ga, or Sb, showing that the strong/weak Ni/Ga motif found in previous work is more general than previously known.<sup>22</sup> Interestingly, combining two weak-binding elements can lead to possibly active surfaces. For example: The strongest binding Ga surfaces and Au surfaces are approximately -0.44 eV and -0.53 eV respectively, based on a combination of DFT and ML predictions. However, a Ga-Ga bridge site on AuGa<sub>2</sub>(100) leads to a near-ideal binding energy of -0.57 eV.

Although the volume of data generated by this study is arguably intractable to study in detail, the size of the data enables certain methods of data analysis. For example: We are now able to assess potential intermetallic performance based solely on the number and distribution of potential active sites instead of the activity of one particular surface at one particular alloying ratio. Figure 5 illustrates this point by showing all 19,644 sites where we performed DFT calculations of  $\Delta E_{CO}$ . The X and Y axes in this figure are a reduced 2-dimensional feature space;<sup>40</sup> reference the Supplementary Methods for details. Clusters of points



**Figure 4 | CO<sub>2</sub> reduction activity map for bimetallics.** Visualization of two component intermetallics whose surfaces have low-coverage CO adsorption energy ( $\Delta E_{CO}$ ) values inside the range of [-0.77, -0.57] eV. White shading indicates an absence of any enumerated surfaces; grey shading indicates that all  $\Delta E_{CO}$  values are outside the range of [-0.77, -0.57] eV; and colored shading indicates possible activity. The  $\Delta E_{CO}$  values used to create the upper half of this figure were calculated by DFT, and the values used to create the bottom half were calculated by the surrogate ML model.



**Figure 5 | Active site motif analysis.** Latent space visualization using t-SNE<sup>40</sup> of all the adsorption sites simulated with DFT. Proximity in this reduced space indicates similarity in the structures of the adsorption sites. Black points represent sites that bind too strongly; dark purple points represent sites with optimal binding; and purple points indicate sites that bind too weakly. Stronger binding sites are overlaid on top of weaker binding sites to indicate that the stronger sites have a greater influence on activity than weaker sites due to their greater thermodynamic stability. We labeled dark purple clusters/materials, because we expect them to be better candidates for further investigation and experimentation.

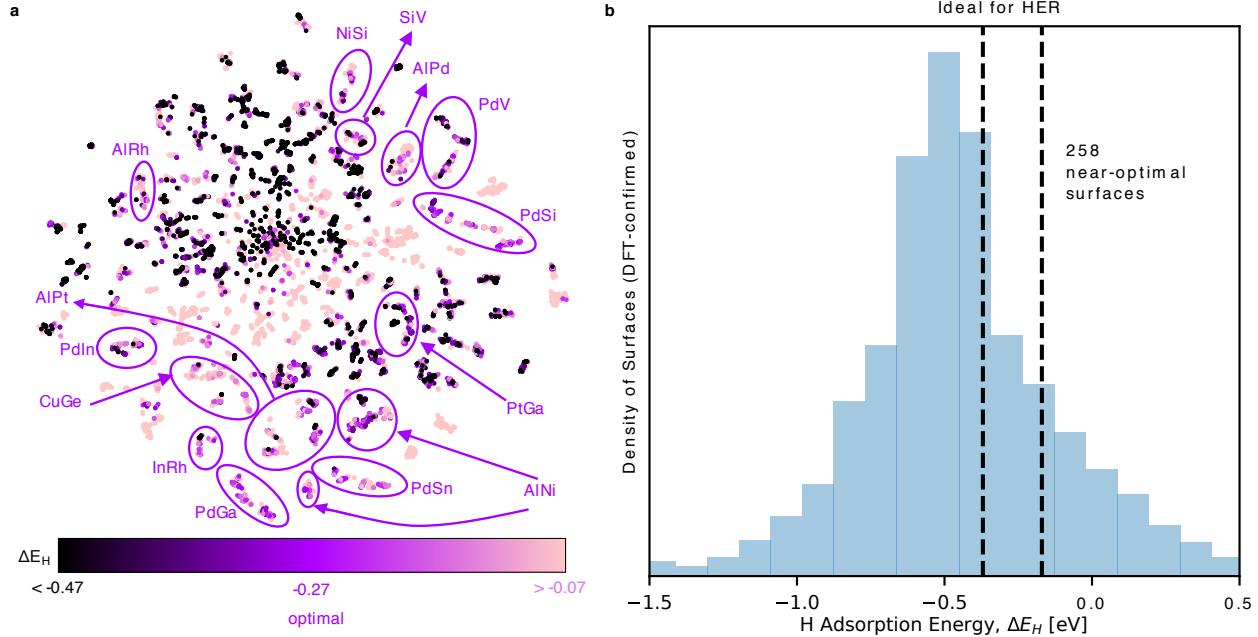
in this reduced space share similarities in site coordination and elemental combinations. Sites are colored by their  $\Delta E_{CO}$  so that regions of strong-binding (black), weak-binding (light purple) or near-ideal binding (dark purple) can be identified.

Clusters that are nearly uniformly dark purple are robust combinations are labeled in Figure 5. These are not the only possible active alloys; instead they represent combinations that are most likely to yield a higher fraction of active adsorption sites than other alloy combinations investigated thus far. This is especially important when matching theory with polycrystalline experiments where the precise active surface may not be known a priori or where there is little control over the surfaces created. Clusters that contain weak-binding sites alongside active binding sites may still be active as the CO will prefer the stronger binding, more active sites. Presence of strong binding sites are the more likely to hide an active site on a surface and should be avoided. Within a cluster, the embedding shows how active site coordination or alloying ratios may affect the activity. For example: The bottom-right panel in Figure 5 shows that Si sites in CuSi alloys tend to bind too weakly, thereby suggesting that higher ratios of Al to Si may improve activity.

The machine learning model in this work also provides activity estimates of surfaces without explicit DFT calculations. The model predicts that approximately 81% of surfaces have non-ideal  $\Delta E_{CO}$  values, defined as outside of the range  $[-0.8 - \text{MAE}, -0.5 + \text{MAE}]$  eV. This considerably narrows the potential experimental search space. Likewise, the search space for bimetallic combinations can be reduced. If at least one surface must be predicted to be near-optimal for the CO<sub>2</sub>RR, the search space can be reduced by 72%. If at least 10% of surfaces must be active (similar to the robust determination from above), then the search space is reduced by 93%.

## Discovering potential catalysts for H<sub>2</sub> evolution

The same types of analyses performed for CO<sub>2</sub>RR can also be performed for HER. Figure 6a illustrates the t-SNE representation of the 23,141 adsorption sites where we used DFT to calculate  $\Delta E_H$ . Figure 6b shows



**Figure 6 | Analysis of results for HER performance.** **a**, t-SNE<sup>40</sup> visualization of all the adsorption sites simulated with DFT. Similar to Figure 5, stronger binding sites are overlaid on top of weaker binding sites, and dark purple clusters/materials are labeled because we expect them to be better candidates for further investigation and experimentation. **b**, Normalized distribution of low-coverage  $\Delta E_H$  values calculated by our DFT workflow. Dashed lines indicate the  $\pm 0.1$  eV range around the optimal  $\Delta E_H$  value of -0.27 eV.

the distribution of DFT-calculated  $\Delta E_H$  values, where we found 258 different surfaces with low-coverage  $\Delta E_H$  values within 0.1 eV of the optimal value of -0.27 eV. All of these surfaces are listed in Supplementary Table 2, and the bimetallic map of HER performance is shown in Supplementary Figure 3. Similar to our analysis for CO<sub>2</sub>RR, a number of the intermetallics that our screening study identified as having surfaces with near-optimal  $\Delta E_H$  values have already been verified by various literature studies.<sup>41–43</sup>

Supplementary Figure 3 shows that, in addition to Pt, there is a band of elements with comparable monometallic adsorption energies that tend to yield intermetallic surfaces with near-optimal  $\Delta E_H$  values: As, Al, Si, Sb, Rh, and Pd. Many of these elements also appear in the t-SNE diagram for HER as well (Figure 6a), suggesting that these elements warrant further study and experimentation.

## Conclusions

We created a framework that produces and stores DFT data continuously and without the need for user intervention. This framework combines task and calculation management software with active machine learning and surrogate-based optimization to enable the automated, systematic selection and execution of DFT calculations. The framework produced 42,785 adsorption energy calculations to identify 130 candidate surfaces across 54 intermetallics with potentially high CO<sub>2</sub> reduction activity and 258 candidate surfaces across 102 intermetallics for hydrogen evolution. A number of the candidate surfaces found here have already been validated by literature experiments,<sup>37–39,41–43</sup> suggesting that the unstudied candidates found in this screening warrant further study. The full list of potential surfaces is shown in Supplementary Tables 1 and 2, and shortened lists of candidate intermetallics are illustrated in Figures 5 and 6.

Our workflow for generating DFT data offers a combination of benefits that we have not yet seen in other literature frameworks. Our task and calculation management systems reduce the amount of time required to configure and process DFT calculations; our database of DFT results enables holistic analyses across numerous adsorption sites, surfaces, and material spaces; and our active machine learning/surrogate-based optimization workflow guides the discovery of candidate catalysts without the need for expert intuition. The flexibility of

the framework also allows for expert-assisted guidance, allowing us to use the high-throughput DFT workflow to study specific sites, surfaces, or systems if needed. The combination of flexibility, automation, and machine learning guidance accelerates the theoretical discovery and study of catalysts for CO<sub>2</sub> reduction, H<sub>2</sub> evolution, or any other chemistry with a descriptor-performance scaling relationship.

A shortfall of our workflow is its heavy reliance on descriptor-performance relationships, which are used to guide the active learning algorithms. For example: this method will have issues with predicting CO<sub>2</sub> reduction activity for materials/surfaces that yield reaction mechanisms where  $\Delta E_{\text{CO}}$  is independent of activity. Additionally, this method does not address other important aspects of catalyst performance, such as surface stability or catalyst cost. These issues are acceptable because this framework is used primarily as a tool to screen for candidate catalysts from a relatively large search space and to supplement experts' intuitions with machine-derived suggestions. Our framework does not replace robust theoretical and experimental studies; it accelerates them by reducing search spaces to more tractable sizes and focusing expensive studies to systems that are more likely to yield interesting results. Future work could still be done to address the issues of diverse reaction mechanisms or multiple aspects of catalyst performance.

## Methods

### Enumerating search space

For each of the 1,499 intermetallic crystals we obtained from the Materials Project,<sup>30</sup> we used pymatgen<sup>17</sup> to enumerate symmetrically distinct facets with Miller indices between -2 and 2. Many intermetallic facets contained asymmetric top/bottom surfaces, and in those cases both were analyzed as well as distinct surfaces arising from the absolute position of the surface cut. In total 1,499 crystal structures were considered resulting in 17,507 unique surfaces and 1,684,908 unique adsorption sites. Surfaces were enumerated using ideal structures from the Materials Project instead of relaxed structures. This can cause differences in the number of enumerated facets, but it allows the enumeration to be completed without DFT relaxations for every bulk structure.

### Addressing models with a variable number of features

One issue with our fingerprinting method is that it yields a variable number of features. For example, two vectors are needed to represent the first shell of a Cu-Al bridge site, but only one vector is needed to represent a Cu-Cu bridge. This issue can be addressed with zero-padding, but can be better modeled using a literature method<sup>32</sup> to make dummy features to replace features that are not populated naturally. To continue with the previous example, the first shell of the Cu-Al site would be represented by a vector of 4 numbers for the Cu element and 4 more numbers for the Al element, but a Cu-Cu site would be represented by 4 numbers for the Cu element and 4 dummy features. These dummy features are the average atomic number ( $\bar{Z}$ ), the average Pauling electronegativity ( $\bar{\chi}$ ), the average median adsorption energy of all elements we studied ( $\bar{\tilde{E}}$ ), and a CN value of zero. Using averages of valid feature values reduces the bias induced from these dummy features while the CN value of zero ensures that no valid configuration could be confounded with the dummy features.

### Calculating optimal adsorption energies

A descriptor/activity relationship<sup>10</sup> was used to predict catalyst activity and selectivity for CO<sub>2</sub>RR given a free energy change,  $\Delta G_{\text{CO}}$ . This relationship shows that a  $\Delta G_{\text{CO}}$  of -0.17 eV yields an optimal activity and selectivity, corresponding to a  $\Delta E_{\text{CO}}$  of -0.67 eV; reference the Supplementary Methods for more details. Similarly, literature relations were adopted to predict HER performance.<sup>12</sup> This relation predicted an optimal  $\Delta G_{\text{H}}$  of -0.03 eV corresponding to a target  $\Delta E_{\text{H}}$  of 0.27 eV.<sup>44</sup>

### Calculating adsorption energies with DFT

The adsorption energy calculation workflow used in this study mimicked typical computational chemistry methods for calculating adsorption energies for sites of interest in the catalog.<sup>11</sup> Crystal structures from

the Materials Project were relaxed using DFT. Relaxed crystal structures were used to generate facets of interest which were then relaxed with free surface atoms and fixed subsurface atoms. Finally the adsorbate was placed on the surface at the relevant site and a final relaxation was completed. Final relaxed structures and their energies were comparable to traditional expert-made structures, as shown in Supplementary Note 3.

We performed all DFT calculations using: the *Vienna Ab initio Simulation Package*<sup>45–48</sup> (*VASP*) implemented in *ASE*,<sup>49</sup> the RPBE functionals,<sup>50</sup> k-point grids of  $4 \times 4 \times 1$ ; an energy cutoff of 350 eV; and the default pseudopotentials supplied by *VASP* version 5.4. Bulk relaxations were performed with a  $10 \times 10 \times 10$  k-point grid and a 500 eV cutoff and only isotropic relaxation were allowed. Surfaces were replicated in the X/Y directions so that each cell vector was at least 4.5 Å. No spin magnetism or dispersion corrections were included. Slabs were replicated in the Z direction to a minimum of 7 Å and at least 20 Å of vacuum was included in between slabs. For some facets this led to slabs with large depth due to constraints in how the facet could be formed. Generally the bottom layers were fixed and defined as those atoms more than 3 Å from the top of the surface in the scaled Z direction. Adsorption energies were calculated relative to gas-phase CO(g) for CO, and relative to gas-phase  $\frac{1}{2}$ H<sub>2</sub>(g) for H.

## Automating DFT calculations

DFT calculations and other calculational tasks were coordinated in parallel and in an automated, high-throughput fashion. Each type of calculation and task was encoded as an interdependent task, and then dependency management software (Luigi<sup>51</sup>) was used to manage the tasks in parallel. For example, an adsorption energy calculation depends on a single surface relaxation which depends on a single bulk relaxation. Requesting an adsorption energy calculation automatically triggers the prerequisite bulk and slab relaxations and then adds the results to a database. When a new adsorption energy calculation is triggered that requires the same surface, the prerequisite bulk and slab results are read from the database instead of being regenerated. This differs from a fixed pipeline approach,<sup>52</sup> because intermediate tasks such as slab relaxations can be shared across multiple pipelines and at different times. DFT tasks were managed by a central FireWorks<sup>19</sup> database that distributed DFT relaxation tasks across multiple computing clusters. This combined Luigi and FireWorks framework enabled high-throughput DFT calculations, because adsorption energies could be queried for any of the 1,684,908 enumerated sites without the need for human management of the intermediate tasks. All DFT relaxations were stored in a Mongo database that contains DFT calculation settings, the identity of the original crystal structure, the Miller indices of the slab, the exact, Cartesian location of the adsorption site, chemical information about the adsorption site such as local coordination, and the adsorption energy.

Our simulations occasionally yielded abnormal relaxations arising from desorptions, dissociations, surface reconstructions, or DFT non-convergence. These abnormalities were omitted from our regressions by excluding data from simulations that met any of the following criteria: simulations whose final maximum equilibrium force between any two atoms exceeded 0.5 eV/Å, where the absolute value of the adsorption energy exceeded 4 eV, where any atom moved more than 0.5 Å during bare slab relaxation, where the adsorbate moved more than 1.5 Å during adsorption relaxation, and where any slab atom moved more than 1.5 Å during adsorption relaxation. These exclusion criteria were used as heuristics to reduce outliers. This approach may induce bias in the dataset if systematic portions of the search space are missing because they often fail for these reasons. Some DFT errors may be treated by automatically tuning the DFT calculation settings,<sup>17</sup> but these approaches are not robust across the full range of calculation errors in adsorption simulations.

The adsorption energy database, which was required to train machine learning models, was initially seeded with  $\Delta E_{\text{CO}}$  and  $\Delta E_{\text{H}}$  calculations for every unique adsorption site on a variety of surfaces, including the (100), (111), and (211) facets of the most stable crystal form of each element included in Supplementary Figure 1. We also added all unique coordinations types (on-top Ni, on-top Fe, bridge Ni-Fe, etc.) up to a coordination number of two and then selected the corresponding surfaces with the smallest number of atoms. This ensured that the original machine learning model contained at least some data to begin with.

## Code availability

The code used to perform this work is available at “<https://github.com/ulissigroup/GASPy>”.

## Data availability

The code and data used to produce the figures in this article are available in the GASpy\_manuscript repository at “[https://github.com/ulissigroup/GASpy\\_manuscript](https://github.com/ulissigroup/GASpy_manuscript)”. An up-to-date dataset can be viewed at “<http://sm1.cheme.cmu.edu/volcano>”. Up-to-date versions of the data are also available in JSON format from the corresponding author on reasonable request.

## References

- [1] World Energy Outlook 2017. Tech. Rep., International Energy Agency (2017). URL <http://www.iea.org/weo2017/>.
- [2] Annual Energy Outlook 2017 with projections to 2050. Tech. Rep., U.S. Energy Information Administration (2017). URL [https://www.eia.gov/outlooks/aeo/pdf/0383\(2017\).pdf](https://www.eia.gov/outlooks/aeo/pdf/0383(2017).pdf).
- [3] Mackay, D.J.C. *Sustainable Energy—without the hot air*, vol. 2 (UIT Cambridge Ltd., 2009).
- [4] Edenhofer, O., Madruga, R.P. & Sokona, Y. *Renewable Energy Sources and Climate Change Mitigation (Special Report of the Intergovernmental Panel on Climate Change)* (Cambridge University Press, 2012), 1 edn.
- [5] Rockström, J. *et al.* A safe operating space for humanity. *Nature* **461**, 472–475 (2009).
- [6] *Fifth Assessment Synthesis Report* (Intergovernmental Panel on Climate Change, 2014). URL [https://www.ipcc.ch/pdf/assessment-report/ar5/syr/SYR\\_AR5\\_FINAL\\_full\\_wcover.pdf](https://www.ipcc.ch/pdf/assessment-report/ar5/syr/SYR_AR5_FINAL_full_wcover.pdf). arXiv:1011.1669v3.
- [7] Lewis, N.S. & Nocera, D.G. Powering the planet: Chemical challenges in solar energy utilization. *Proceedings of the National Academy of Sciences* **104**, 15729–15735 (2007).
- [8] Seh, Z.W. *et al.* Combining theory and experiment in electrocatalysis: Insights into materials design. *Science* **355** (2017).
- [9] Montoya, J.H. *et al.* Materials for solar fuels and chemicals. *Nature Materials* **16**, 70–81 (2016).
- [10] Liu, X. *et al.* Understanding trends in electrochemical carbon dioxide reduction rates. *Nature Communications* **8** (2017).
- [11] Nørskov, J.K., Studt, F., Abild-Pedersen, F. & Bligaard, T. *Fundamental Concepts in Heterogeneous Catalysis* (John Wiley & Sons, Inc., 2015).
- [12] Greeley, J., Jaramillo, T.F., Bonde, J., Chorkendorff, I. & Nørskov, J.K. Computational high-throughput screening of electrocatalytic materials for hydrogen evolution. *Nature Materials* **5**, 909–913 (2006).
- [13] Hansen, H.A., Shi, C., Lausche, A.C., Peterson, A.A. & Nørskov, J.K. Bifunctional alloys for the electroreduction of CO<sub>2</sub> and CO. *Physical Chemistry Chemical Physics* **18**, 9194–9201 (2016).
- [14] Li, Z., Wang, S., Chin, W.S., Achenie, L.E. & Xin, H. High-throughput screening of bimetallic catalysts enabled by machine learning. *Journal of Materials Chemistry A* **5**, 24131–24138 (2017).
- [15] Hummelshøj, J.S., Abild-Pedersen, F., Studt, F., Bligaard, T. & Nørskov, J.K. CatApp: A web application for surface chemistry and heterogeneous catalysis. *Angewandte Chemie - International Edition* **51**, 272–274 (2012).
- [16] Scheffler, M. & Draxl, C. Computer Center of the Max-Planck Society, Garching, The NoMaD Repository (2014).
- [17] Ong, S.P. *et al.* Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science* **68**, 314–319 (2013).
- [18] Montoya, J.H. & Persson, K.A. A high-throughput framework for determining adsorption energies on solid surfaces. *Nature Partner Journals Computational Materials* **3**, 14 (2017).
- [19] Jain, A. *et al.* FireWorks: a dynamic workflow system designed for high- throughput applications. *Concurrency Computation Practice and Experience* **22**, 685–701 (2010).

- [20] Meredig, B. *et al.* Combinatorial screening for new materials in unconstrained composition space with machine learning. *Physical Review B - Condensed Matter and Materials Physics* **89**, 1–7 (2014).
- [21] Ward, L. *et al.* Including crystal structure attributes in machine learning models of formation energies via Voronoi tessellations. *Physical Review B* **96**, 1–12 (2017).
- [22] Ulissi, Z.W. *et al.* Machine-Learning Methods Enable Exhaustive Searches for Active Bimetallic Facets and Reveal Active Site Motifs for CO<sub>2</sub> Reduction. *ACS Catalysis* **7**, 6600–6608 (2017).
- [23] Peterson, A.A. Acceleration of saddle-point searches with machine learning. *Journal of Chemical Physics* **145** (2016).
- [24] Boes, J.R. & Kitchin, J.R. Modeling Segregation on AuPd(111) Surfaces with Density Functional Theory and Monte Carlo Simulations. *Journal of Physical Chemistry C* **121**, 3479–3487 (2017).
- [25] Han, Z.H. & Zhang, K.S. Surrogate-Based Optimization. In Roeva, O. (ed.) *Real-World Applications of Genetic Algorithms*, chap. 17 (InTech, 2012). 1605.09522.
- [26] Settles, B. *Active Learning* (Morgan & Claypool, 2012).
- [27] Gómez-Bombarelli, R. *et al.* Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach. *Nature Materials* **15**, 1120–1127 (2016).
- [28] Warmuth, M.K. *et al.* Active learning with support vector machines in the drug discovery process. *Journal of Chemical Information and Computer Sciences* **43**, 667–673 (2003).
- [29] Gubaev, K., Podryabinkin, E.V. & Shapeev, A.V. Machine learning of molecular properties: Locality and active learning. *Journal of Chemical Physics* **148**, 1–9 (2018).
- [30] Jain, A. *et al.* The materials project: A materials genome approach to accelerating materials innovation. *APL Materials* **1**, 1–11 (2013).
- [31] Lukasz, M. Mendeleev (2014). URL <https://bitbucket.org/lukaszmentel/mendeleev>.
- [32] Davie, S.J., Di Pasquale, N. & Popelier, P.L. Kriging atomic properties with a variable number of inputs. *Journal of Chemical Physics* **145**, 1–11 (2016).
- [33] Calle-Vallejo, F., Loffreda, D., Koper, M.T.M. & Sautet, P. Introducing structural sensitivity into adsorption-energy scaling relations by means of coordination numbers. *Nature Chemistry* **7**, 403–410 (2015).
- [34] Zhang, Y. & Ling, C. A strategy to apply machine learning to small datasets in materials science. *npj Computational Materials* **25**, 28–33 (2018).
- [35] Olson, R.S. *et al.* Automating Biomedical Data Science Through Tree-Based Pipeline Optimization. In *European Conference on the Applications of Evolutionary Computation*, 123–137 (Springer International Publishing, Porto, Portugal, 2016).
- [36] Hyndman, R.J. & Athanasopoulos, G. *Forecasting: Principles and practice* (otexts.com, 2014), 1 edn. URL <http://otexts.com/fpp>.
- [37] Morimoto, M. *et al.* Electrodeposited Cu-Sn Alloy for Electrochemical CO<sub>2</sub> Reduction to CO/HCOO-. *Electrocatalysis* 1–10 (2017).
- [38] Torelli, D.A. *et al.* Nickel-Gallium-Catalyzed Electrochemical Reduction of CO<sub>2</sub> to Highly Reduced Products at Low Overpotentials. *ACS Catalysis* **6**, 2100–2104 (2016).
- [39] Kortlever, R. *et al.* Palladium-gold catalyst for the electrochemical reduction of CO<sub>2</sub> to C<sub>2</sub>-C<sub>5</sub> hydrocarbons. *Chemical Communications* **52**, 10229–10232 (2016).
- [40] Maaten, L.V.D. Accelerating t-SNE using Tree-Based Algorithms. *Journal of Machine Learning Research* **15**, 1–21 (2014). 1307.1662.

- [41] Cherepanov, P.V., Ashokkumar, M. & Andreeva, D.V. Ultrasound assisted formation of Al-Ni electrocatalyst for hydrogen evolution. *Ultrasonics Sonochemistry* **23**, 142–147 (2015).
- [42] Yamauchi, M., Abe, R., Tsukuda, T., Kato, K. & Takata, M. Highly selective ammonia synthesis from nitrate with photocatalytically generated hydrogen on CuPd/TiO<sub>2</sub>. *Journal of the American Chemical Society* **133**, 1150–1152 (2011).
- [43] Liao, H. *et al.* A Multisite Strategy for Enhancing the Hydrogen Evolution Reaction on a Nano-Pd Surface in Alkaline Media. *Advanced Energy Materials* **7**, 1–7 (2017).
- [44] Nørskov, J.K. *et al.* Trends in the Exchange Current for Hydrogen Evolution. *Journal of The Electrochemical Society* **152**, J23 (2005).
- [45] Kresse, G. & Hafner, J. Ab initio molecular dynamics for liquid metals. *Physical Review B* **47**, 558–561 (1993). 0927-0256(96)00008.
- [46] Kresse, G. & Hafner, J. Ab initio molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium. *Physical Review B* **49**, 14251–14269 (1994).
- [47] Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational Materials Science* **6**, 15–50 (1996).
- [48] Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical Review B* **54**, 11169–11186 (1996).
- [49] Hjorth Larsen, A. *et al.* The atomic simulation environment—a Python library for working with atoms. *Journal of Physics: Condensed Matter* **29**, 273002 (2017).
- [50] Hammer, B., Hansen, L.B. & Nørskov, J. Improved adsorption energetics within density-functional theory using revised Perdew-Burke-Ernzerhof functionals. *Physical Review B* **59**, 7413–7421 (1999).
- [51] Bernhardsson, E., Freider, E. & Rouhani, A. Luigi, a Python package that builds complex pipelines of batch jobs (2012). URL <https://github.com/spotify/luigi>.
- [52] Mathew, K. *et al.* Atomate: A high-level interface to generate, execute, and analyze computational materials science workflows. *Computational Materials Science* **139**, 140–152 (2017).

## Acknowledgements

This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. We thank Karen Chan for helpful discussions about descriptor targets, as well as Phil de Luna and Edward T. Sargent for helpful discussions about analysis.

## Author Contributions

K.T. and Z.W.U. contributed to the scientific workflow software and DFT calculations. K.T. and Z.W.U. made the regression models and analysis. K.T. performed the clustering analysis. K.T. and Z.W.U. wrote the manuscript. Z.W.U. conceived the idea.

## Competing Interests

The authors declare no competing interests.