Digital Discovery

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/d5dd00298b

Adsorb-Agent: autonomous identification of stable adsorption configurations *via* a large language model agent

Janghoon Ock,[ab] Radheesh Sharma Meda,[c] Tirtha Vinchurkar,[a] Yayati Jadhav[c] and Amir Barati Farimani[*c]

Adsorption energy is a key reactivity descriptor in catalysis. Determining adsorption energy requires evaluating numerous adsorbate–catalyst configurations, making it computationally intensive. Current methods rely on exhaustive sampling, which must navigate a large search space without guaranteeing the identification of the global minimum energy. To address this, we introduce Adsorb-Agent, a Large Language Model (LLM) agent designed to efficiently identify stable adsorption configurations corresponding to the global minimum energy. Adsorb-Agent leverages its built-in knowledge and reasoning to strategically explore configurations, significantly reducing the number of initial configurations required while improving the energy prediction accuracy. In this study, we also evaluated the performance of different LLMs—GPT-4o, GPT-4o-mini, Claude-3.7-Sonnet, and DeepSeek-Chat—as the reasoning engine for Adsorb-Agent, with GPT-4o showing the strongest overall performance. Tested on twenty diverse systems, Adsorb-Agent identifies comparable adsorption energies for 84% of cases and achieves lower energies for 35%, particularly excelling in complex systems. It identifies lower energies in 47% of intermetallic systems and 67% of systems with large adsorbates. These findings demonstrate Adsorb-Agent's potential to accelerate catalyst discovery by reducing computational costs and enhancing prediction reliability compared to exhaustive search methods.

Received 8th July 2025
Accepted 11th December 2025

DOI: 10.1039/d5dd00298b

rsc.li/digitaldiscovery

Introduction

The design of optimal catalyst materials for targeted reaction processes plays an essential role in advancing chemical processes.[1–3] In particular, addressing the dual challenge of meeting the growing global energy demand while combating climate change necessitates the development of efficient, low-cost catalysts to enable the broader use of renewable energy sources.[3] Traditionally, the search for optimal catalysts has relied on either labor-intensive experimental methods or computationally expensive quantum chemistry calculations. Because of the vast material design space, screening strategies often focus on identifying key descriptors that effectively predict catalytic performance.

Adsorption energy, defined as the change in energy upon the binding of a molecule to a catalytic surface, is one of the most widely used descriptors in computational catalysis due to its direct correlation with catalytic reactivity.[4,5] The adsorption energy, corresponding to the most stable adsorption configuration, serves as a key descriptor of catalyst performance and plays a crucial role in estimating the reactivity of various catalysts.[4,6–8] Furthermore, adsorption energy is a foundational parameter in constructing free energy diagrams, which are used to identify the energetically preferred reaction pathways on catalyst surfaces.

The adsorption energy, ΔE_{ads}, is mathematically defined as the global minimum energy among all possible adsorption configurations.[9,10] It is calculated as the difference between the total energy of the adsorbate–catalyst system ($E_{sys,i}$), the energy of the clean surface (slab) (E_{slab}), and the energy of the gas-phase adsorbate or reference species (E_{gas}):

$$\Delta E_i = E_{sys,i} - E_{slab} - E_{gas} \quad (1a)$$

$$\Delta E_{ads} = \min_i(\Delta E_i) \quad (1b)$$

Accurately determining adsorption energy presents significant challenges. The complex electron-level interactions that govern chemical bonding make it impractical to predict the most stable configuration based solely on atomic-level information. As a result, determining the global minimum adsorption energy typically requires enumerating and evaluating a vast number of possible configurations.[10–12] This process becomes computationally prohibitive when using quantum chemistry methods such as density functional theory (DFT).[10,13] The difficulty is

[a]Department of Chemical Engineering, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213, USA. E-mail: barati@cmu.edu
[b]Department of Chemical and Biomolecular Engineering, University of Nebraska–Lincoln, Lincoln, NE 68588, USA
[c]Department of Mechanical Engineering, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213, USA

further compounded by the combinatorial explosion of potential binding sites, variations in surface geometries, and diverse orientations the adsorbate can adopt. Despite exhaustive configuration searches, there is no guarantee of reliably identifying the true global minimum energy configuration. These challenges underscore the need for more efficient and accurate approaches to streamline adsorption energy determination, paving the way for faster and more reliable catalyst design.

Recent advances in machine learning (ML) have introduced promising alternatives to conventional quantum chemistry methods, significantly improving the efficiency of adsorption energy prediction tasks. In particular, Graph Neural Networks (GNNs) have demonstrated exceptional performance in predicting energy and forces for atomic systems. For adsorbate–catalyst systems,[14] GNNs achieve a high level of precision, predicting adsorption energies with an error of approximately 0.2 eV and forces with an error of 0.013 eV $Å^{-1}$.[15,16] These capabilities make GNNs effective surrogates for DFT calculations in tasks such as geometry optimization and energy prediction. Building on this foundation, Lan *et al.* introduced the AdsorbML method, which achieved a 2000× speedup in adsorption energy determination while retaining 87.36% of the accuracy of full DFT calculations by integrating GNNs with DFT.[10] In their approach, GNNs are used to relax structures from initial adsorption configurations, after which DFT is employed for further relaxation or single-point energy calculations to obtain DFT-validated adsorption energies. Despite this progress, the placement of adsorbates on the surface and the sampling of adsorption sites remain reliant on exhaustive enumeration, posing a significant challenge in reducing the initial search space and improving the computational efficiency.

Moving forward, large language models (LLMs) are emerging as transformative tools in scientific problem-solving by leveraging their pre-trained knowledge and planning capabilities, as well as their seamless human–machine interaction.[17] LLM agents, powered by LLMs, have demonstrated remarkable potential in reshaping scientific workflows. For example, Boiko *et al.* proposed Coscientist, an LLM agent that automates experimental design and execution, significantly enhancing productivity while reducing manual effort.[18] Similarly, Szymanski *et al.* demonstrated the application of LLMs in an autonomous laboratory framework for proposing synthesis recipes.[19] In the field of additive manufacturing, Jadhav *et al.* introduced the LLM-3D Print framework, which streamlines the design-to-manufacturing process by autonomously generating, validating, and optimizing 3D printing instructions with an LLM agent at its core.[20] These breakthroughs demonstrate how LLMs are reshaping the landscape of scientific discovery and innovation.

In this study, we introduce Adsorb-Agent, an LLM-based agent designed to determine adsorption energy efficiently. Adsorb-Agent predicts initial adsorption configurations that are likely to be close to the most stable configuration and relaxes them to identify the minimum energy state. While human researchers can propose plausible stable adsorption configurations for specific systems based on domain knowledge—such as chemical bonding and surface science—it remains exceedingly difficult to derive a universal theorem for predicting the most stable configuration across diverse adsorbate–catalyst systems. Moreover, in high-throughput screening, where millions of candidate systems must be evaluated, it is infeasible to manually propose stable configurations for each individual system.[21] Adsorb-Agent addresses these challenges by autonomously deriving stable configurations, relying solely on the LLM's built-in knowledge and emergent reasoning capabilities. Because it operates purely through inference from pre-trained models, it is readily applicable to large-scale tasks.

This study has two primary objectives: first, to reduce the computational cost of adsorption energy identification by minimizing the number of initial configurations required; second, to enhance the accuracy of adsorption energy predictions by generating refined initial configurations that are closer to the global minimum while maintaining human interpretability. By bridging state-of-the-art LLM capabilities with catalytic configuration challenges, Adsorb-Agent represents a significant step toward broader adoption of AI-driven methods in materials science and catalysis, accelerating the discovery and design of optimal catalysts. Furthermore, when integrated with other LLM models and tools for catalyst design,[13,22,23] Adsorb-Agent can be extended to a wider range of applications in optimal catalyst development.

Agentic framework

Workflow overview

Adsorb-Agent is an LLM-powered agent designed to identify the most stable adsorption configuration and its corresponding adsorption energy. It consists of three core LLM modules—the Solution Planner, Critic, and Binding Indexer—all powered by LLMs, as illustrated in Fig. 1a. The process begins with a user query that specifies the adsorbate's SMILES, the catalyst's chemical symbol, and the surface orientation. The core functionality of Adsorb-Agent is to narrow the configuration search space by identifying promising adsorption candidates for the specified adsorbate–catalyst system. The agent retrieves the adsorbate molecule and catalyst surface structure from a database, places the adsorbate according to the predicted configuration, and then relaxes the system using designated computational tools to calculate the adsorption energy. By comparing the relaxed energies, the agent identifies the configuration with the lowest energy, which serves as a key reactivity descriptor in catalysis.

To reduce hallucination and improve reliability, we adopt two safeguards. First, instead of using the full free-form LLM output, we extract only four essential descriptors, such as site type, surface binding atoms, adsorbate orientation, and adsorbate binding atoms. They are the minimal elements needed to define a configuration. This limits sensitivity to wording or irrelevant reasoning. Second, a Critic module checks the internal consistency of the proposed configuration and filters out contradictory outputs. These measures ensure that only coherent and well-defined configurations proceed to the relaxation stage.

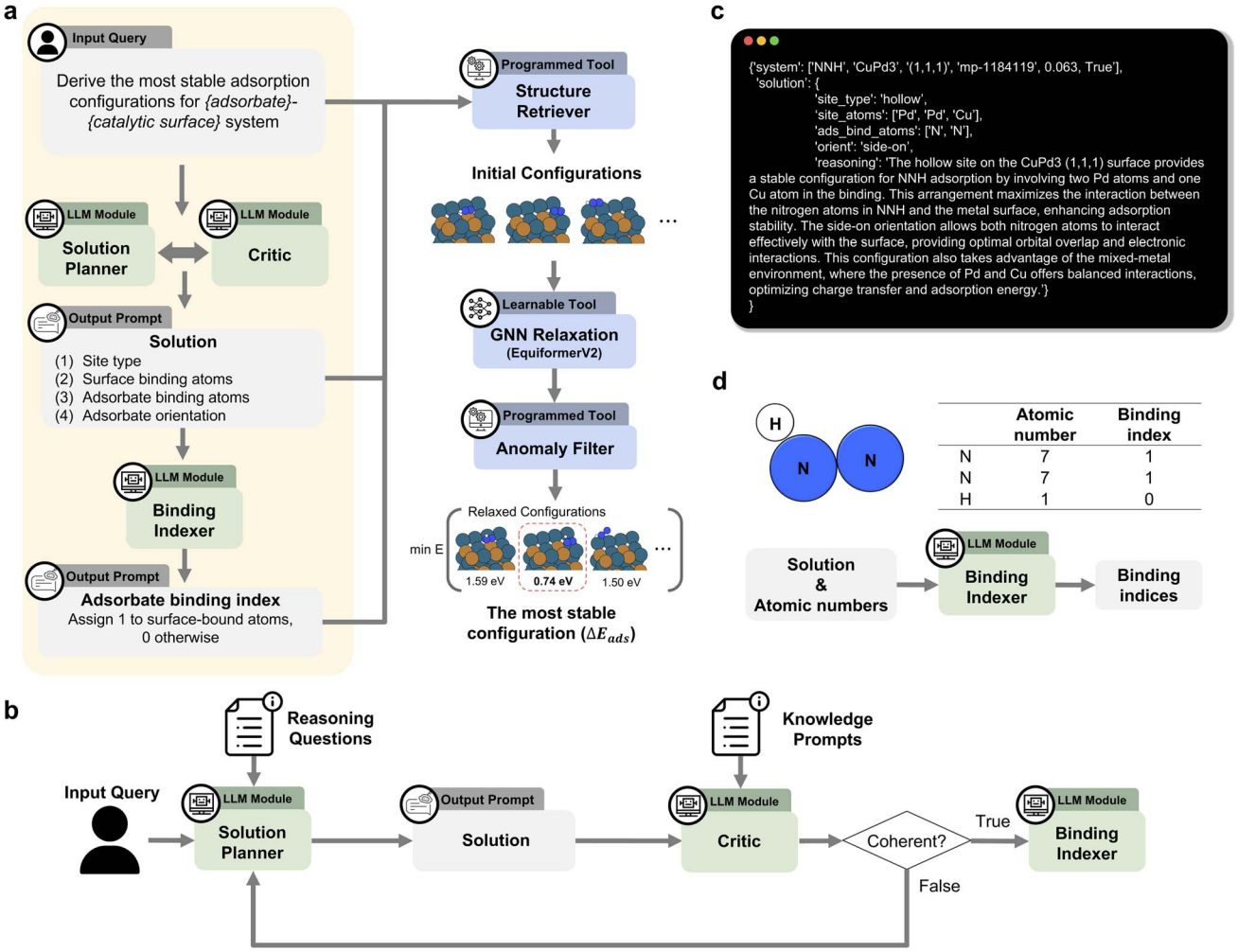

Fig. 1 Adsorb-Agent framework. (a) Overall process for identifying the most stable configuration by combining the LLM modules and tools. The left yellow box represents the part where the LLM modules are involved, comprising three components—Solution Planner, Critic, and Binding Indexer. These modules are powered by LLM APIs such as GPT-4o, GPT-4o-mini, Claude-3.7-Sonnet, and DeepSeek-Chat. (b) Illustration of the iterative interaction between the Solution Planner and Critic modules. (c) Example solution for the NNH–$CuPd_3$ (1, 1, 1) system. The system metadata comprise the following in order: the SMILES representation of the adsorbate molecule, the chemical composition of the bulk catalyst, the Miller index, the Materials Project ID, the shift, and the top. (d) Example of binding index derivation from the Binding Indexer module.

Initial configuration candidates

Based on the natural language input query, the Solution Planner derives the description for the most probable stable adsorption configuration. This process is primarily conducted based on the LLM's built-in knowledge and reasoning capability. The solution output includes four pieces of information: the type of adsorption site, the binding atoms on both the surface and the adsorbate, and the orientation of the adsorbate. This derivation is guided by a structured list of reasoning questions provided by the user, as shown in Fig. 1b. These questions reflect the typical thought process of human researchers—for example, "Are the bonds between the adsorbate and surface strong enough to ensure stability?"—while deliberately excluding system-specific knowledge to ensure general applicability (see SI Fig. S1). The Solution Planner can also provide reasoning statements that explain how it arrived at its prediction, as illustrated in Fig. 1c. A full set of outputs for 20 systems is included in the SI Sections S3 and S4. The effects of internal and external prompts are discussed in SI Sections S5 and S6.

To ensure the logical coherence of the solution output, the Critic module evaluates the initial solution generated by the Solution Planner. It takes the solution as input and uses a knowledge prompt that clarifies the terms within the solution to guide its review (see Fig. 1b). Details of the knowledge prompt are provided in SI Fig. S1. The module focuses on two aspects: (1) the coherence between the adsorption site type and the binding atoms on the surface and (2) the alignment between the adsorbate's binding atoms and its orientation. For instance, if the solution specifies a bridge site, it must involve two binding surface atoms. Similarly, for adsorbate orientation, if the adsorbate is described as end-on, it should have only one binding adsorbate atom. If any incoherence is identified, the Critic rebuts the solution, and the Solution Planner is re-initialized to produce a revised solution. This iterative

interaction ensures the final adsorption configuration is logically coherent and self-consistent.

Once a coherent solution is generated, the Binding Indexer module assigns indices to the adsorbate's binding atoms based on the solution. This step translates the human-readable configuration into a numerical format suitable for computational processing. As illustrated in Fig. 1d, the Binding Indexer takes the identified binding atoms and orientation from the solution, along with the adsorbate's atomic number array, to generate a binding index array. This array specifies which atoms in the adsorbate are involved in binding to the surface. Using this array, the adsorbate can be positioned on the catalytic surface, reflecting both the surface orientation and binding atom information. This automation removes the need for manually pre-defining binding atoms—commonly required in datasets like the Open Catalyst Project, which explicitly mark binding atoms with asterisks (*e.g.*, NNH). In addition, we introduce a new placement strategy capable of handling side-on adsorbates (see the Methods section).

Energy determination

The following steps are carried out without the involvement of the LLM modules, but using the pre-written computation scripts. The catalytic surface and adsorbate molecule are retrieved using the Open Catalyst Project demo API, based on the adsorbate SMILES, catalyst bulk composition, and Miller indices provided in the user query. Details about the Open Catalyst Project demo are provided in the Methods section. The adsorbate is then placed onto the catalytic surface to generate initial adsorbate–catalyst structures, guided by the predicted configuration and the binding index array. If the LLM proposes a physically unrealistic configuration, such cases are automatically filtered out during this structure generation stage handled by the structure retriever. As illustrated in the atomic visualization in Fig. 1, multiple initial configurations remain possible; however, their number is significantly reduced compared to conventional configuration enumeration approaches.

These initial structures are subsequently relaxed to determine the minimum energy configuration, as relaxed energies are necessary for meaningful comparison. In this study, we employ a GNN model, specifically EquiformerV2 trained on the Open Catalyst 2020 dataset,[14,16] although other machine learning models or quantum chemistry simulations could also be used. Details of the GNN-based relaxation process are provided in the Methods section. During relaxation, even initially similar configurations can evolve into distinct final structures with different energies. Structures exhibiting anomalies such as extensive surface reconstruction, adsorbate dissociation, or desorption may occur, and any structures exhibiting these anomalies are filtered out.[10,24,25] Among the remaining valid configurations, the one with the lowest energy is identified as the most stable configuration. This energy is recognized as the adsorption energy, which serves as a critical reactivity descriptor for the adsorbate–catalyst combination.

Performance evaluation

Theoretically, the most stable adsorption configuration corresponds to the global minimum energy among all possible configurations. However, due to the vast configurational space, it is practically impossible to exhaustively identify the true global minimum. Conventional approaches instead rely on enumeration algorithms, such as `random` and `heuristic` algorithms, which sample numerous configurations to approximate

Table 1 Comparison of Adsorb-Agent and the algorithmic approach. The adsorption energy corresponds to the minimum energy among configurations. Results for individual runs are provided in SI Table S1

			Adsorption energy [eV]		Number of initial sets	
No.	Adsorbate	Catalyst	Adsorb-Agent (↓)	Algorithm	Adsorb-Agent (↓)	Algorithm
1	H	Mo_3Pd (111)	−0.764 ± 0.113	−0.941 ± 0.002	6.7 ± 2.1	59
2	NNH	Mo_3Pd (111)	−1.265 ± 0.158	−0.903 ± 0.117	9.3 ± 3.7	51
3	H	$CuPd_3$ (111)	−0.380 ± 0.003	−0.398 ± 0.017	16.7 ± 1.2	98
4	NNH	$CuPd_3$ (111)	0.745 ± 0.006	0.867 ± 0.072	17.3 ± 3.1	78
5	H	Cu_3Ag (111)	−0.019 ± 0.041	−0.072 ± 0.002	21.3 ± 4.1	98
6	NNH	Cu_3Ag (111)	1.504 ± 0.057	1.500 ± 0.002	16.7 ± 2.6	56
7	H	Ru_3Mo (111)	−0.587 ± 0.002	−0.586 ± 0.050	17.0 ± 2.2	94
8	NNH	Ru_3Mo (111)	−0.498 ± 0.013	−0.276 ± 0.003	18.7 ± 0.5	81
9	OH	Pt (111)	0.990 ± 0.001	0.990 ± 0.071	7.0 ± 1.6	54
10	OH	Pt (100)	0.991 ± 0.001	0.991 ± 0.001	10.3 ± 4.2	54
11	OH	Pd (111)	0.814 ± 0.000	0.814 ± 0.001	20.0 ± 5.7	54
12	OH	Au (111)	1.408 ± 0.002	1.409 ± 0.002	23.3 ± 5.2	54
13	OH	Ag (100)	0.440 ± 0.001	0.463 ± 0.009	23.7 ± 4.5	53
14	OH	CoPt (111)	−0.208 ± 0.015	−0.166 ± 0.046	41.3 ± 1.2	120
15	CH_2CH_2OH	Cu_6Ga_2 (100)	−2.338 ± 0.833	−3.077 ± 0.062	28.7 ± 15.5	66
16	CH_2CH_2OH	Au_2Hf (102)	−2.761 ± 0.592	−3.761 ± 0.129	28.0 ± 4.5	78
17	$OCHCH_3$	Rh_2Ti_2 (111)	−4.561 ± 0.007	−4.275 ± 0.086	29.0 ± 4.3	62
18	$OCHCH_3$	Al_3Zr (101)	−4.616 ± 0.014	−4.325 ± 0.052	22.0 ± 2.4	68
19	$OCHCH_3$	Hf_2Zn_6 (110)	−5.922 ± 0.209	−5.443 ± 0.037	18.0 ± 2.2	67
20	$ONN(CH_3)_2$	Bi_2Ti_6 (211)	−3.454 ± 0.337	−2.441 ± 0.103	33.0 ± 3.6	139

the global minimum adsorption energy. Details of these enumeration procedures are provided in the Methods section.

Adsorb-Agent's ability to identify the most stable configuration is evaluated against exhaustive enumeration algorithms. Performance is assessed based on three key criteria: (i) efficiency in reducing the configuration search space, (ii) accuracy in identifying adsorption configurations with energies comparable to those found by enumeration algorithms, and (iii) consistency of results across independent trials.

The evaluation is conducted on 20 adsorbate–catalyst systems selected for their practical importance, particularly in nitrogen production and fuel cell applications.[26–28] For example, the electrocatalytic nitrogen reduction reaction (NRR) offers a potential route for sustainable nitrogen fixation. However, its performance is often limited by the high activation energy required to cleave the inert N≡N bond and by strong competition from the hydrogen evolution reaction (HER).[29] Likewise, the oxygen reduction reaction (ORR) is a vital reaction process in fuel cell operation and holds a central position in the broader field of electrocatalysis.[28]

Eight of the selected systems are associated with the NRR and HER. These include four catalysts proposed by Zhou *et al.*,[29] each interacting with two key adsorbates: NNH for the NRR and H for the HER, resulting in a total of eight systems. An additional six systems are associated with the ORR, with OH as the key adsorbate, selected from the experimentally verified sets reported by Kulkarni *et al.*[28,30,31] To broaden the evaluation beyond simple adsorbates, six more systems featuring larger molecules—such as CH_2CH_2OH, $OCHCH_3$, and $ONN(CH_3)_2$—and intermetallic catalysts were included. These complex systems were randomly selected from the Open Catalyst 2020-Dense (OC20-Dense) dataset.[10] A complete list of adsorbate–catalyst systems is provided in Table 1, with detailed slab information available in SI Table S1.

Results and discussion

Search space reduction demonstration

Adsorb-Agent effectively reduces the configuration search space for further energy determination by specifying both the adsorption site and the binding atoms, thereby limiting the number of initial configurations to be evaluated. In search problems, beginning from an initial point proximal to the optimal solution is essential, as this can significantly reduce the search space and improve the convergence efficiency.[32] Fig. 2 illustrates three example cases: NNH–$CuPd_3$ (111), OH–Pt (111), and $OCHCH_3$–$HfZn_3$ (110), which represent NRR-related, ORR-related, and large adsorbate-containing systems, respectively. Among these, the OH–Pt (111) system serves as an example of a monometallic surface, which is more homogeneous compared to intermetallic surfaces. In atomic visualizations,

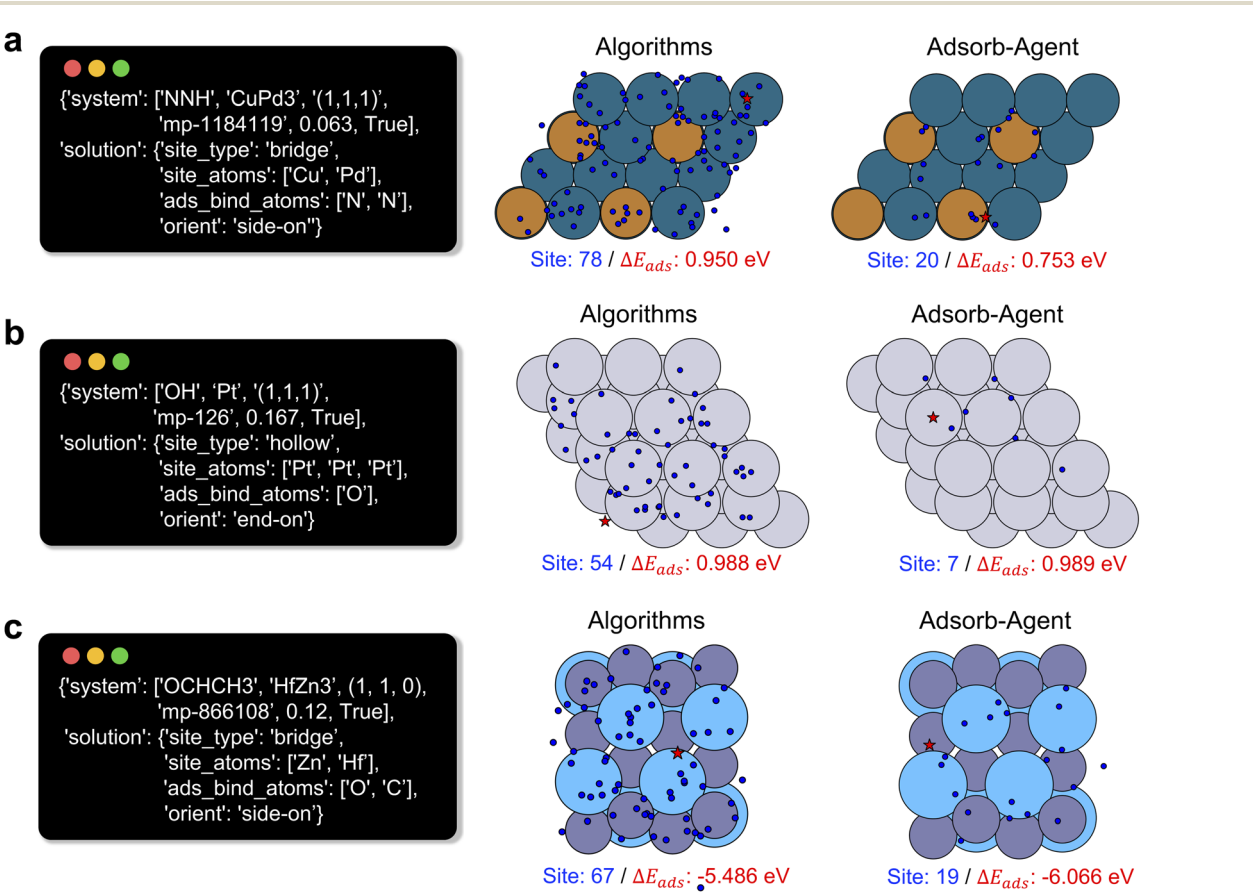

Fig. 2 Comparison of search space between the algorithmic approach and Adsorb-Agent. Left panels show solutions derived by Adsorb-Agent, while right panels present atomic visualizations of surfaces with initial adsorption sites for (a) NNH–$CuPd_3$ (111), (b) OH–Pt (111), and (c) $OCHCH_3$–$HfZn_3$ (110) (blue dots: initial sites; red stars: relaxed stable sites).

blue dots denote the initial adsorption sites, while red stars mark the relaxed adsorption sites corresponding to the most stable configurations.

Adsorb-Agent significantly reduces the number of initial configurations by focusing on likely adsorption sites, compared to the exhaustive enumeration algorithmic approaches. For example, Adsorb-Agent predicts the bridge site at the intersection of Cu and Pd atoms as the primary adsorption site for NNH–$CuPd_3$ (111). Similarly, it identifies the hollow site of the Pt (111) surface as the optimal adsorption site for OH and the bridge site between Zn and Hf atoms for the $OCHCH_3$–$HfZn_3$ (110) system. Occasional deviations from the predicted solutions arise due to the distance margin used to define targeted adsorption sites. By focusing on specific sites, Adsorb-Agent effectively reduces the search space.

For the NNH–$CuPd_3$ (111) and $OCHCH_3$–$HfZn_3$ (110) systems, the relaxed adsorption sites of the most stable configurations precisely match the solutions proposed by Adsorb-Agent. In both cases, Adsorb-Agent achieves adsorption energies lower than those obtained *via* algorithmic approaches. However, for the OH–Pt (111) system, the adsorption site of the most stable relaxed configuration identified by both the algorithmic methods and Adsorb-Agent is an ontop site, differing from the initial Adsorb-Agent prediction. Despite this discrepancy, the initial site suggested by Adsorb-Agent successfully guided the system to the most stable configuration during the relaxation process.

Additionally, the solution generated by Solution Planner includes reasoning prompts beyond the four essential pieces of information required for determining the initial configuration (see Fig. 1c). Although the adsorption site and surface-binding details suggested are not explicitly reported in prior literature, the general concept of side-on orientation preference is reasonable. The idea that a "side-on orientation allows both nitrogen atoms to interact effectively with the surface, providing optimal orbital overlap and electronic interactions" is plausible in general surface science contexts but lacks direct support in CuPd nitrate reduction. Orientation effects (tilt/side-on *vs.* end-on) have been studied for NO and NH_x species on metal (111) surfaces, where orbital overlap and back-donation govern adsorption.[33] The proposal that the mixed-metal Cu–Pd environment "offers balanced interactions, optimizing charge transfer and adsorption energy", is supported by multiple studies of Cu–Pd alloys and intermetallic catalysts. For example, ordered B2 CuPd nanocubes can break classical adsorption-energy scaling by stabilizing early oxy-nitrogen intermediates (*e.g.*, $*NO_3$) while weakening late nitrogen-containing fragments (*e.g.*, *N) *via* subsurface Pd electronic effects.[34,35] Similarly, Cu–M alloys (M = Pd, Zn, *etc.*) have been reported to tune intermediate binding energies and charge transfer in nitrate reduction reactions.[36]

Adsorption energy identification

The performance of the Adsorb-Agent is evaluated against conventional algorithmic approaches (`random` and `heuristic`), as summarized in Table 1. To quantify the effectiveness of Adsorb-Agent in identifying adsorption energies, three key metrics are defined, and their mathematical formulations are provided in the Methods section. Success Ratio (SR) assesses the ability of Adsorb-Agent to identify adsorption energies comparable to those found by the algorithmic approaches. Lower Energy Discovery Ratio (LEDR) measures the capability of Adsorb-Agent to discover adsorption energies lower than those identified by the algorithmic approaches. Reduced Search Space Ratio (RSR) quantifies the reduction in the number of initial configurations required by Adsorb-Agent compared to the algorithmic approaches.

These metrics provide a comprehensive framework for evaluating Adsorb-Agent's effectiveness in identifying the most stable configurations relative to conventional methods. Specifically, the RSR reflects the efficiency of the search process, while the SR and LEDR reflect the accuracy of energy identification. An increase in the number of initial configurations, as reflected by a higher RSR, often corresponds to improvements in SR and LEDR values. Therefore, all three metrics should be considered collectively to thoroughly assess Adsorb-Agent's performance in comparison to algorithmic approaches.

As shown in Fig. 3a, Adsorb-Agent successfully identifies adsorption energies comparable to those found by the algorithmic approach in 83.7% of cases and discovers lower energies in 35.0% of cases. Remarkably, it achieved these results while using only 6.8–63.6% of the initial configurations required by the algorithmic methods (see Fig. 3b). As discussed earlier, increasing the number of initial configurations is likely to improve both the SR and LEDR. To ensure a fair comparison in this study, the number of initial configurations used by Adsorb-Agent is scaled relative to the algorithmic approach, resulting in a reduction of 6.8–63.6% of the original, with an average of 26.9% across three independent runs. For practical applications, the number of initial configurations can be further increased, particularly for systems with lower RSRs, to improve the performance of adsorption energy determination.

An analysis of specific system categories reveals that the results vary with system complexity, as shown in Fig. 3a. The systems are categorized based on the composition of the catalytic surface (monometallic and intermetallic) and the size of the adsorbate molecule (small and large molecules). Large adsorbates are defined as those with more than three atoms, such as CH_2CH_2OH, $OCHCH_3$, and $ONN(CH_3)_2$.

For systems with monometallic catalysts, Adsorb-Agent consistently identifies adsorption energies comparable to those found by the algorithmic approach across all three trials. However, it does not achieve lower adsorption energies, indicating that the algorithmic approach successfully identifies the adsorption configuration with the global minimum energy. This outcome is likely due to the relatively homogeneous nature of monometallic surfaces, which consist of a single atom type (see Fig. 2b). These findings suggest that extensive site enumeration is unnecessary for monotonous surfaces.

In contrast, for systems with intermetallic catalysts, Adsorb-Agent demonstrates a distinct advantage. While the SR slightly decreases to 82.2% compared to monometallic systems, the LEDR significantly improves to 46.7%. This highlights the ability of Adsorb-Agent to uncover new global minima through targeted searches, which the algorithmic approach cannot

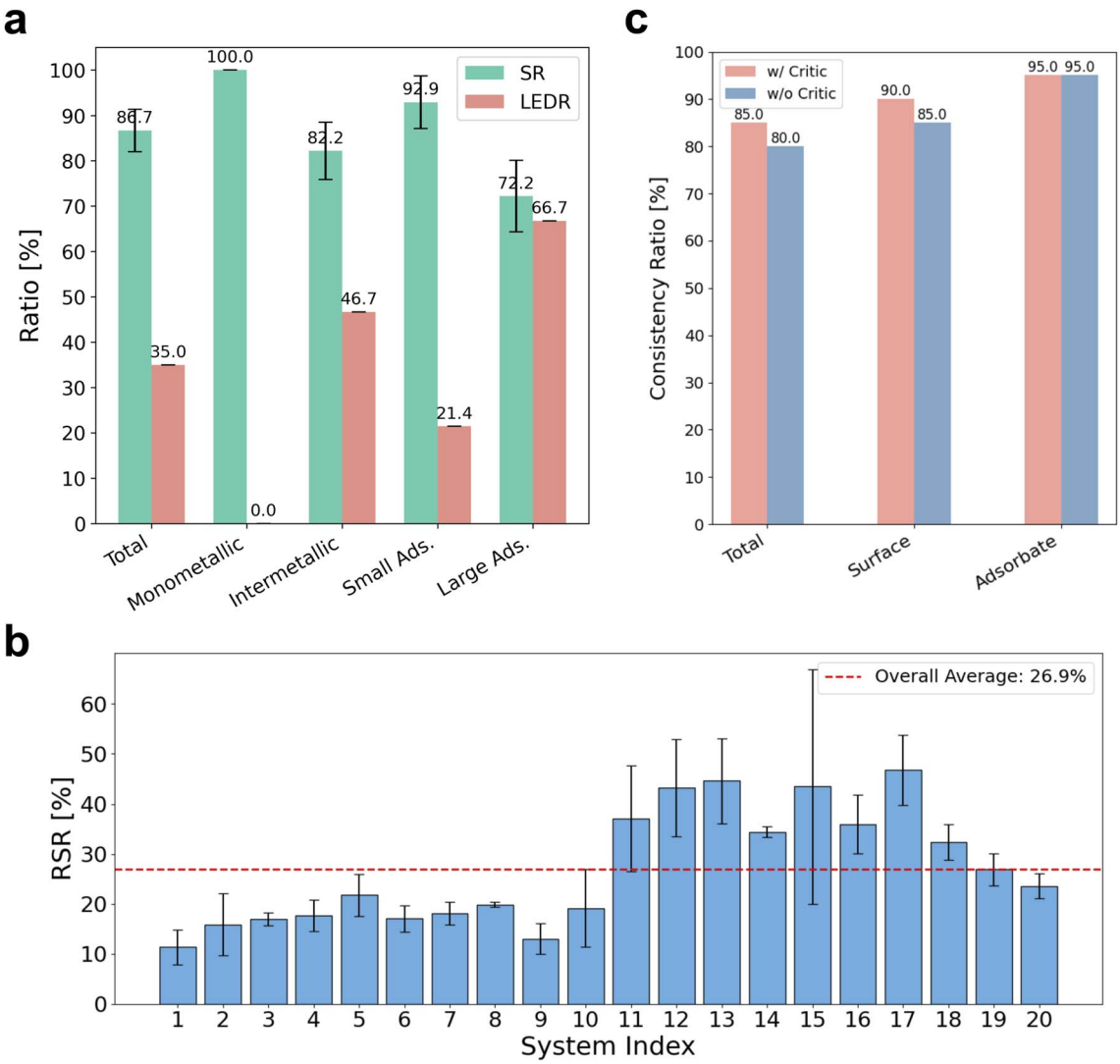

Fig. 3 Quantitative analysis of Adsorb-Agent results. (a) Success Rate (SR) and Lower Energy Discovery Rate (LEDR) across different categories; (b) Reduced Search Space Rate (RSR) obtained from three independent runs. For each system, the number of configurations is averaged over the three runs, and the error bars indicate the standard deviation; (c) consistency rate across independent runs.

achieve through simple enumeration because of the increased complexity and heterogeneity of intermetallic surfaces.

A similar trend is observed when analyzing systems based on adsorbate complexity. Systems with large adsorbates exhibit a lower SR but a significantly higher LEDR compared to those with small adsorbates. Notably, systems with large adsorbates achieve the highest LEDR of 66.7%, underscoring the effectiveness of the targeted search approach employed by Adsorb-Agent. These results suggest that simple enumeration is relatively less effective at identifying the global minimum in systems containing complex adsorbates. Furthermore, this finding reinforces the effectiveness of Adsorb-Agent in addressing these challenges.

Consistency across independent trials

As LLMs are inherently non-deterministic, ensuring consistent and reproducible outputs is important. The energy distribution panels in Fig. 4 demonstrate that Adsorb-Agent consistently identifies lower-energy configurations, rather than doing so by chance. For the NNH–$CuPd_3$ (111) and $OCHCH_3$–Hf_2Zn_6 (110) systems, the frequency of identifying lower-energy configurations is significantly higher compared to the algorithmic approach. This suggests that Adsorb-Agent effectively targets adsorption configurations closer to the global minimum. For the OH–Pt (111) system, where the algorithmic approach already exhibits a high frequency of lower-energy identifications, Adsorb-Agent preserves this trend. These results indicate that Adsorb-Agent systematically identifies configurations with energies near the global minimum, demonstrating its capability rather than relying on chance.

Furthermore, as shown in Table 1, the standard deviations of adsorption energies across multiple implementations of Adsorb-Agent remain within an acceptable range for most systems. The only exceptions are two systems involving CH_2CH_2OH, where higher deviations are observed. This variability highlights the need for further refinement in handling complex adsorbates while affirming the overall robustness of Adsorb-Agent.

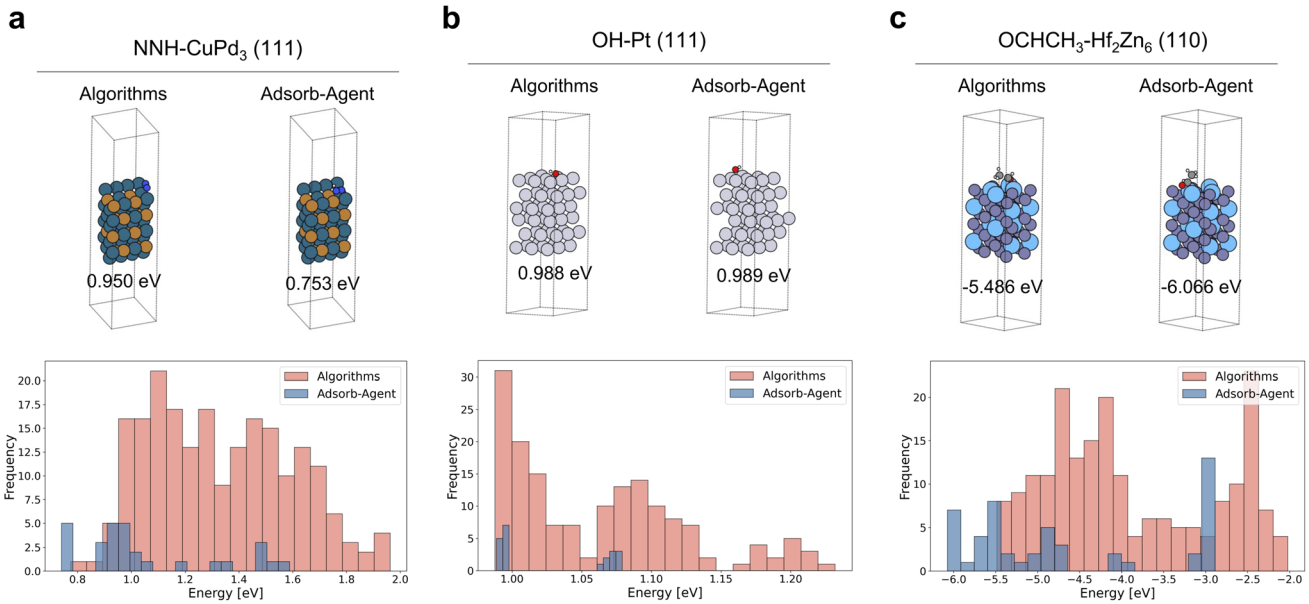

Fig. 4 Comparison of relaxed adsorption configurations (upper panels) and energy distributions (lower panels). (a) NNH–$CuPd_3$ (111); (b) OH–Pt (111); (c) $OCHCH_3$–Hf_2Zn_5 (110). The energy distributions are generated by combining results from all three independent runs.

To quantitatively evaluate consistency across independent trials, we introduce the consistency ratio, which measures the proportion of systems that yield consistent solutions across three independent trials. The consistency ratio is defined as:

$$\text{Consistency ratio}[\%] = \frac{\sum_{i=1}^{N} \mathbb{1}\left(S_{\text{trial}_1,i} = S_{\text{trial}_2,i} = S_{\text{trial}_3,i}\right)}{N} \quad (2)$$

Here, N represents the total number of systems evaluated, which is set to 20, and $S_{\text{trial}_j,i}$ denotes the solution obtained for the i-th system in the j-th trial.

Consistency is evaluated separately for surface-related and adsorbate-related information. A solution is deemed consistent if the binding atom arrays across the three trials either match exactly or differ by no more than one atom, with the shorter array being a subset of the longer array. This criterion is applied to both surface binding atoms and adsorbate binding atoms using the algorithm detailed in Algorithm 1. A solution is considered fully consistent only if both surface-related and adsorbate-related information meet these criteria.

Adsorb-Agent demonstrates reasonable reliability, producing consistent solutions for 17 out of 20 systems when the Critic module is applied. Specifically, only one system fails to achieve consistency in adsorbate-related solutions, while two systems fail in surface-related solutions. This highlights Adsorb-Agent's strong performance in generating reliable adsorbate-related solutions. Without the Critic module, one additional system fails to achieve consistency in surface-related solutions, indicating the Critic module's potential role in enhancing solution reliability. Although the limited size of the test set makes it challenging to generalize the effectiveness of the Critic module, these results suggest that the Critic module may help improve consistency by filtering out incoherent

Algorithm 1 Consistency check algorithm

```
Initialize: σ_s ← True                                  ▷ σ: consistency flag
for all a, b ∈ S do                                     ▷ S: list of binding atom groups
    if a ≠ b then
        if |len(a) − len(b)| > 1 then
            σ ← False
        end if
        shorter, longer ← min(a, b), max(a, b)          ▷ Based on length
        if set(shorter) ⊈ set(longer) then
            σ ← False
        end if
    end if
end for
```

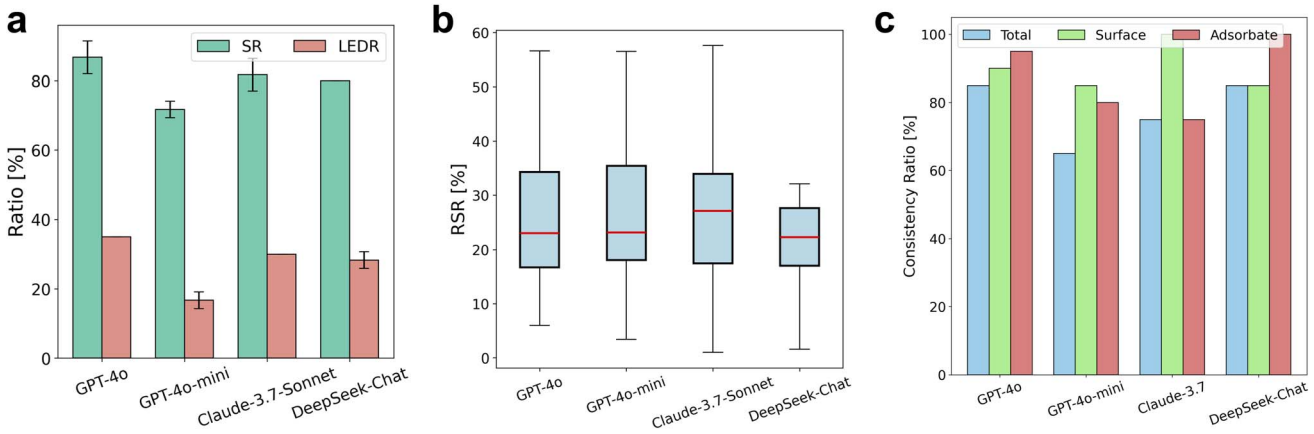

Fig. 5 Comparison of Adsorb-Agent performance using different language models. (a) Success Rate (SR) and Lower Energy Discovery Rate (LEDR) across models; (b) box plots of Reduced Search Space Rate (RSR) for each language model over three iterations, where the red line indicates the median, the box represents the interquartile range (IQR), and the whiskers extend to the minimum and maximum values within 1.5 times the IQR; (c) consistency ratios for total (surface and adsorbate categories) across independent runs.

solutions within the tested systems. Further discussion of the Critic module performance is provided in SI Section S7.

Comparison across language models

We evaluated the performance of various LLMs, including GPT-4o, GPT-4o-mini, Claude-3.7-Sonnet, and DeepSeek-Chat, serving as the "brain" of Adsorb-Agent. Among them, GPT-4o demonstrated the strongest overall performance for both SR and LEDR, as shown in Fig. 5a, although its improvements over Claude-3.7-Sonnet and DeepSeek-Chat were marginal. In contrast, GPT-4o-mini exhibited the weakest performance, with an average SR of 77% and an LEDR of 16.7%. DeepSeek-Chat showed the most aggressive reduction of the configuration space, achieving the lowest average RSR of 20.8%, while the other models yielded average RSR values in the range of 26–28%, as illustrated in Fig. 5b. Here, the average RSR refers to the mean value calculated across 20 samples over three iterations. Both GPT-4o and DeepSeek-Chat attained a high consistency ratio of 85%, whereas GPT-4o-mini displayed a substantially lower consistency ratio of 65%, as shown in Fig. 5c. In summary, GPT-4o consistently delivered the best overall performance, while GPT-4o-mini underperformed relative to other models. Notably, despite being a free and open-source model, DeepSeek-Chat achieved performance comparable to the proprietary LLMs.

Conclusion

We introduced Adsorb-Agent, an LLM-powered agent designed to efficiently explore adsorption configuration spaces and accurately identify adsorption energies. Adsorb-Agent streamlines the computational process by significantly reducing the number of initial configurations required while improving the efficiency and accuracy of minimum adsorption energy predictions, thereby mitigating the computational cost of extensive DFT simulations.

A key aspect of our approach is that the agent is built on a standard LLM trained on general language corpora, rather than domain-specific data. By leveraging its built-in understanding of chemical bonding, surface chemistry, and emergent reasoning capabilities, Adsorb-Agent autonomously proposes plausible adsorption configurations tailored to specific systems. Adsorb-Agent demonstrates a strong ability to identify configurations with energies closer to the global minimum, particularly in complex systems such as intermetallic surfaces and large adsorbate molecules. This capability highlights a critical advantage of our approach in addressing computationally intensive and chemically complex systems.

Methods

GPT-4o

The GPT-4o model is an optimized variant of OpenAI's GPT-4, which builds on the advancements of the Generative Pretrained Transformer (GPT) series.[37,38] Like its predecessors, GPT-4 is a large-scale language model based on the transformer architecture, utilizing self-attention mechanisms[39] to effectively model long-range dependencies in text. The GPT-4o model retains the core advantages of GPT-4, including its powerful transformer architecture that ensures high accuracy in language-based tasks. Additionally, GPT-4o has been optimized for more efficient processing, making it well-suited for tasks that demand reduced computational resources without compromising its ability to understand and generate human-like text. This optimization is especially valuable for multi-step reasoning, problem-solving, and decision-making applications. In this study, we used GPT-4o with its default settings: a temperature of 1.0 and top_p of 1.0.

GPT-4o-mini

GPT-4o-mini is a smaller and more lightweight variant of OpenAI's GPT-4o model. While it shares the same underlying transformer architecture and core design principles as GPT-4 and GPT-4o,[37,38] GPT-4o-mini is specifically optimized for faster inference and lower computational costs. This makes it suitable for resource-constrained environments or applications requiring rapid response times. Its reduced parameter count may lead to

trade-offs in accuracy and reasoning depth compared to the full GPT-4o model. In this study, we used GPT-4o-mini with its default settings: a temperature of 1.0 and top_*p* of 1.0.

Claude-3.7-Sonnet

Claude-3.7-Sonnet is a large language model developed by Anthropic as part of the Claude 3 family.[40] Claude-3.7-Sonnet is positioned as a mid-sized model in the Claude 3 lineup, balancing performance and speed. While its exact number of parameters is not publicly disclosed, it is intended to offer strong capabilities for multi-step reasoning and code generation, with faster responses compared to larger models like Claude-3.7-Opus. Claude-3.7-Sonnet is trained using Anthropic's reinforcement learning from human feedback (RLHF) framework, with particular emphasis on controllability and safety to reduce harmful or biased outputs. In this study, we used Claude-3.7-Sonnet through the Anthropic API with its default settings: a temperature of 1.0 and a top_*p* value of 0.999.

DeepSeek-Chat

DeepSeek-Chat is an open-source large language model developed by DeepSeek AI, released as part of the DeepSeek LLM series.[41] It is based on a transformer decoder-only architecture, similar to models like GPT, and is trained on large-scale internet datasets to handle a wide range of natural language tasks, including text generation, reasoning, and dialogue. The version used in this study is DeepSeek-V3-0324 7B, which contains approximately 7 billion parameters. It is trained with supervised fine-tuning and reinforcement learning from human feedback (RLHF) to enhance its instruction-following capabilities.[42] In this study, we used DeepSeek-Chat *via* the Hugging Face Transformers library with its default generation settings: a temperature of 1.0 and a top_*p* value of 1.0.

Adsorbate placement

A commonly used adsorbate placement involves the `heuristic` site enumeration algorithm, which leverages surface symmetry.[11,12] Using Pymatgen's `AdsorbateSiteFinder`, this algorithm identifies the most energetically favorable sites, such as ontop, bridge, or hollow sites. The adsorbate is then placed at the selected sites with a random rotation about the *z*-axis and minor adjustments along the *x*- and *y*-axes, ensuring the binding atom is positioned at the site.[14]

To expand the configuration space beyond the `heuristic` algorithm, Lan *et al.* proposed a `random` algorithm.[10] This method uniformly samples surface sites at random. After performing a Delaunay triangulation of the surface atoms, random sites are selected within each triangle. At each randomly selected site, the adsorbate is placed with random rotations about the *x*-, *y*-, and *z*-axes, ensuring alignment of the center of mass with the target site.[14]

In this paper, we introduce a new placement strategy specifically designed to accommodate side-on oriented adsorbates, offering a key differentiation from the above methods. Our approach determines the placement center as the weighted center of the binding atoms and orients the adsorbate to maximize the exposure of its binding atoms to the surface. This method differs from the above placement strategies, which position a single binding atom at the site and apply stochastic rotations. Although our strategy operates similarly to existing methods for adsorbates with a single binding atom, it specifically enables side-on placement for bidentate adsorbates.

GNN relaxation

To evaluate the ground-state properties of an atomic structure, it is necessary to optimize its geometry by minimizing the system's energy. This optimization is typically achieved by calculating interatomic forces and adjusting atomic positions accordingly while minimizing the total energy of the system. Traditionally, this process is performed using quantum chemistry methods, such as DFT. However, GNNs have recently emerged as a fast and cost-effective surrogate for DFT-based relaxation.[10,14]

In GNN-based relaxation, the atomic system is represented as a graph, where atoms are treated as nodes and interatomic bonds or interactions as edges. The GNN iteratively performs message passing, where neighboring atoms exchange information, allowing the network to capture local chemical environments. Through these learned representations, the GNN predicts both the system's energy and the atomic forces. The atomic positions are then updated using optimization algorithms based on the predicted forces, and this process is repeated until the system converges to a stable, low-energy configuration.

In this study, we use EquiformerV2 as the core GNN model due to its demonstrated high accuracy (~0.2 eV MAE). We specifically use the checkpoint "EquiformerV2-31M-S2EF-OC20-All+MD", available from the Equiformer repository.[16] This checkpoint corresponds to a model trained on the OC20 S2EF (Structure-to-Energy-and-Forces) dataset, including both the full OC20 training set and additional MD trajectories.

After the GNN relaxation, the energy of the adsorbate–catalyst system can be predicted. Since a single adsorbate–surface system may have multiple possible adsorption configurations, it is necessary to evaluate the energies of all candidate configurations to identify the most stable one. The adsorption energy is determined by selecting the configuration with the lowest energy, offering insights into the energetically favorable configuration. The calculation is defined as eqn (1).

Open Catalyst demo

Conventional GNN-based relaxation was conducted using the Open Catalyst demo (OC-demo) API. The OC-demo is an interactive platform that allows users to explore and optimize binding sites for adsorbates on catalyst surfaces.[43] It supports 11 427 catalyst materials and 86 adsorbates, with the catalyst crystal structures sourced from the Materials Project[44] and the Open Quantum Materials Database.[45] The OC-demo helps identify the adsorption energy of the selected adsorbate and catalytic surface by generating multiple initial configurations and evaluating them using state-of-the-art graph neural networks, like GemNet-OC and EquiformerV2. In our workflow,

we use the OC-demo to place adsorbates on catalyst surfaces using enumeration algorithms, perform structure relaxation with EquiformerV2, and obtain the corresponding adsorption energies. These results serve as the baseline for the conventional algorithmic approach, against which we compare our LLM-agent-based method.

Evaluation metrics

The Success Ratio (SR) estimates Adsorb-Agent's ability to identify adsorption energies comparable to those found by the algorithmic approach. Given the inherent error margins in the energy and force predictions by EquiformerV2, a successful identification is defined as the adsorption energy predicted by Adsorb-Agent falling within a predefined tolerance (ε) of the algorithmic approach. The SR is mathematically expressed as:

$$\mathrm{SR}[\%] = \frac{\sum_{i=1}^{N} \mathbb{1}\left(\left|E_{\mathrm{agent},i} - E_{\mathrm{algorithm}}\right| \leq \varepsilon\right)}{N} \times 100 \quad (3)$$

Here, N is the total number of adsorbate–catalyst systems evaluated, which is set to 20 in this study. The threshold ε is defined as 0.1 eV, approximately half the energy prediction error of EquiformerV2.[16] $E_{\mathrm{algorithm}}$ is computed as the average of three independent trials conducted using the algorithmic approaches. The indicator function $\mathbb{1}(\cdot)$ is defined to be 1 if the condition inside is satisfied and 0 otherwise.

The Lower Energy Discovery Ratio (LEDR) assesses Adsorb-Agent's ability to identify adsorption energies that are lower than those found by the algorithmic approaches. It is defined as:

$$\mathrm{LEDR}[\%] = \frac{\sum_{i=1}^{N} \mathbb{1}\left(E_{\mathrm{agent},i} \leq E_{\mathrm{algorithm}} - \varepsilon\right)}{N} \times 100 \quad (4)$$

The Reduced Search Space Ratio (RSR) quantifies the reduction in the number of initial configurations required by Adsorb-Agent ($N_{\mathrm{init,agent}}$) compared to the algorithmic approaches ($N_{\mathrm{init,algorithm}}$). A lower RSR indicates a greater reduction in the search space. It is defined as:

$$\mathrm{RSR}[\%] = \frac{N_{\mathrm{init,agent}}}{N_{\mathrm{init,algorithm}}} \times 100 \quad (5)$$

Author contributions

J. O., R. S. M., T. V., Y. J., and A. B. F. designed the research study. J. O., R. S. M., and T. V. developed the method, wrote the code, and performed the analysis. All authors wrote and approved the manuscript.

Conflicts of interest

The authors have no conflicts to disclose.

Data availability

The code supporting the findings of this study is publicly available at the following GitHub repository: https://github.com/hoon-ock/CatalystAIgent and has also been archived on Zenodo: https://doi.org/10.5281/zenodo.17585022.

Supplementary information (SI) is available. See DOI: https://doi.org/10.1039/d5dd00298b.

Acknowledgements

The authors gratefully acknowledge support from the H. Robert Sharbaugh Presidential Fellowship. We extend our gratitude to the Meta Fundamental AI Research (FAIR) Chemistry team for making the Open Catalyst demo service and the Open Catalyst Project dataset publicly available. ChatGPT was utilized in preparing the preprint version of this manuscript, specifically for assistance with grammar and typographical corrections. All authors have thoroughly reviewed, verified, and approved the content of the manuscript to ensure its accuracy and integrity.

References

1 J. Nørskov, F. Studt, F. Abild-Pedersen and T. Bligaard, *Fundamental Concepts in Heterogeneous Catalysis*, John Wiley & Sons, Ltd, 2014.
2 J. A. Dumesic, G. W. Huber and M. Boudart, *Handbook of Heterogeneous Catalysis*, John Wiley & Sons, Ltd, 2008.
3 C. L. Zitnick, *et al.*, An Introduction to Electrocatalyst Design using Machine Learning for Renewable Energy Storage, *arXiv*, 2020, preprint, arXiv:2010.09435, DOI: 10.48550/arXiv.2010.09435, https://arxiv.org/abs/2010.09435.
4 J. Nørskov, T. Bligaard, J. Rossmeisl and C. Christensen, Towards the Computational Design of Solid Catalysts, *Nat. Chem.*, 2009, **1**, 37–46.
5 B. Yang, R. Burch, C. Hardacre, G. Headdock and P. Hu, Understanding the Optimal Adsorption Energies for Catalyst Screening in Heterogeneous Catalysis, *ACS Catal.*, 2014, **4**, 182–186.
6 S. Wang, B. Temel, J. Shen, G. Jones, L. C. Grabow, F. Studt, T. Bligaard, F. Abild-Pedersen, C. H. Christensen and J. K. Nørskov, Universal Brønsted-Evans-Polanyi Relations for C–C, C–O, C–N, N–O, N–N, and O–O Dissociation Reactions, *Catal. Lett.*, 2011, **141**, 370–373.
7 J. E. Sutton and D. G. Vlachos, A Theoretical and Computational Analysis of Linear Free Energy Relations for the Estimation of Activation Energies, *ACS Catal.*, 2012, **2**, 1624–1634.
8 Z. W. Ulissi, A. J. Medford, T. Bligaard and J. K. Nørskov, To Address Surface Reaction Network Complexity Using Scaling Relations Machine Learning and DFT Calculations, *Nat. Commun.*, 2017, **8**, 14621.
9 J. Ock, T. Tian, J. Kitchin and Z. Ulissi, Beyond independent error assumptions in large GNN atomistic models, *J. Chem. Phys.*, 2023, **158**, 214702.
10 J. Lan, A. Palizhati, M. Shuaibi, B. M. Wood, B. Wander, A. Das, M. Uyttendaele, C. L. Zitnick and Z. W. Ulissi, AdsorbML: a leap in efficiency for adsorption energy calculations using generalizable machine learning potentials, *npj Comput. Mater.*, 2023, **9**, 172.

11 J. R. Boes, O. Mamun, K. Winther and T. Bligaard, Graph Theory Approach to High-Throughput Surface Adsorption Structure Generation, *J. Phys. Chem. A*, 2019, **123**, 2281–2285.
12 S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson and G. Ceder, Python Materials Genomics (Pymatgen): A Robust, Open-source Python Library for Materials Analysis, *Comput. Mater. Sci.*, 2013, **68**, 314–319.
13 A. Kolluru and J. R. Kitchin, AdsorbDiff: Adsorbate Placement via Conditional Denoising Diffusion, *arXiv*, 2024, preprint, arXiv:2405.03962, DOI: 10.48550/arXiv.2405.03962, https://arxiv.org/abs/2405.03962.
14 L. Chanussot, *et al.*, Open Catalyst 2020 (OC20) Dataset and Community Challenges, *ACS Catal.*, 2021, **11**, 6059–6072.
15 Open Catalyst Project Leaderboard, https://opencatalystproject.org/leaderboard.html, accessed: 2024-10-18.
16 Y.-L. Liao, B. Wood, A. Das and T. Smidt, EquiformerV2: Improved Equivariant Transformer for Scaling to Higher-Degree Representations, *arXiv*, 2023, preprint, arXiv:2306.12059, DOI: 10.48550/arXiv.2306.12059, https://arxiv.org/abs/2306.12059.
17 C. Lu, C. Lu, R. T. Lange, J. Foerster, J. Clune and D. Ha, The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery, *arXiv*, 2024, preprint, arXiv:2408.06292, DOI: 10.48550/arXiv.2408.06292, https://arxiv.org/abs/2408.06292.
18 D. A. Boiko, R. MacKnight, B. Kline and G. Gomes, Autonomous chemical research with large language models, *Nature*, 2023, **624**, 570–578.
19 N. J. Szymanski, B. Rendy, Y. Fei, R. E. Kumar, T. He, D. Milsted, M. J. McDermott, M. Gallant, E. D. Cubuk and A. Merchant, others An autonomous laboratory for the accelerated synthesis of novel materials, *Nature*, 2023, **624**, 86–91.
20 Y. Jadhav, P. Pak and A. B. Farimani, LLM-3D Print: Large Language Models To Monitor and Control 3D Printing, *Addit. Manuf.*, 2025, **114**, 105027.
21 B. Wander, K. Broderick and Z. W. Ulissi, Catlas: an automated framework for catalyst discovery demonstrated for direct syngas conversion, *Catal. Sci. Technol.*, 2022, **12**, 6256–6267.
22 J. Ock, C. Guntuboina and A. Barati Farimani, Catalyst Energy Prediction with CatBERTa: Unveiling Feature Exploration Strategies through Large Language Models, *ACS Catal.*, 2023, **13**, 16032–16044.
23 J. Ock, S. Badrinarayanan, R. Magar, A. Antony and A. Barati Farimani, Multimodal language and graph learning of adsorption configuration in catalysis, *Nat. Mach. Intell.*, 2024, 1–11.
24 A. A. Peterson, Global optimization of adsorbate–surface structures while preserving molecular identity, *Top. Catal.*, 2014, **57**, 40–53.
25 H. Jung, L. Sauerland, S. Stocker, K. Reuter and J. T. Margraf, Machine-learning driven global optimization of surface adsorbate geometries, *npj Comput. Mater.*, 2023, **9**, 114.
26 F. Wang, J. Guo, Y. Quan, S. Wang and Q. Wang, Recent Advances in Electrocatalytic Nitrogen Reduction to Produce Ammonia Under Ambient Conditions, *Nat. Environ. Pollut. Technol.*, 2022, **21**(4), 1847–1855.
27 J. H. Montoya, C. Tsai, A. Vojvodic and J. K. Nørskov, The challenge of electrochemical ammonia synthesis: a new perspective on the role of nitrogen scaling relations, *ChemSusChem*, 2015, **8**, 2180–2186.
28 A. Kulkarni, S. Siahrostami, A. Patel and J. K. Nørskov, Understanding Catalytic Activity Trends in the Oxygen Reduction Reaction, *Chem. Rev.*, 2018, **118**, 2302–2312.
29 J. Zhou, X. Chen, M. Guo, W. Hu, B. Huang and D. Yuan, Enhanced catalytic activity of bimetallic ordered catalysts for nitrogen reduction reaction by perturbation of scaling relations, *ACS Catal.*, 2023, **13**, 2190–2201.
30 D. Friebel, V. Viswanathan, D. J. Miller, T. Anniyev, H. Ogasawara, A. H. Larsen, C. P. O'Grady, J. K. Nørskov and A. Nilsson, Balance of Nanostructure and Bimetallic Interactions in Pt Model Fuel Cell Catalysts: In Situ XAS and DFT Study, *J. Am. Chem. Soc.*, 2012, **134**, 9664–9671.
31 K. J. J. Mayrhofer, B. B. Blizanac, M. Arenz, V. R. Stamenkovic, P. N. Ross and N. M. Markovic, The Impact of Geometric and Surface Electronic Properties of Pt-Catalysts on the Particle Size Effect in Electrocatalysis, *J. Phys. Chem. B*, 2005, **109**, 14433–14440.
32 J. Ock, P. Mollaei and A. Barati Farimani, GradNav: Accelerated Exploration of Potential Energy Surfaces with Gradient-Based Navigation, *J. Chem. Theory Comput.*, 2024, **20**, 4088–4098.
33 Z. Jiang, P. Qin and T. Fang, Theoretical study of NH3 decomposition on Pd-Cu (111) and Cu-Pd (111) surfaces: A comparison with clean Pd (111) and Cu (111), *Appl. Surf. Sci.*, 2016, **371**, 337–342.
34 Q. Gao, H. S. Pillai, Y. Huang, S. Liu, Q. Mu, X. Han, Z. Yan, H. Zhou, Q. He, H. Xin and H. Zhu, Breaking adsorption-energy scaling limitations of electrocatalytic nitrate reduction on intermetallic CuPd nanocubes by machine-learned insights, *Nat. Commun.*, 2022, **13**, 2338.
35 S. S. Borkar and M. Shetty, Density Functional Theory Investigation into Modulating Surface–Adsorbate Interactions with Strain for Ammonia Synthesis on a Pd (111) Surface, *J. Phys. Chem. C*, 2024, **128**, 12916–12930.
36 R. Zhang, S. Zhang, H. Cui, Y. Guo, N. Li and C. Zhi, Electrochemical nitrate reduction to ammonia using copper-based electrocatalysts, *Next Energy*, 2024, **4**, 100125.
37 A. Radford and K. Narasimhan, Improving Language Understanding by Generative Pre-Training, 2018, https://api.semanticscholar.org/CorpusID:49313245.
38 *OpenAI GPT-4 Technical Report*, 2023, https://cdn.openai.com/papers/gpt-4.pdf.
39 A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser and I. Polosukhin, Attention is all you need, *Advances in Neural Information Processing Systems*, 2017, vol. 30.
40 Anthropic Introducing Claude 3 Models, 2024, https://www.anthropic.com/news/claude-3-family.

41 H. Lu, W. Liu, B. Zhang, B. Wang, K. Dong, B. Liu, J. Sun, T. Ren, Z. Li, Y. Sun, C. Deng, H. Xu, Z. Xie and C. Ruan, *DeepSeek-VL: Towards Real-World Vision-Language Understanding*, 2024.

42 DeepSeek-AI, *et al.*, DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning, *arXiv*, 2025, preprint, arXiv:2501.12948, DOI: 10.48550/arXiv.2501.12948, https://arxiv.org/abs/2501.12948.

43 Open Catalyst Project – Demo, https://opencatalyst.metademolab.com/, accessed: 2025-04-17.

44 A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder and K. A. Persson, Commentary: The Materials Project: A materials genome approach to accelerating materials innovation, *APL Mater.*, 2013, **1**, 011002.

45 S. Kirklin, J. E. Saal, B. Meredig, A. Thompson, J. W. Doak, M. Aykol, S. Rühl and C. Wolverton, The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies, *npj Comput. Mater.*, 2015, **1**, 15010.