

# ChemReasoner: Heuristic Search over a Large Language Model's Knowledge Space using Quantum-Chemical Feedback

Henry W. Sprueill<sup>1</sup>, Carl Edwards<sup>2</sup>, Khushbu Agarwal<sup>1</sup>, Mariefel V. Olarte<sup>1</sup>, Udishnu Sanyal<sup>1</sup>, Conrad Johnston<sup>3</sup>, Hongbin Liu<sup>3</sup>, Heng Ji<sup>2</sup>, Sutanay Choudhury<sup>1</sup>

Pacific Northwest National Laboratory<sup>1</sup>, University of Illinois Urbana-Champaign<sup>2</sup>, Microsoft Research Azure Quantum<sup>3</sup>

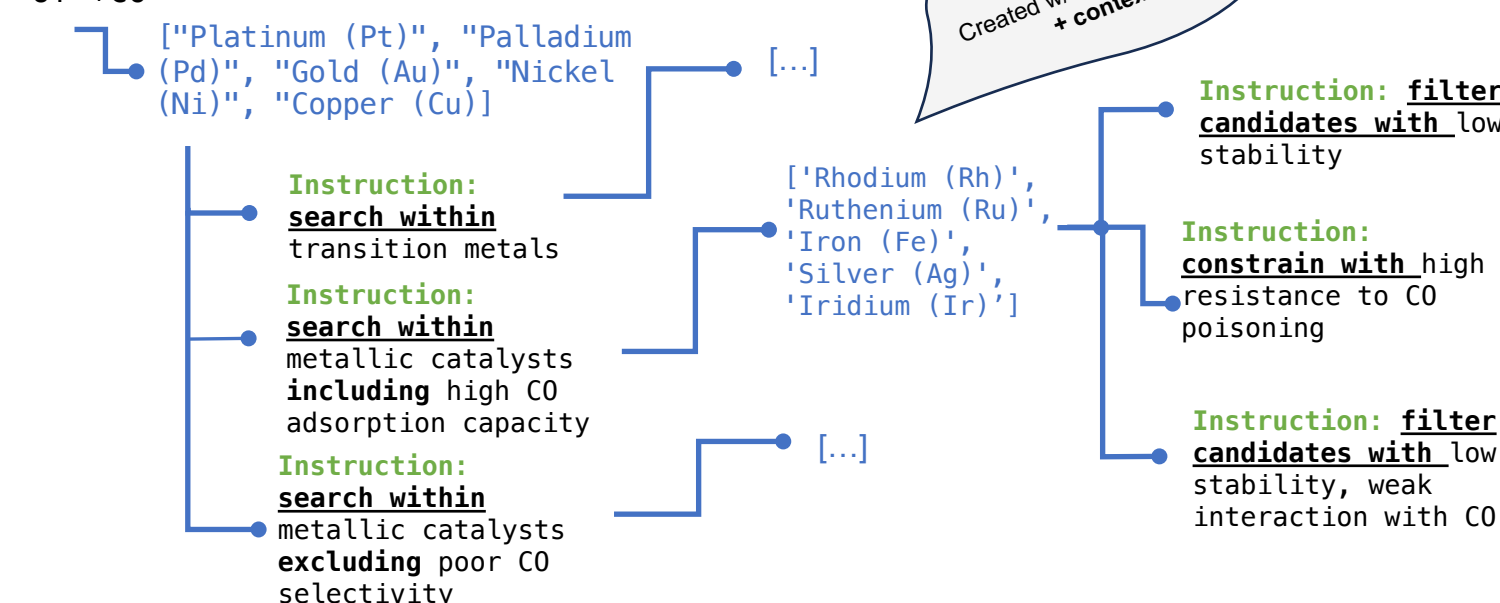
Paper Link

GitHub



## Heuristic Tree Search with LLM-based Descriptors

Instruction: Generate a list of top-5 metallic catalysts for the adsorption of \*CO



Each path in the tree is a "reasoning pathway." LLM generated catalyst descriptors narrow the search to more specific answers

## LLM-Guided Planner

Question: Generate a list of top-5 catalysts that exhibit high activity for the CO<sub>2</sub> to methanol conversion reaction

Let's consider the following factors:

**Catalyst Type:** The root prompt asks for metallic catalysts, so we don't need to change the type of catalyst. However, to narrow down our search, we could specify a subclass of metallic catalysts that are known to be effective in this reaction. For example, transition metal catalysts such as copper, zinc, and palladium are known to be active in CO<sub>2</sub> to methanol conversion.

**Inclusion Criteria:** Since we're looking for catalysts that exhibit high activity for the CO<sub>2</sub> to methanol conversion reaction, we should include this as a criteria. We might also want to include catalysts that are stable under reaction conditions and have a high selectivity towards methanol.

**Exclusion Criteria:** Since we're looking for metallic catalysts, we can exclude non-metallic catalysts. Additionally, we might want to exclude catalysts that degrade quickly or are toxic.

**Relationship to Candidate List:** We're looking for catalysts that are similar to those in our candidate list. This means they should have similar properties and performance in the CO<sub>2</sub> to methanol conversion reaction.

To address the root question, we need to modify the search state to find metallic catalysts that are effective in the CO<sub>2</sub> to methanol conversion reaction. Given the existing state, we need to add inclusion criteria that align with this requirement.

Here are my suggestions for a plan:

**Catalyst Type:** ["transition metal catalysts"]  
**Inclusion Criteria:** ["high activity in CO<sub>2</sub> to methanol conversion", "stability under reaction conditions", "high selectivity towards methanol"]  
**Exclusion Criteria:** ["non-metallic catalysts", "quickly degrading catalysts", "toxic catalysts"]  
**Relationship to Candidate List:** ["similar to"]

The LLM uses context from the previous prompts and candidates to determine the next set of catalyst descriptors with scientific reasoning

Catalyst candidates sent for computational screening

LLM Answer

Copper-Zinc Catalyst

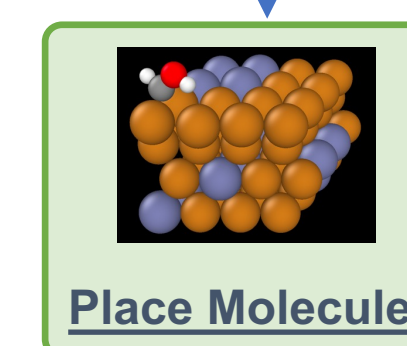
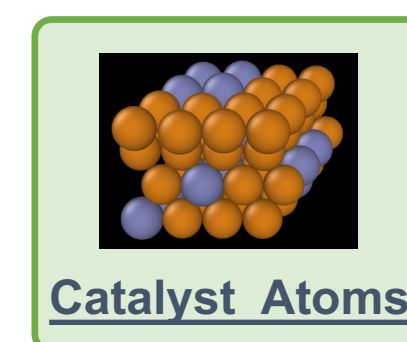
Parse Symbols

["Cu", "Zn"]

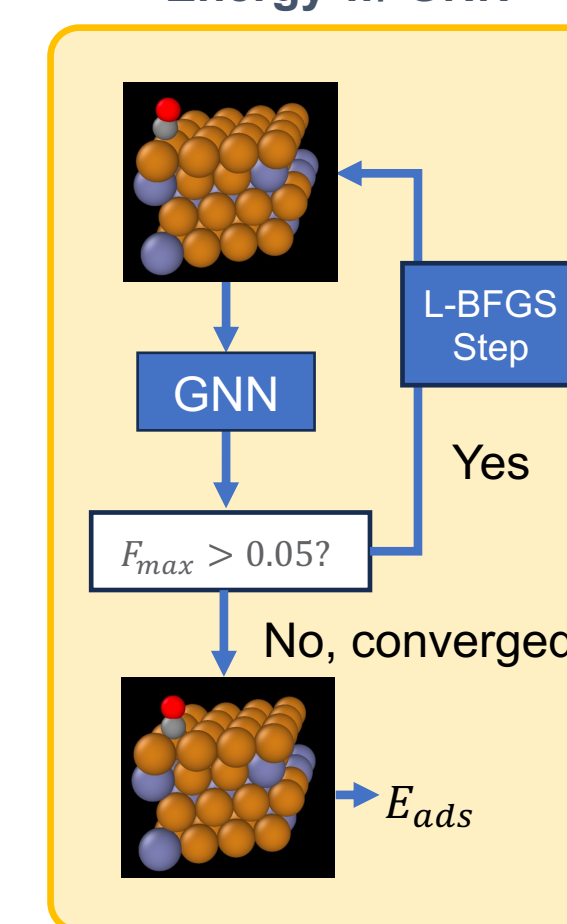
## Translation to Computational Chemistry

II

Known Reference Values



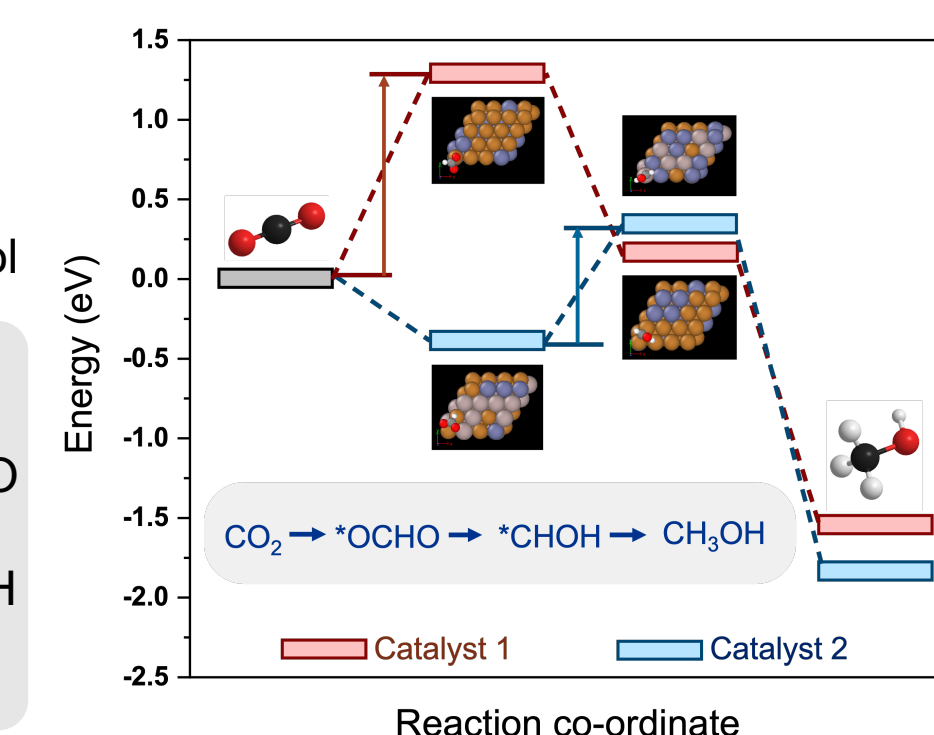
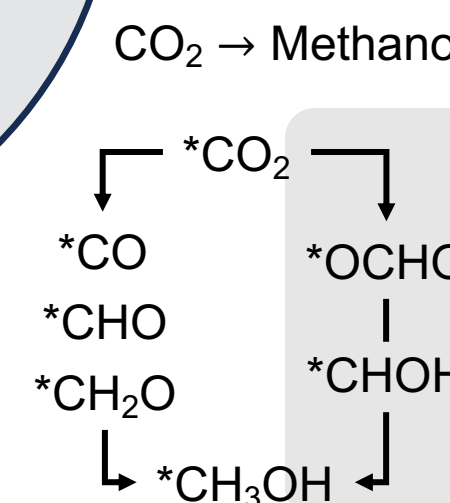
Minimize Potential Energy w/ GNN



LLM answers are translated to 3D atomic structures for evaluation. The GNN is used to relax the structures and find the minimum energy

## Reward Estimation for Chemical Reaction Pathways

III

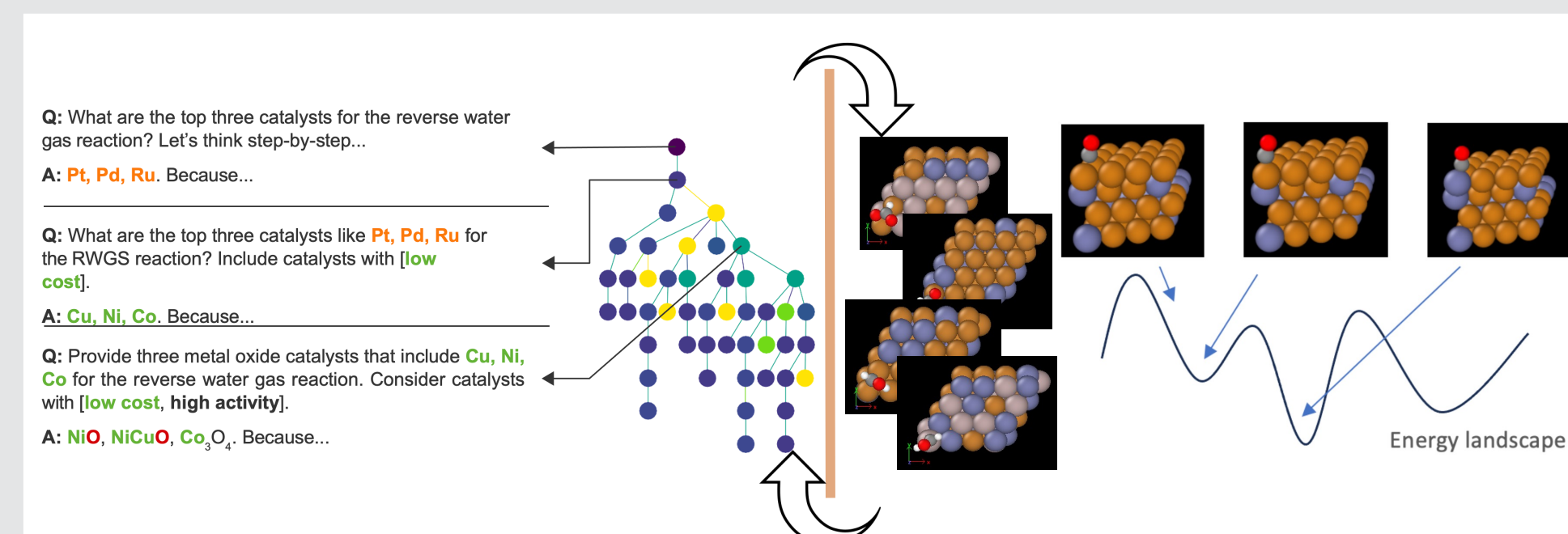


Chemical reactions occur in branching pathways of intermediate states. The reaction will take the path with the smallest hill to climb

## Motivation: Chemical Design

Exploration of the chemical space via LLM-based Descriptors

Generate 3D structures and estimate reward via DFT-surrogate models



The accelerated development of innovative clean energy technologies is imperative to meet the US Government's aggressive 2030 and 2050 net-zero CO<sub>2</sub> emission target. Catalysts play a ubiquitous role in producing renewable fuels and chemicals to achieve world-wide net-zero goal

Monolithic AI-foundational models lack the ability to propose novel configurations with sufficient mechanistic understanding of elementary steps. To tackle this issue, we propose **ChemReasoner**, which...

- reasons about design choices across macroscopic and microscopic scales, bridging complex linguistic concepts with atomistic representation
- steers exploration into novel configurations using feedback from a graph neural network (GNN) trained from ab-initio simulations

Top-k candidates selected. Plan future research

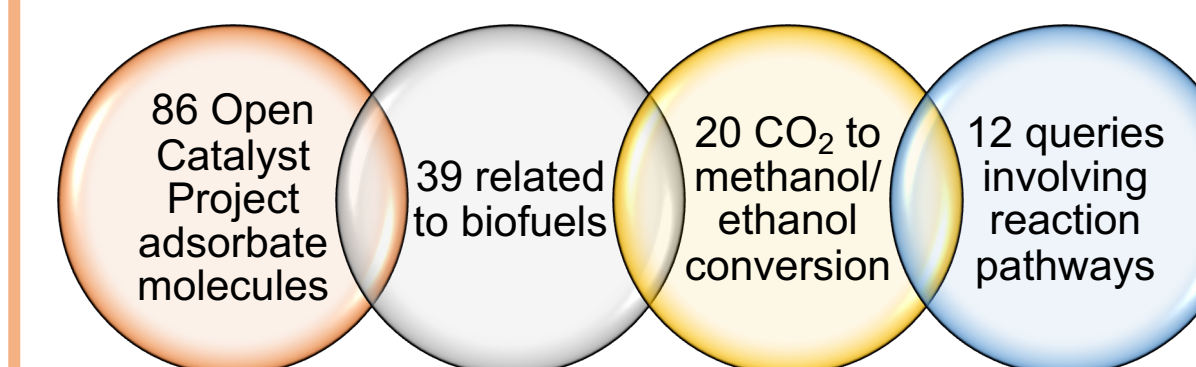
Final Reward Value

$$r(c) = - \min_{p \in \text{Paths}} \left( \max_{\text{ads}_t \in p} (E_{\text{ads}_t} - E_{\text{ads}_{t-1}}) \right)$$

## Acknowledgements

- Accelerating Foundation Models Research award, Microsoft
- Seed LDRD Program, Pacific Northwest National Laboratory
- Generative AI for Science LDRD Program, Pacific Northwest National Laboratory
- GNN Models: <https://opencatalystproject.org>

## Analysis of Search Methods



	OpenCatalyst		BioFuels		CO <sub>2</sub> -Conversion	
	GPT-4	GPT-3.5	GPT-4	GPT-3.5	GPT-4	GPT-3.5
Chain-of-Thought	0.37	0.66	2.08	2.10	-0.62	-0.54
Self Consistency	0.73	0.76	2.08	2.12	-0.54	-0.36
CHEMREASONER-Expert	1.90	2.11	3.90	3.79	0.45	<b>0.78</b>
CHEMREASONER-Planner	<b>2.36</b>	2.16	<b>4.15</b>	3.29	0.01	0.49

Analysis of rewards for 4 different prompting methods:

- **Chain-of-thought:** Standard prompting with "Let's think step-by-step"
- **Self consistency:** Highest reward answer from 10 chain-of-thought prompts
- **ChemReasoner Expert:** Tree search with expert specified catalyst descriptors
- **ChemReasoner Planner:** Tree search with LLM planner-guided catalyst descriptors

ChemReasoner methods outperform chain-of-thought and self consistency methods across each dataset and each LLM

## GNN Compared to DFT

Surface	Pathway	Adsorbate	Adsorption Energy (eV)	
			GNN	DFT
CuZn	Methanol	CO <sub>2</sub>	0.384	-0.066
		CHOH	0.552	5.951
		OCHO	0.577	5.836
		OHCH <sub>3</sub>	-1.160	2.699
CuAlZn	Methanol	CO <sub>2</sub>	0.265	6.816
		CHOH	0.609	-1.824
		OCHO	-0.125	2.820
		OHCH <sub>3</sub>	-1.589	-5.615

## Lessons Learned

- Many structures which produce low GNN-based binding energies may be unstable when calculated with DFT.
- DFT calculations may not converge for poorly mixed alloys, while GNN calculations are less sensitive mixing
- Generating 3D materials structures remains a challenge in AI4Science