CatMaster: An Agentic Autonomous System for Computational Heterogeneous Catalysis Research

Honghao Chen [1] **Jiangjie Qiu** [1] **Yi Shen Tew** [1] **Xiaonan Wang** [1]

[1]*Beijing Key Laboratory of Artificial Intelligence for Advanced Chemical Engineering Materials, Department of Chemical Engineering, Tsinghua University, Beijing, China.* Correspondence to: Xiaonan Wang wangxiaonan@tsinghua.edu.cn.

1. Introduction

While machine learning has revolutionized prediction and design procedures in catalysis, the conduction of rigorous research campaigns remains a labor-intensive, human-centric process. Heterogeneous catalysis workflows are high-cost, strongly procedural, and evidence-heavy: researchers must manually chain tools, convert data formats, monitor convergence, and make frequent micro-decisions (e.g., parameter adjustment, failure recovery, and validation selection). This operational overhead severely limits the complexity and scale of attainable research questions.

The promise of agentic catalysis research is to shift the human role from workflow operator to research architect: defining high-level scientific goals while intelligent agents handle trivial procedural execution, error recovery, and evidence curation. Here we present **CatMaster**, an autonomous agentic system designed to bridge natural-language research intent and high-fidelity simulation procedures and execution driven by large language models. Unlike generic LLM agents, CatMaster targets catalysis-specific scenarios: reliable orchestration across tightly coupled steps, long-horizon campaign management, and the generation of *auditable scientific evidence* (structures, inputs, outputs, logs, and reports). We report the current status of CatMaster's *Compute* module, where *a user specifies only the task requirement in natural language* and the system autonomously instantiates the protocol, executes on HPC resources, parses results, and produces a self-contained report, with a future scope of unifying literature/evidence reasoning modules toward a catalysis theory co-scientist.

2. System Architecture

CatMaster is a catalysis-oriented, end-to-end agentic computation engine whose design target is for *project-level campaign* rather than a single standard job (e.g., a geometry optimization). It adopts a *file-centric* execution contract: the primary state is a structured workspace of artifacts (structures, inputs, outputs, manifests, and traces), enabling restartable long-horizon runs and post-hoc scientific auditing.

A central *Orchestrator* compiles a natural-language request into a dependency-aware task graph and executes it via a schema-validated tool layer. To prevent goal drift and support multi-day iterations, CatMaster maintains (i) an execution memory for independent subtask runners servres as short-term memory for sub-task completion and (ii) a structured "whiteboard" that records facts, files, and physical constraints as an evidence world model and serves as the long-term memory. For scalability in realistic catalysis workloads, CatMaster provides batch-capable primitives (slab generation, adsorption enumeration, surrogate relaxation), a Tool Registry/Tool Graph for tool selection and composition, and heterogeneous execution across local or remote cluster backends. A controlled Python execution interface covers long-tail post-processing, while optional human-in-the-loop checkpoints support task plan revision from wrong modeling choices or execution failures (Fig. 1).

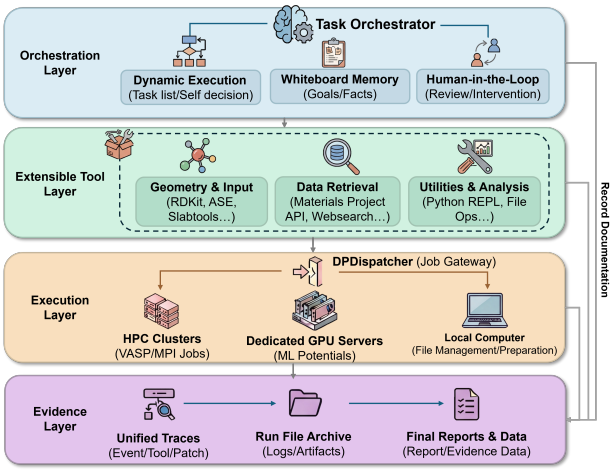

Fig. 1: CatMaster *Compute* module architecture. A natural-language request is compiled by an Orchestrator into a dependency-aware task graph, executed via a schema-validated tool graph across heterogeneous backends, and summarized into an auditable evidence trail (workspace artifacts + whiteboard state), with optional HITL recovery.

3. Related work

Recent agentic systems for chemistry and materials computation often emphasize either task-level demonstrations, broad tool wrappers, or benchmark-centric evaluation under externally scripted scaffolds.[1, 2, 3, 4, 5] Heterogeneous catalysis, however, requires *project-level* autonomy: tightly coupled multi-step planning and orchestration ability, heterogeneous execution across compute backends, and an auditable evidence chain for results tracing.

CatMaster is positioned as a catalysis-first hierarchical system that couples (i) task-graph execution

and long-horizon memory, (ii) an evidence world model with artifact-first provenance, and (iii) multi-fidelity tools for quick/rigourous tasks. This focus aims to move agentic computation from benchmark-style demonstrations toward real catalysis research campaigns.

4. Case Studies

4.1 O_2 Spin Multiplicity Comparison

To validate end-to-end integrity, we first tasked CatMaster with determining the ground state of O_2 in a vacuum box. From a simple natural-language request, the system autonomously generated VASP input files with restricted spin configurations, executed structural relaxations, parsed energies and bond lengths, and produced a final report. CatMaster correctly identified the triplet state as energetically favorable over the singlet state by $\Delta E \approx 0.381$ eV, with relaxed O–O bond lengths of 1.233–1.234 Å. This test provides a strict sanity check for autonomous input generation, convergence handling, and post-processing.

4.2 BCC Fe Surface Energies

We further evaluated CatMaster on a standard surface thermodynamics protocol: computing surface energies for BCC Fe (110), (100), and (111). Crucially, the user provided only the protocol-level requirement; CatMaster autonomously queried Materials Project, distinguished the appropriate BCC bulk reference from competing phase possibilities, selected cubic Im$\bar{3}$m Fe (`mp-13`), performed lattice relaxation, constructed symmetric slabs (15 Å slab + 15 Å vacuum), and computed surface energies using the symmetric-slab formulation:

$$\gamma = \frac{E_{\text{slab}} - NE_{\text{bulk}}}{2A}. \tag{1}$$

The computed surface-energy ordering is $\gamma_{110} < \gamma_{100} < \gamma_{111}$, consistent with representative literature reports. The magnitudes also agree well, with relative errors within 3%, which is typical given differences in computational setups. (Table 1) This benchmark demonstrates that CatMaster can instantiate and execute a multi-step thermodynamic workflow *from natural language alone,* while preserving complete computational evidence.

Table 1: BCC Fe surface energies: CatMaster vs. representative references. Values in parentheses are relative errors computed as $(\gamma_{\text{ours}} - \gamma_{\text{ref}})/\gamma_{\text{ref}}$. Units in J/m^2.

Facet	CatMaster	Heiniger[6]	Tan[7]	MP[8]
(110)	2.48	2.56(-2.98%)	2.46(+0.96%)	2.45(+1.50%)
(100)	2.56	2.63(-2.48%)	2.50(+2.59%)	2.50(+2.63%)
(111)	2.76	2.80(-1.41%)	2.73(+1.12%)	2.73(+1.12%)

4.3 High-Throughput Screening of HER Catalyst in Tenary Pt–Ni–Cu Alloy Space

To demonstrate utility in complex screening scenarios, we applied CatMaster to screen Hydrogen Evolution Reaction (HER) catalysts in the Pt–Ni–Cu alloy space. The objective was to identify candidates with near-thermoneutral H* adsorption ($\Delta G_{H^*} \approx 0$). In one run over Pt–Ni, Pt–Cu, and Ni–Cu binaries, CatMaster retrieved 10 metastable hits, deduplicated to 8 unique compositions, constructed 35 slabs across common facets, enumerated 347 symmetry-unique H* adsorption structures, and completed 383 surrogate relaxations. Candidates were ranked by $|\Delta G_{H^*}|$, as defined:

$$\Delta E_{H^*} = E(\text{slab+H}) - E(\text{slab}) - \tfrac{1}{2}E(H_2), \tag{2}$$

$$\Delta G_{H^*} = \Delta E_{H^*} + 0.24 \text{ eV}. \tag{3}$$

The top four were triggered for DFT validation by CatMaster. The final results (Table 2) show that three of these candidates achieve near-thermoneutral adsorption ($|\Delta G_{H^*}| < 0.1$ eV), successfully demonstrating the complete "retrieval → screening → validation" autonomous loop.

Table 2: Top-4 HER candidates with DFT validation (sorted by $|\Delta G_{H^*}|$ from DFT). Values are ΔG_{H^*} in eV. $G_{\text{error}} = \Delta G^{\text{MACE}}_{H^*} - \Delta G^{\text{DFT}}_{H^*}$.

Rank	Candidate	Facet	Site	$\Delta G^{\text{MACE}}_{H^*}$	$\Delta G^{\text{DFT}}_{H^*}$	G_{error}
1	Cu_3Pt	(100)	bridge_0	−0.005	−0.038	+0.033
2	Cu_3Pt	(110)	bridge_0	+0.005	−0.093	+0.098
3	$NiPt_3$	(110)	bridge_4	+0.009	+0.100	−0.090
4	$CuPt_7$	(110)	bridge_7	−0.001	+0.148	−0.149

5. Conclusion

CatMaster advances autonomous catalysis computation by prioritizing domain-specific protocol depth and artifact-based scientific evidence over generic agent breadth. Across three case studies, CatMaster demonstrates that natural-language task specifications can be translated into executable simulation protocols, executed end-to-end, and summarized into auditable reports and validation-ready artifacts.

Future work will focus on completing the "catalysis theory co-scientist" loop by integrating more parallel modules, for example, a literature-grounded reasoning module (evidence retrieval, protocol justification, and citation-aware reporting) and adding evaluator agents for robust quality control (mitigating early stopping, parameter drift, and long-horizon memory degradation). We also foresee to incorporate active learning and uncertainty-aware budgeting to drive experiment-facing recommendations, together towards the *co-scientist* for real scientific outcomes.

Acknowledgments

This work was supported by the National Key R&D Program of China (No. 2022ZD0117501), the Scientific Research Innovation Capability Support Project

for Young Faculty (ZYGXQNJSKYCXNLZCXM-E7), the Tsinghua University Initiative Scientific Research Program, and the Carbon Neutrality and Energy System Transformation (CNEST) Program led by Tsinghua University.

References

[1] Yunheng Zou, Austin H. Cheng, Abdulrahman Aldossary, Jiaru Bai, Shi Xuan Leong, Jorge Arturo Campos-Gonzalez-Angulo, Changhyeok Choi, Cher Tian Ser, Gary Tom, Andrew Wang, Zijian Zhang, Ilya Yakavets, Han Hao, Chris Crebolder, Varinia Bernales, and Alán Aspuru-Guzik. El agente: An autonomous agent for quantum chemistry. *Matter*, 8(7), July 2025.

[2] Thang D. Pham, Aditya Tanikanti, and Murat Keçeli. Chemgraph: An agentic framework for computational chemistry workflows, 2025.

[3] Samuel Rothfarb, Megan C. Davis, Ivana Matanovic, Baikun Li, Edward F. Holby, and Wilton J. M. Kort-Kamp. Hierarchical multi-agent large language model reasoning for autonomous functional materials discovery, 2025.

[4] Zeyu Xia, Jinzhe Ma, Congjie Zheng, Shufei Zhang, Yuqiang Li, Hang Su, P. Hu, Changshui Zhang, Xingao Gong, Wanli Ouyang, Lei Bai, Dongzhan Zhou, and Mao Su. An agentic framework for autonomous materials computation, 2025.

[5] Aikaterini Vriza, Uma Kornu, Aditya Koneru, Henry Chan, and Subramanian K. R. S. Sankaranarayanan. Multi-agentic AI framework for end-to-end atomistic simulations. *Digital Discovery*, 2026. Advance Article.

[6] Nathan Heiniger. A first-principles study of the surface energy of FeCrAl alloys. Technical Report TRITA–ITM-EX 2025:286, KTH Royal Institute of Technology, Stockholm, Sweden, 2025. Degree project (first cycle), 15 credits.

[7] Xiaochao Tan, Jicheng Zhou, and Yinqiao Peng. First-principles study of oxygen adsorption on Fe(110) surface. *Applied Surface Science*, 258(22):8484–8491, September 2012.

[8] Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, and Kristin A. Persson. The materials project: A materials genome approach to accelerating materials innovation. *APL Materials*, 1(1):011002, 2013.