LLM agent for mechanism and reaction discovery in heterogeneous electrocatalysis

Adem Aouichaoui [1], **Davide Tisi** [1], **Peter Beck** [1], **Emiliana Fabbri** [2], **Philippe Schwaller** [3], **Ivano E. Castelli** [1, †]

[1]Department of Energy Conversion and Storage, Technical University of Denmark, Kgs. Lyngby, 2800, Denmark
[2]PSI Center for Energy and Environmental Sciences, Paul Scherrer Institute, Villigen, 5232, Switzerland
[3]Laboratory of Artificial Chemical Intelligence, Ecole Polytechnique Fédérale de Lausanne, Lausanne, 1015, Switzerland
†corresponding author's email: ivca@dtu.dk

In computational catalysis, proposing reaction mechanisms remains a major bottleneck to discovery, as the process still relies on manual effort and on strong assumptions, such as considering only proton-coupled electron transfer steps. Universal machine-learning interatomic potentials have made the computational energetics accessible [1], so the bottleneck is no longer computing a pathway but proposing one. Large language models (LLMs) are increasingly used to automate such decisions, reasoning over chemistry while established software performs the calculation [2]. So far they have mostly scored mechanisms that were written down in advance. We present an agent that proposes them. Given only a catalyst surface and the name of a reaction, it discovers candidate mechanisms and scores them from atomistic simulations. Seven stages (see Figure 1) run sequentially, three driven by an LLM and four fully deterministic. The model selects which transformation rules apply, enumerates surface intermediates by expanding a mechanism graph, and filters proposed species for chemical plausibility. Deterministic software then constructs pathways satisfying an exact electron-count constraint, relaxes every intermediate with a universal potential, evaluates the computational hydrogen electrode [3], and ranks catalysts by limiting potential. When proposing the mechanism, the LLM is free to include chemical, potential-independent steps that hand-built pathways exclude by construction.

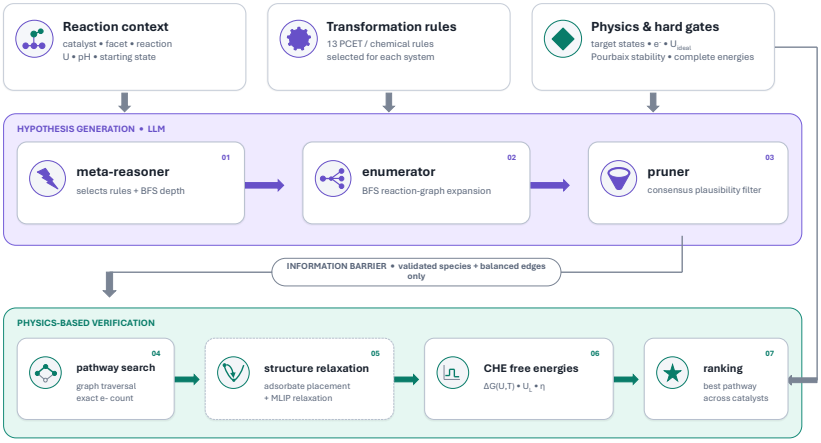

Figure 1: The seven stages. Shaded stages call an LLM, hatched contains one gated call, the rest are deterministic. The reaction definition (dashed) never enters a prompt.

Across four different reactions, the agent recovers established surface chemistry that it was never given. It reconstructs the competing carboxyl- and formate-initiated branches of CO_2 reduction, and moves the ORR limiting step to the first reduction on the weaker-binding Au surface, tracking binding-energy trends rather than reproducing a single assumed pathway. For a fixed mechanism, the pipeline is essentially deterministic, reproducing the limiting potential to within 0.5 mV across independent runs and three model checkpoints, and the selected mechanisms score marginally better than the hand-written ones evaluated through the same stack. On Pt(111) ORR the interatomic potential and electrode model account for 0.518 V of the deviation from experiment, while the choice of mechanism moves the limiting potential by 18 mV. Autonomous discovery therefore costs accuracy very little.

Repeated runs then expose challenges that are characteristic of LLM-driven mechanism discovery, showing a clear distinction between statistical accuracy and reproducibility. Run-to-run variability arises from the agent selecting between competing mechanisms that are each internally valid. On Pt(111) ORR, ten repeats partition exactly four and four between two limiting steps, at +0.265 and +0.314 V, so averaging them describes nothing physical. We also show that the limiting potential becomes ill-defined on discovered mechanisms, since it is blind to chemical steps.

References

[1] L. Chanussot et al., *ACS Catalysis* **11**, 6059 (2021).
[2] A.M. Bran, S. Cox, O. Schilter, C. Baldassari, A.D. White, and P. Schwaller, *Nature Machine Intelligence* **6**, 525 (2024).
[3] J.K. Nørskov, J. Rossmeisl, A. Logadottir, et al., *The Journal of Physical Chemistry B* **108**, 17886 (2004).