Integrating Multimodal Knowledge Mining and Autonomous Experimentation for Accelerated Electrosynthesis Discovery

Hao Han[a,b], **Sergio Pablo-Garcia**[a,b], **Shi Xuan Leong**[a,c], **Yang Cao** [a,b]

[a] , *University of Toronto, Lash Miller Chemical Laboratories, 80 St. George Street, ON M5S 3H6, Toronto, Canada.* sp.garcia@utoronto.ca, hann.hao@utoronto.ca, yangcyril.cao@utoronto.ca

[b] *Acceleration Consortium, 80 St. George St., M5S 3H6, Toronto, Canada.*

[c] *School of Chemistry, Chemical Engineering and Biotechnology, Nanyang Technological University, 21 Nanyang Link, Singapore.* shixuan.leong@ntu.edu.sg

1. Introduction

Empowered by advanced automation and cutting-edge AI techniques, self-driving labs (SDLs) have revolutionized contemporary scientific discovery, especially in the optimization and adoption of novel synthetic methodologies in automated workflows.[1] For example, Yoshikawa *et al.* successfully convinced a long-standing text-book protocol, namely “8” shape electrode polishing, is unnecessary using a robotic arm, which is further embedded in ORGANA, a robot assistant for automated chemistry experimentation and characterization.[2][3] Meanwhile, SDLs have also demonstrated their supremacy in discovery in high dimensional space, as established by their vertune in high entropy alloy for electrocatalysis. Nevertheless, most reactions involved are relatively simple, leaving the utilization of electrochemistry to access complicated chemical space underexplored. Herein, we present a combinational effort to populate machine-readable high quality data for the development of advanced electrosynthesis from both original experimental results and published literature.

2. Substantial section

One major challenge in fully leveraging the potential of modern SDLs is the lack of capability to learn from resourceful literatures, which are presented as heterogeneous data scattered across text, tables and figures. We explored multimodal large language models (MLLMs) to automatically mine the literature publications around electrosynthesis. Benchmarking MERMES-T24 dataset, we evaluate the parsing and data resolving challenges with MLLMs, which in turn lead to MERMES (multimodal reaction mining pipeline for electrosynthesis) toolkit for the end-to-end analysis and knowledge extraction of electrochemistry information.

The digitization nature of SDLs with robotics sheds light on automated, seamless translation *in silico* knowledge to reality implementations and backward data extraction. To assure the accessibility and tunability of robotic electrochemistry experiments, we developed an affordable platform based on the MEDUSA synthesis engine and homebrew potentiostat. With modular design and full API control, we demonstrated unsupervised synthesis and electrochemistry analysis under the orchestration of ChemOS 2.0, providing high quality metal complex voltammetry data with unprecedented volume and quality.

As a joint effort of both modules, we extended the capability of our tools beyond generating homogeneous datasets. Using literature knowledge informed AI engine for experiment design, we present preliminary electrochemistry synthesis and analysis using our automation platform to collect real-world data for closed-loop electrosynthesis development, revealing further potentials using generative design for novel synthetic conditions.

With our modular approach, each component—from data curation to hardware execution—contributes to a robust, autonomous workflow. By fusing LLM-derived insights with state-of-the-art automation, our platform paves the way for more efficient electrocatalysis optimization and provides a versatile framework for future synthetic applications.

Acknowledgments

The authors acknowledge Prof **Alán Aspuru-Guzik** for help on instructing, CFREF for funding

References

[1] Tom, G.; Schmid, S. P.; Baird, S. G.; Cao, Y.; Darvish, K.; **Hao, H.**; Lo, S.; Pablo-García, S.; Rajaonson, E. M.; Skreta, M.; Yoshikawa, N.; Corapi, S.; Akkoc, G. D.; Strieth-Kalthoff, F.*; Seifrid, M.*; Aspuru-Guzik, A.* Self-Driving Laboratories for Chemistry and Materials Science. *Chem. Rev.* **2024**, *124*, 9633-9732

[2] Yoshikawa, N.; Akkoc, G. D.; Pablo-García, S.; Cao, Y.; **Hao, H.***; Aspuru-Guzik, A.* Does one need to polish electrodes in an eight pattern? Automation provides the answer *Digit. Discov.* **2025**. *4*, 326-330.

[3] Darvish, K.*; Skreta, M.; Zhao, Y.; Yoshikawa, N.; Som, S.; Bogdanovic, M.; Cao, Y.; **Hao, H.**; Xu, H.; Aspuru-Guzik, A.*; Garg, A.*; Shkurti, F. ORGANA: A robotic assistant for automated chemistry experimentation and characterization *Matter*, **2024**, *2*, 101897

[4] Leong, S. X.; Pablo-Garcia, S.; Zhang, Z.; Aspuru-Guzik, A. Automated electrosynthesis reaction mining with multimodal large language models (MLLMs) *Chem. Sci.* **2024**, *15*, 17881-17819

[5] Pablo-García, S.; García, Á.*; Akkoc, G. D.; Sim, M.; Cao, Y.; Somers, M.; Hattrick, C.; Yoshikawa, N.; Dworschak, D.; **Hao, H.***; Aspuru-Guzik, A.* An Affordable Platform for Automated Synthesis and Electrochemical Characterization. *Device* **2024**, 100567