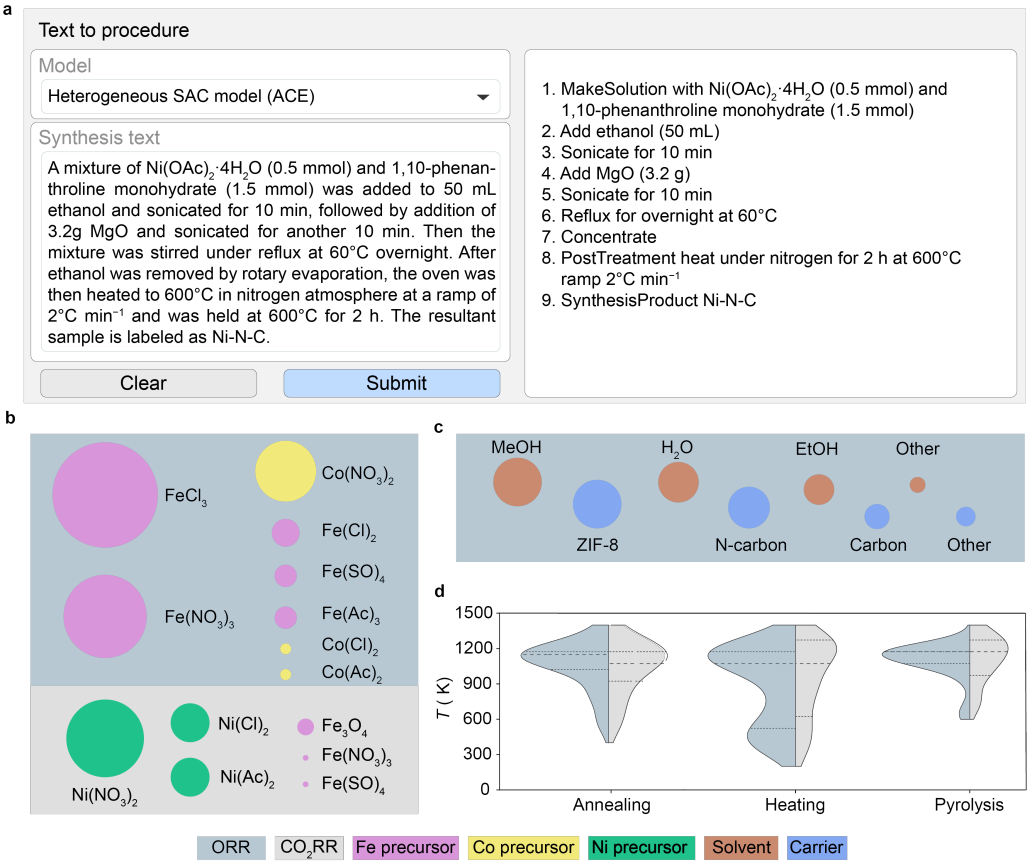
Digitally-guided synthesis of single-atom heterogeneous catalysts by leveraging language models and protocol standardization

M. Suvarna[1], A. Vaucher[2], S. Mitchell[1], T. Laino[2]*, J. Pérez-Ramírez[1]*

[1]ETH Zurich, [2]IBM Research Europe

The synthesis of single-atom catalysts (SACs) is rapidly evolving, offering precise nanoscale control over structures optimized for diverse applications[1]. However, keeping up with new developments is challenging due to the vast number of publications and lack of data standards. We develop a transformer-based language model that extracts synthesis protocols of SACs from texts and converts them into action sequences along with their associated parameters, reflecting all procedures necessary to replicate the corresponding synthesis[2] (**Fig. 1a**). We use the output of the model to perform statistical analysis of the trends in SAC synthesis conditions for the widely investigated oxygen reduction and carbon dioxide reduction reactions (**Fig. 1b-d**) and demonstrate the model's generalizability to other families of heterogeneous catalysts. Our analysis highlights a lack of standardization in reporting synthesis procedures, which limits machine-reading capabilities. We take vital steps toward implementing data-driven synthesis in heterogeneous catalysis by providing guidelines on machine-readable formats and open-source our technology as a web app for greater access by the research community.



[1] Sharon Mitchell, Javier Pérez-Ramírez, *Nat. Rev. Mater.* **2021**, *6*, 969.
[2] Alain Claude Vaucher, Federico Zipoli, Joppe Geluykens, Vishnu H Nair, Philippe Schwaller, Teodoro Laino, *Nat. Commun.* **2020**,*11*,3601.