

Data-driven Materials Science for Energy-Sustainable Applications

Jacqueline M. Cole¹

¹University of Cambridge, Ray Dolby Centre, Cavendish Laboratory, J. J. Thomson Avenue, Cambridge, UK.

jmc61@cam.ac.uk

Abstract

This talk presents the development and application of data-extraction tools and machine-learning models that are enabling data-driven materials science for energy-sustainable applications. The presentation describes how to source high-quality data to build custom databases[1,2] and apply them to: (a) train machine-learning models; (b) realize data-driven materials discovery;[3] (c) equip manufacturing processes with data-driven optimization strategies;[4] (d) map phase diagrams of materials;[5] (e) forecast trends in materials innovation;[6] (f) create energy-efficient domain-specific language models for materials science to feed agentic AI and help democratize AI for the materials research community.[7-9] The case studies presented track a theme of energy-sustainable materials.

References

- [1] M. C. Swain, J. M. Cole, *J. Chem. Inf. Model.* (2016) 56, 1894.
 - [2] J. Mavracic, C. J. Court, T. Isazawa, S. R. Elliott, J. M. Cole, *J. Chem. Inf. Model.* (2021) 61, 4280.
 - [3] C. B. Cooper, J. M. Cole et al, *Adv. Energy Mater.* 9, 1802820.
 - [4] E. J. Beard, J. M. Cole, *Sci. Data* (Springer Nature) (2022) 9, 329.
 - [5] C. J. Court, J. M. Cole, *npj Comput. Mater.* (2020) 6, 18.
 - [6] O. Sierpeklis, J. M. Cole, *Sci. Data* (Springer Nature) (2022) 9, 648.
 - [7] S. Huang, J. M. Cole, *J. Chem. Inf. Model.* (2022) 62, 6365.
 - [8] Z. Li, J. M. Cole, *Digital Discovery* (2025) 4, 998.
 - [9] O. Sierpeklis, J. M. Cole, *J. Chem. Inf. Model.* (2025) 65, 8579.
-