

Large Scale Datasets and Machine Learning for Direct Air Capture: The Open DAC Project and Beyond

Andrew J. Medford

Georgia Institute of Technology

ABSTRACT

Direct air capture (DAC) with porous adsorbents has the potential to aid large-scale decarbonization, but identifying useful sorbents for capturing CO₂ from humid air remains a formidable challenge given the vast chemical space of candidate materials such as metal–organic frameworks (MOFs). This talk surveys how AI and machine learning, powered by large, high-fidelity computational datasets, are reshaping the discovery pipeline for DAC sorbents using the Open DAC (ODAC) project as a central example. The earlier Open DAC 2023 (ODAC23) dataset established the approach with roughly 38 million density functional theory (DFT) calculations of CO₂ and H₂O adsorption across more than 8,000 MOFs. The new Open DAC 2025 (ODAC25) dataset comprises nearly 60 million DFT single-point calculations for CO₂, H₂O, N₂, and O₂ adsorption in more than 15,000 sorbent structures, introducing chemical and configurational diversity through functionalized MOFs, high-energy GCMC-

derived placements, and synthetically generated frameworks. ODAC25 substantially improves both the accuracy of DFT calculations and the treatment of flexible MOFs relative to ODAC23. Alongside the datasets, we release state-of-the-art machine learning interatomic potentials, evaluate them on adsorption energy and Henry's law coefficient predictions, and show how these models can be integrated into classical molecular simulation workflows to move toward economical and scalable DAC processes. The talk will cover these developments and briefly outline challenges and opportunities in the path from atomistic predictions to process-relevant sorbent screening.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>

