






OPEN

DATA DESCRIPTOR

A Dataset for Investigations of Amine-Impregnated Solid Adsorbent for Direct Air Capture

Eryu Wang¹, Liping Luo², Jiachuan Wang³, Jiaxin Dai¹, Shuangyin Li⁴  , Lei Chen^{2,3} & Jia Li⁵ 

Amine-impregnated solid adsorbents are widely explored for point source capture and direct air capture (DAC) to address climate change. Existing literature serves as a valuable source for the investigation of amine-functionalized solid adsorbents. This study selected 52 articles from bibliographic platforms using GPT-assisted data source screening. A total of 1,336 data points were manually collected. Each data point is characterized by 28 features including the CO₂ capture performance of various adsorbents from diluted to concentrated sources, resulting in 29,857 records. The methodology addresses inconsistencies in units and terminologies in the published articles and demonstrates database reliability, regularity and integrity through statistical analysis. The diverse types of amines and mesoporous solids in the database offer innovation potential for future research. In addition, two machine learning models were trained to promote dataset reuse by scientists from lab-based research and cheminformatics. This study provides opportunities to explore the use of machine learning on small databases and encourages data sharing and uniform reporting among DAC communities.

Background & Summary

The Paris Agreement was adopted by 196 Parties at the UN Climate Change Conference (COP 21) to limit the global average temperature increase¹. The global demand for a carbon-neutral future drives research efforts to understand climate technology at various scales^{2,3}. At COP 28, Parties decided to improve climate technology development and transfer through The Technology Mechanism, including the use of artificial intelligence (AI) in technology needs assessments for 2023–2027⁴. The roles of AI in facilitating climate technology implementations are widely discussed in aspects of carbon capture and storage (CCUS) and carbon dioxide removal (CDR) technologies^{5–8}. Direct air capture (DAC) technology is one of the CDR technologies that removes CO₂ from the atmosphere using chemisorption or physisorption. The common sorbent involves alkaline solutions, amine-functionalized adsorbents, ion-exchange resins, metal-organic frameworks (MOFs), etc. Among all these materials, the crystalline structural information of MOFs can be stored in CIF (Crystallographic Information File) format, which allows easier archiving, analysis, and visualization for data interoperability, sharing and reuse in compliance with the FAIR Principles^{9–11}. In contrast, the datasets available for amine-functionalised solids as DAC adsorbents are mostly confined to the laboratory stage.

Supported polyethyleneimine (PEI) adsorbents have been studied for carbon capture since 2002¹². The amine-impregnated adsorbent is a class of supported amines where small amines, polyamines or mixed amines and additives are impregnated on the surface of the porous solid through non-covalent interactions¹³. It is one of the most widely studied DAC adsorbents among previous research, owing to their accessibility and reactivity towards CO₂¹⁴. The physical impregnation method offers lower costs, simpler operations, and higher amine content compared to other types of supported amines¹⁵. However, challenges remain regarding amine leaching and the stability of the adsorbents¹⁶. Amine-grafted adsorbent is another class of widely studied adsorbents

¹Innovation, Policy and Entrepreneurship Thrust, Society Hub, The Hong Kong University of Science and Technology (Guangzhou), No.1 Duxue Road, Nansha, Guangzhou, 511453, China. ²Data Science and Analytics Thrust, Information Hub, The Hong Kong University of Science and Technology (Guangzhou), No.1 Duxue Road, Nansha, Guangzhou, 511453, China. ³Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Clearwater Bay, Kowloon, Hong Kong SAR, 999077, China. ⁴School of Computer Science, South China Normal University, Guangzhou, China. ⁵School of Interdisciplinary Studies, Lingnan University, Tuen Mun, Hong Kong SAR. ✉e-mail: shuangyinli@scnu.edu.cn; jia.li@ln.edu.hk

that possess higher stability due to the covalent bonding of amines to the solid supports¹⁷. Additionally, the amine-grafted adsorbent was widely involved in computational studies for investigations in CO₂ adsorption mechanism or kinetic modeling for engineering design^{18–23}. The simulation study related to amine-impregnated adsorbent is complex due to the diverse molecular structure of amines, the varied compositions and types of porous solids (supports), the movement of amines on different supports, and the synergistic interactions between amines and the supports^{24–26}. The limited availability of dataset for amine-impregnated adsorbent has hindered the application of advanced AI technology in studying their performance for DAC. Although, the powered amine-functionalised adsorbents are not suitable for pilot DAC applications, the investigation in their compositions provided valuable insights in CO₂ adsorption processes^{27–29}. Current research focuses on enhancing the efficiency, stability, and adsorption kinetics of fabricated amine-functionalized adsorbents for direct air capture (DAC)³⁰. Additionally, efforts are being made to develop simulation techniques and adsorption units suitable for larger-scale applications of related DAC adsorbents^{31,32}. The significance of sharing a computational-ready database for amine-impregnated solid adsorbents include: (1) Enabling AI models for designing and discovering amine-functionalized solid sorbents. (2) Enhancing the understanding of structure-property relationships of amine-impregnated solid sorbents through data-driven analysis and explanatory machine learning. (3) Offering opportunities for incorporating large language models in the data extractions of related amine-functionalised adsorbents. (4) Accelerating the identifications of promising amines and porous solid sorbents for DAC studies and reducing unnecessary laboratory work. (5) Providing examples for creating and training small datasets associated with material science³³. (6) Encouraging uniform data reporting, data sharing and communications of research ideas among the DAC communities.

Previous studies demonstrated dataset construction procedures through experimental data collection and data extraction from the literature^{34,35}. These datasets are suitable for training machine learning models that make predictions on the performance of the CO₂ adsorbent materials. For example, Zhang *et al.* constructed a database containing 1200 data points of porous carbon physisorbent for CO₂ capture. A deep neural network model (DNN) was trained³⁶. The model prediction strongly correlates with the experimentally measured CO₂ capacity of the selected porous carbon. In addition, a private dataset of amine-functionalised MCM-41 and SBA-15 for CO₂ capture was constructed by Yildiz *et al.* The dataset comprises 1039 data points extracted from 28 publications spanning from 2002 to 2017. Statistical analysis and decision tree-based classification over the dataset offered valuable recommendations on optimal amine type and operational conditions for achieving high amine efficiencies on carbon capture³⁷. Another private dataset was constructed based on the properties of CO₂ adsorbents prepared from Polyethylenimine (PEI) and Tetraethylenepentamine (TEPA) impregnated separately on six commercially available porous solids³⁸. A random forest model was trained to predict the CO₂ capacity, amine efficiency, and cyclic stability of the PEI and TEPA-impregnated mesoporous silica KYX using the private dataset. More recently, Li *et al.* demonstrated the training of a ChatGPT 4o model with in-context learning to predict CO₂ adsorption uptake based on input features of adsorbent properties and adsorption measurement conditions with approximately 200 data points³⁹. The findings suggest that context-based modeling is effective in reducing prediction errors when compared to conventional machine learning models in the predictive task. The more comprehensive database derived from this work would further support investigations into the effects of training data size and data diversity on the performance of machine learning models, facilitating a better understanding of how to utilize AI models for lab based research⁴⁰.

The previous research provided valuable adsorbent characteristics and performance indicators as effective features to describe different amine-functionalised solid adsorbents. Inspired by the promising results, this study aims to collect the data associated with amine-impregnated solid adsorbents with expansions in the range of data source, data size and data diversity. One of the new categories of amine-impregnated solid adsorbent is the adsorbent for DAC. Even though the research for DAC only started to emerge in the last ten years, the accumulated publications are promising in creating a compact set of data for amine-impregnated solid adsorbents with diverse chemical and structural properties. Based on the manual data collection processes introduced in previous research^{35,37}, the new dataset further considers the specific adsorbent characteristics for DAC, such as the presence of hydroxy groups or surfactants, adsorption kinetics, sorbent stability and performance at sub-ambient temperatures. These characteristics are crucial for understanding carbon capture science and advancing the amine-functionalised adsorbents for DAC.

To the best of our knowledge, there is no publicly available database on amine-impregnated solid adsorbents for CO₂ capture from both diluted and concentrated sources. Despite all the expected contributions of a publicly available dataset of amine-impregnated solids for carbon capture, many of the chemical and textural properties, and the performance indicators of those adsorbents were reported in inconsistent units or technical terms. Sometimes, the chemical and textural properties of the adsorbents are missing. Moreover, the kinetic data reported are based on adsorption modeling equations or contained in figures. Furthermore, there are multiple approaches to defining promising amine-impregnated solid sorbents, which can lead to variations in the reported types of performance indicators and missing data. In addition, some research focuses on the mechanisms rather than applications of supported amines, where different types of measurement results are reported. Therefore, the initial data collection and processing of amine-functionalised solid adsorbents relies on human expertise. Based on previous research data, this work demonstrates the construction of an amine-impregnated porous solid database for both lab and AI-related research in carbon capture science. The data were extracted from 52 articles (52 cases), with 1,336 data points and 28 features for each data point. The methods used during data collection and data processing are reported in the following sections (Fig. 1). By aggregating and statistically validating the data from various research teams, the computational-ready dataset is expected to support predictive and exploratory tasks across different laboratories, enhancing collaborative research efforts in DAC technology.

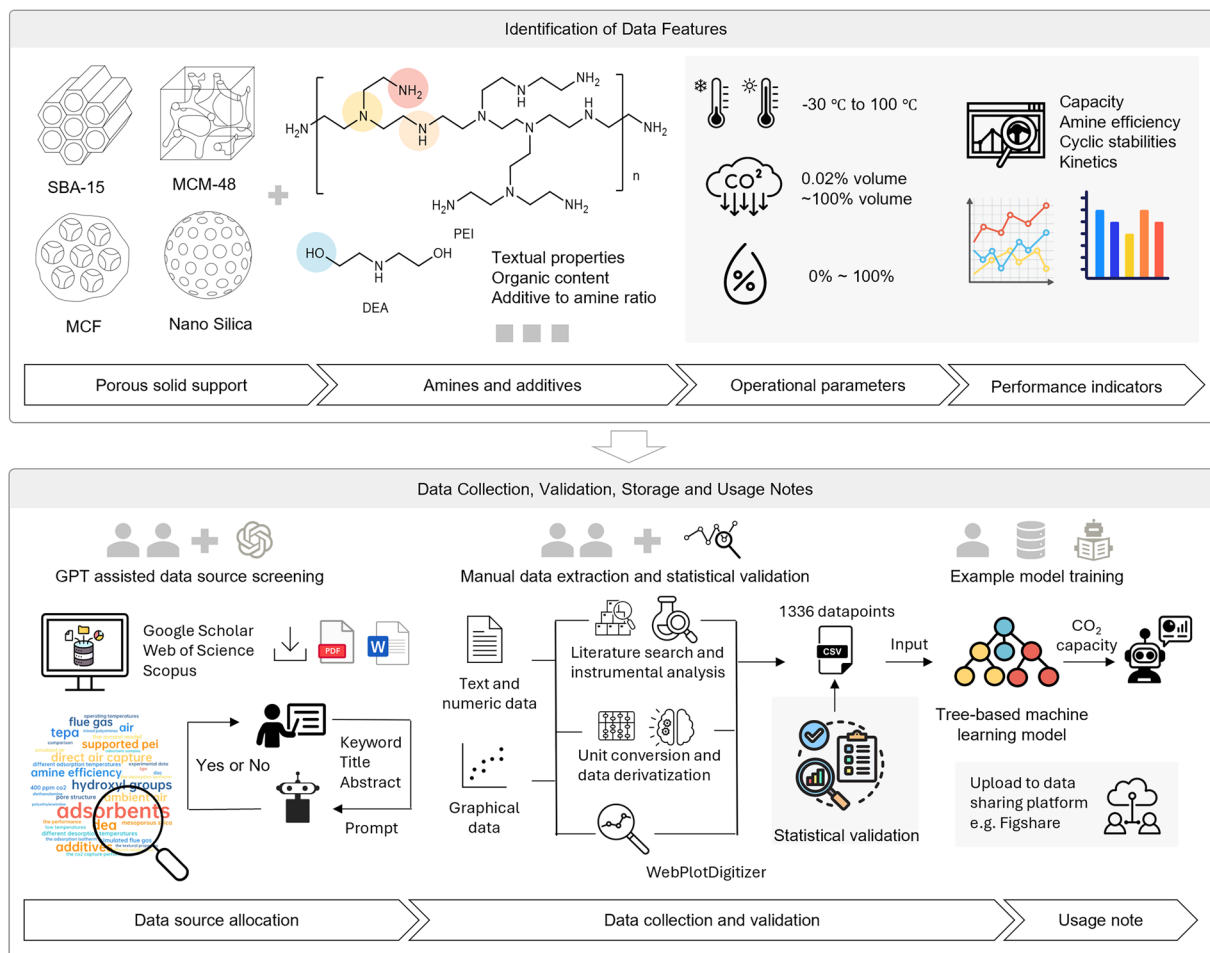


Fig. 1 The workflow of constructing the database of amine-impregnated adsorbents. The identified features of each data point were collected from the bibliometric database, and the data sources were initially allocated based on expert knowledge. Selected articles were screened by ChatGPT. The data collection process involved, data extraction from text and figures, derivatization of data, literature search and measurement of missing contextual data. The collection methods were validated through statistical analysis, and machine learning models were trained for future applications of the database.

Methods

Data source selection via manual screening and prompt engineering. A comprehensive search was conducted on Web of Science, Scopus, and Google Scholar to identify reliable and relevant sources on amine-impregnated solid adsorbents for DAC (Fig. 2a). The keywords used for the search of targeted literature varied depending on the responses from each search engine. Web of Science yields 325 articles with keywords “amine” and “direct air capture” as topic indexes and “amine efficiency” as added keywords during the literature search. Scopus yields 595 research articles with “amine” “air” and “capture” as topic indexes. The search results suggested publication numbers of research in amine functionalised adsorbent for DAC increased dramatically from 2005. Therefore, the time frame of 2005–2024 was used for the search on Google Scholar. The terms “amine”, “impregnated,” and “direct air capture” resulted in 791 articles overall. The exported literature was filtered with defined rules based on domain knowledge. The exclusion criteria are summarized in descending order based on their level of ease of recognition (Table 1).

The first round of data source screening resulted in approximately 30 articles associated with an amine-impregnated adsorbent for DAC. Noteworthy, 10 out of the 30 selected articles involved studies of supported amines not only under DAC conditions but also under simulated flue gas or pure CO₂ stream conditions. The partial pressure of CO₂ is a crucial factor influencing the performance of amine-impregnated adsorbents¹⁵. Since DAC is relatively new compared to point source carbon capture, the performance of supported amines under pure CO₂ and flue gas conditions was studied earlier, and more literature is available. These previous work provide valuable insights for investigating novel DAC adsorbents. Therefore, the database could be expanded by including studies on amine-impregnated solid adsorbents under different partial pressures of CO₂. As a challenge, the fixed-bed breakthrough experiments are one of the major techniques to assess the performance of amine-impregnated solid adsorbents in earlier publications. In addition, the focus on adsorbent properties changes with the demands of climate technologies over time. These articles are not included in the final dataset



Fig. 2 The detailed workflow on the method for data allocation. **(a)** The initial data screening was carried out by scientists in amine-impregnated solid adsorbents and **(b)** The second round of screening involves prompt engineering and interactions with GPT where the prompt used is shown in panel b. **(c)** The output by the GPT model were verified and processed by scientists in detail to collect the relevant data. Overall, there are 71 articles in total, and 52 articles were selected for final data collection. The accuracy of the GPT model was assessed using the formulas indicated in panel c.

to prevent bias introduced by the measurement techniques of the adsorbent. This will degrade the completeness of data ranges in the dataset as a trade-off. Expert-defined rules were applied during the second-round search for expansion of the datasets. The earliest research on amine-grafted solid adsorbents under both 0.04% and 10% volume CO₂ conditions was published in 2010, while the first study on amine-impregnated solid adsorbents under these conditions was published in 2011^{21,41}. Therefore, the timeframe for the second-round search is from 2011 to 2024. The investigation of the adsorbent properties under varying CO₂ conditions is anticipated to yield results that are more consistent with those from earlier studies. The same keywords were used on Web of Science and Scopus. The keywords used on Google Scholar include "amine", "impregnation", "CO₂", "adsorption", "flue gas" and "thermal gravimetric analysis (TGA)" with exclusions of "graft", "grafting" and "grafted", resulting in

Exclusion Criteria	Explanations and Limitations
Excluded Literatures	
Techno-economic analysis or life cycle assessments	Does not contain the desired data.
Sorbent fabrications, system design and simulation studies	Utilizing a standard adsorbent or a single adsorbent as examples which contain limited data associated with this dataset.
Mechanistic and ageing studies of the adsorbents	Contains distinct focuses on adsorbent properties, instrumental methods and operational conditions
Discovery of optimal preparation method or synthesis of novel adsorbents	Some research offers valuable data on optimal synthetic conditions for amine-impregnated solid adsorbents ⁵⁵ . However, some of those studies focused on only one or two adsorbent properties.
Class II and Class III amine-functionalized adsorbents	Class II (Amine-grafted adsorbents) and Class III adsorbents (made from <i>in-situ</i> polymerization of amine containing monomers) are distinct classes of supported amines with different preparation methods, chemical properties, and overall performance ¹¹⁶ .
TGA is not involved in the measurement of CO ₂ uptake	Different instrumental methods require varying sample quantities and operational parameters, resulting in slightly different measurement results.
Missing key data or overlapping points in graphical datasets	A few articles didn't contain the ratio of amine to additives or additive loading within the adsorbent ^{117,118} .
Excluded Parameters	
Sample preparation methods	The experimental parameters for impregnation are crucial for laboratory chemists to prepare optimal adsorbents. Key factors include solvent type, amine concentration, reaction scale, temperature, time, and the techniques used. While solvent types and reaction conditions are generally available, details on methods such as rotary evaporation, submersion, or sonication are often reported inconsistently ⁶⁸ . Additionally, some articles lack information on reactant quantities and ratios. Omitting these parameters may limit the dataset's usability for adsorbent preparations.
Activations	Adsorbents are pretreated by TGA before gas adsorption measurements, requiring key activation parameters such as vacuum drying conditions, activation time and temperature, ramping rate, purge gas composition, and purge gas flow rate. Excluding these parameters may reduce the dataset's comprehensiveness, but many are often underreported.
Instrument for CO ₂ adsorption	Various articles have employed different instruments for CO ₂ adsorption studies, including gas adsorption analyzers, surface area and porosimetry analyzers, fixed bed breakthrough experiments, and custom systems designed for specific conditions, particularly in humid environments ^{62,69,70} . Omitting these details may impact the dataset's accuracy for model training.
TGA temperature profiles	The TGA setup parameters, such as purge gas, the gas used to balance the feed gas during adsorption measurement, feed gas flow rate, stabilization time, and ramping rate, can significantly influence the adsorption and desorption performance of the adsorbent ⁵⁷ . Those parameters vary based on the objectives of the gas adsorption studies, the purge gas employed, and the temperatures used during each measurement. Some of these factors are inconsistently reported. Additionally, the TGA setup for cyclic studies depends on the working capacity of amine-functionalized adsorbents, adding complexity to data collection.
Energy related parameters	The kinetic and equilibrium parameters are crucial for modeling of fixed-bed column performance of amine-functionalized adsorbents ^{20,46} . Additionally, the thermodynamic parameters such as enthalpy (ΔH), entropy change (ΔS), sensible heat and activation energy are also critical for evaluating the feasibility of an adsorbent for CO ₂ capture ¹¹⁹ . These parameters provide a comprehensive representation of the adsorbent's performance, enabling engineers to assess their potential for energy-related applications. However, a significant drawback is that these data are often absent in most existing literature on amine-functionalized adsorbents and are not considered as chemical descriptors in this dataset.

Table 1. Explanation and limitations of excluding specific literature and parameters related to DAC adsorbents.

724 articles. The second round of searches focuses on literature containing various supports, amines, and additives to enhance dataset diversity. Additionally, it checks for any missed articles from the first round.

To accelerate the second round of data source screening and allow a time-saving validation of all the selected articles, a prompt was developed based on the features of a few selected articles and fed to the large language model GPT-3.5-turbo (Fig. 2b)⁴². To integrate GPT-3.5-turbo into the experimental pipeline, API calls were configured with a system prompt that defined the assistant's role, while user queries were formatted as message objects. The titles, keywords, and abstracts of all 71 potential articles identified through web searches were input into the GPT model. Based on the provided prompt, the model identified 54 articles as "Yes" and 17 articles as "No." The articles assigned "Yes" by the GPT model were thoroughly processed by scientists to extract relevant data. During this review, the scientists found that 11 out of the 54 articles were incorrectly assigned (False Positive: FP = 11, True Positives: TP = 43). The articles assigned "No" by the GPT model were checked by the scientists, and it was determined that 9 out of the 17 were incorrectly classified (False Negatives: FN = 9, True Negatives: TN = 8). After validating and finalizing all responses from the GPT model, 52 articles were selected for database construction. The performance of the GPT model was evaluated using precision and recall, resulting in an F1 score of 0.81, as indicated in Fig. 2c. The input and output for the GPT model, along with information on the selected articles, can be found in Table S1 and Table S2, respectively (see supplementary xls file).

Identification of chemical descriptors and unit conversions. The selection of appropriate performance indicators for amine-impregnated adsorbents is crucial for ensuring the dataset's quality for future research (Table 2). Both CO₂ capacity and amine efficiency are essential for evaluating an adsorbent performance based on laboratory research which also offer insights into CO₂ interactions with specific amines^{43–45}. Those two parameters were also frequently reported in former research and investigated by machine learning models^{35,38}. The kinetic parameters and heat of adsorptions are performance indicators which assist the understanding of

Features	Explanations
Chemical Properties of Amines:	
Amine type (Amine 1)	Polyamines and small amines selected for wet impregnation
1°, 2°, 3° Amine Ratio	The ratio of the quantities of primary, secondary, and tertiary amines, reported by the vendors or previous research.
Molecular Weight (g/mmol)	Number averaged molecular weight (M_n) of polyamine or molecular weight of small amines reported by the supplier or measure by GPC.
Organic Content (%)	The weight percentage of organic molecules in the adsorbents.
Amine Content (mmol/g)	The moles of nitrogen per gram of adsorbent, obtained by elemental analysis or theoretical calculations ⁶⁸ .
Additives	Additional amine molecules (e.g., diethanolamine) or additives (e.g., surfactants, templating agents) which were incorporated with amine 1 during the impregnation process to enhance the adsorbent's efficiency. Typically, either one additive (additive 2) or one other amine (amine 2) was used; however, there were instances where two additives (additive 2 and another additive 3) were mixed with amine 1.
Additive to Amine Ratio	The weight ratio of additives to amine 1 in the adsorbent.
OH to N ratio	The ratio of OH groups to amine groups in the adsorbent, typically measured through elemental analysis.
Textural Properties of the Solid Supports:	
Porous Solid Support	A wide range of porous support for amine impregnations such as mesoporous silica, hierarchical silica, MOFs and zeolites.
BET Surface Area (cm ² /g)	Obtained from BET (Brunauer-Emmett-Teller) surface area analyzer.
Total Pore Volume (cm ³ /g)	The liquid volume adsorbed at a given pressure ($P/P_0 = 0.99$) by pristine solid supports. Some research reported both the micropore and mesopore volumes separately, while the majority only reported the total pore volume.
Average Pore Diameters (nm)	Calculated from NLDFT or BJH methods or estimated based on reported surface area and pore volume (assuming the pores are cylindrical).
Operational Parameters:	
Adsorption Temperature (°C)	The predetermined temperature during the adsorption of CO ₂ .
CO ₂ concentration (vol%)	CO ₂ concentrations are 0.04% in nitrogen, helium or argon for DAC conditions, 10–30% in nitrogen, helium or argon for flue gas, and 100% in pure CO ₂ .
Adsorption time (t)	The set time for CO ₂ adsorption measurement.
Flow Rate (mL/min)	The set speed of the feed gas during the adsorption test on TGA.
Desorption temperature (°C)	The temperature required for the adsorbent to desorb CO ₂ .
Relative Humidity (%)	The percentage of water content in the fed gas for adsorption measurements.
Number of Cycles	The number of cycles performed during adsorption-desorption cyclic tests.
Performance Indicators:	
CO ₂ Capacity (mmol/g)	The maximum capacity of CO ₂ (absolute capacity) that an adsorbent can adsorb within a specified time.
CO ₂ /N Amine Efficiency (mmol/mmol)	The moles of CO ₂ adsorbed per mole of amine group within a solid amine adsorbent.
Time to half saturation (min)	The time required for the solid adsorbent to reach half of its saturation capacity (absolute capacity).
Time to 90% Saturation (min)	The time required for the solid adsorbent to reach 90% of its saturation capacity (absolute capacity).
Weight Loss Stability (%)	The percentage change in the weight of the adsorbents in the final adsorption cycle compared to their initial CO ₂ adsorption capacity.
Capacity loss stability (%)	The percentage change in CO ₂ uptake capacity of the adsorbents was measured at the final adsorption cycle compared to their initial CO ₂ adsorption capacity.
Heat of Adsorption (kJ/mmol)	The heat of adsorption data indicates the strength of interaction between CO ₂ and the amine-impregnated adsorbents. It was mainly estimated based on the measurement of TGA differential scanning calorimetry (DSC), a few heat of adsorption data were obtained using the Clausius-Clapeyron equation or dual-site temperature-dependent Toth isotherm ^{46,71,72} . One article reported it as heat of desorption which is considered interchangeable with heat of adsorption ^{47,73} .

Table 2. The selected features within the dataset. GPC: Gel Permeation Chromatography, NLDFT: non-local density functional theory, BJH: Barrett-Joyner-Halenda.

adsorption phenomena^{46,47}. Although these parameters were not widely reported for amine-impregnated adsorbents, they are valuable for the simulation and modeling of adsorption breakthrough curves and regeneration heat during the engineering design processes^{46,48}. The time to half saturation and 90% saturation of the adsorbent is chosen since the adsorption behaviour of most amine-impregnated adsorbents was not represented by kinetic models and the rate coefficient were not widely reported. As a drawback, apart from the time to reach CO₂ saturations the reported energy-related parameters were limited especially for amine-impregnated adsorbents. The limitations for excluding energy-related factors are summarised in Table 1. The cyclic stability related to thermal, hydrothermal, and chemical parameters of the adsorption process is another crucial performance indicator to assess the sustainability and economic feasibility of an adsorbent⁴⁹. Even though the cyclic stability of powdered adsorbents measured by TGA provides insights into the performance of adsorbents under practical conditions, stability-related data may vary under breakthrough conditions and was sometimes not reported in the selected studies^{50,51}. This dataset only includes commonly reported temperatures for the adsorption-desorption process, the number of adsorption-desorption cycles, and the weight and capacity loss of adsorbents.

Previous lab research has predominantly focused on optimizing adsorbents by employing different amines and solid supports to achieve an optimal balance among several performance indicators. However,

the preparation of each individual adsorbent typically requires time from 6 to 12 hours in the laboratory, with adsorption measurements extending the overall timeframe even further. The selected chemical descriptor of this database focuses on the amine and support properties for lab-based research to identify potential combinations of starting materials to enhance target performance indicators (Table 2). Applying machine learning or advanced AI models to this dataset would help chemists identify patterns and optimal feature combinations to achieve target properties, thereby reducing the learning curve for researchers across various disciplines.

The composition and structure of different amine molecules are key features influencing the performance of adsorbents. For the dataset to be deemed suitable for computational tasks, the input datasets should encompass more than abbreviated amine names. Describing the molecular structure of amine molecules with numerical input can make the dataset more practical for scientists without significant coding knowledge. The adsorbents preparation uses various linear polyamines, branched polyamines or small amines, which exhibit different mobility, viscosity, and volatility when impregnated within porous solids^{14,24}. Chemically, the ratio of primary, secondary, and tertiary amines within a specific polyamine molecule interacts with CO₂ through different mechanisms, making it a key factor influencing the efficiency and oxidative stability of the adsorbents^{52,53}. The molecular weight of the polyamines is related to the viscosity of the adsorbents and correlates with the performance in CO₂ capacity and amine efficiency⁵⁴. Furthermore, the presence of additives could also influence CO₂ capacity, amine efficiency, adsorption kinetic and adsorption heat of amine-impregnated adsorbents⁵⁵. In terms of the porous support, their surface chemistry, pore size and pore geometries are also crucial factors to influence the amine loading, capacity, kinetics or stabilities of the adsorbent^{14,56}.

Apart from features of amine and porous support the preparation and activation of amines and instrumental setups for CO₂ adsorption also influence the properties and performance indicators of the adsorbents. These experimental parameters are critical for achieving the desired amine loading and textural properties of amine-impregnated adsorbents, as well as for optimizing their performance⁵⁵. The operational parameters such as adsorption temperature, humidity and flow rate of purge gas are critical for the measured adsorbent performances^{21,57,58}. The information about the methods used for measuring adsorption performance is essential for ensuring reliable comparisons and interpretations across different studies. The explanations and limitations for excluding these factors are outlined in Table 1. Additionally, Table S3 (see supplementary xlsx file) summarizes missing experimental and instrumental parameters, highlighting that the availability of data for these factors in constructing a complete dataset involves trade-offs.

The following section address issues in inconsistently reported chemical features. The unit conversion and data derivatization process ensure that the dataset is aligned with the original research while enhancing the integrity of the datasets. In some articles, the term “amine loading” was used to describe the percentage weight of amine impregnated on a solid adsorbent; however, others employ the term “organic content” instead. The term “amine loading”, “amine content”, or “nitrogen content (*N content*)” were also used to denote the quantity of nitrogen per gram of the adsorbent. Furthermore, the *N content* is sometimes reported as a percentage of the adsorbent's weight ($N_{\text{adsorbent}}$), these values were converted to moles of nitrogen per gram of adsorbent (**Formula 1**). The molar mass of nitrogen is 14.01 g/mmol. $M_{\text{adsorbent}}$ represents the weight of the adsorbent. This term can be omitted from the formula as it does not impact the calculation.

$$N \text{ content} = \frac{N_{\text{adsorbent}} \cdot M_{\text{adsorbent}}}{14.01 \cdot M_{\text{adsorbent}}} \quad (1)$$

Notably, within our dataset, “organic content” is employed to represent the weight percentage of organic molecules impregnated on the solid adsorbents, a term considered more precise, particularly when the solid adsorbents encompass both amine molecules and additives.

Most of the obtained data was reported in consistent units, but some literature presented inconsistent units that required further processing by scientists. One common inconsistency is the use of “per gram adsorbent” or “per gram solid support” as the denominator for organic content (%), pseudo-equilibrium CO₂ uptake (mmol), and textural properties of the solid adsorbents (6 cases). Jones *et al.* explained that reporting textural properties per gram of solid support allows for more insightful structural comparisons. They also reported CO₂ capacities in both units for easier comparison⁵⁹. Other studies also use different CO₂ capacity units such as “mg/g_{adsorbent}” (3 cases) and wt% of the adsorbent (1 case). Since most research uses “per gram adsorbent” as the denominator for representing properties and performance of amine-impregnated adsorbents, CO₂ uptake capacities per mass of adsorbent were converted to mmol/g⁶⁰.

As a challenge, the units for the textural properties of the solid adsorbents cannot be converted directly without instrumental measurement and analysis. To address this problem, we only report the textural properties of the pristine solid supports as textural property features in our final datasets. Moreover, utilising the textural properties of pristine solid supports as the input for AI model training is justifiable, as these properties are more fundamental than the textural properties of impregnated solid supports in laboratory-based research. The textural properties of commercially available porous solids may be provided by the vendors, and no experiment is required for such input parameters. When the organic content (*x*) of an adsorbent is reported as percentage per gram of the solid support (x_{support}), conversions are made using the following formula to obtain the percentage value per gram of adsorbent ($x_{\text{adsorbent}}$). Noteworthy the mass of solid support (M_{support}) can be disregarded (**Formula 2**).

$$x_{\text{adsorbent}} = \frac{x_{\text{support}} \cdot M_{\text{support}}}{M_{\text{support}}(1 + x_{\text{support}})} \quad (2)$$

Calculations of nitrogen content and amine efficiency. The chemical feature “Nitrogen Content” is often missing; however, this feature can be easily calculated when both the CO₂ capacity and the amine efficiency are provided by the article (**Formula 3**).

$$\text{Amine Efficiency} = \frac{\text{CO}_2 \text{ Capacity}}{\text{N Content}} \quad (3)$$

In some of the previous research neither the N content nor the amine efficiencies are provided, the N content can still be estimated based on the molecular structure of the amine. When small amine is used for impregnation, the N content is determined by the number of N ($N_{\text{small amine}}$) in the molecule and the number of moles of amine per gram of adsorbent (**Formula 4**). The moles of amines per gram adsorbent can be calculated from the mass of amines ($\text{Organic Content} \cdot M_{\text{adsorbent}}$) over the molecular weight of the amine (MW).

$$\text{N content} = \frac{N_{\text{small amine}} \cdot \text{Organic Content} \cdot M_{\text{adsorbent}}}{MW \cdot M_{\text{adsorbent}}} \quad (4)$$

When polyamine is used for impregnation, the degree of polymerization of the polymer (DP) is first determined via the number average molecular weight (M_n) of the polyamine over the molecular weight of the repeating unit (\bar{m}) (**Formula 5**). The number of N within a polyamine ($N_{\text{polyamine}}$) can be calculated based on the number of N atoms within the repeating unit ($N_{\text{repeating unit}}$) and DP . The N content can then be estimated following the similar procedure as the small amine-impregnated solid adsorbent (**Formula 6**). The simplified formula showed that only the molecular weight of the repeating unit (\bar{m}) of a polyamine and the organic content are needed to estimate the N content per gram of the adsorbent (**Formula 7**).

$$DP = \frac{M_n}{\bar{m}} \quad (5)$$

$$\text{N content} = \frac{(DP \cdot N_{\text{repeating unit}}) \cdot \text{Organic Content} \cdot M_{\text{adsorbent}}}{M_n \cdot M_{\text{adsorbent}}} \quad (6)$$

$$\text{N content} = (\bar{m} \cdot N_{\text{repeating unit}}) \cdot \text{Organic Content} \quad (7)$$

Estimation of additive to amine 1 ratio and OH to amine ratio. When the amine-impregnated solid adsorbent contains additives, the organic content involves both amines and additives. However, a few articles do not contain the organic content of the adsorbents. In one case, the article reported the percentage weight of amine 1 (x_1) per gram of the adsorbent ($M_{\text{adsorbent}}$) and the percentage of additive (x_2) per gram of the support (M_{support})⁶¹. The *Organic Content* (**Formula 8**) and the additive to amine 1 ratio (R_a) can be calculated (**Formula 9**). Detailed calculations are provided in Table S4 (see supplementary xls file).

$$\text{Organic Content} = 1 - \frac{1 - x_1}{1 + x_2} \quad (8)$$

$$R_a = \frac{x_2}{x_1} \cdot \frac{1 - x_1}{1 + x_2} \quad (9)$$

In a few cases, the OH to amine ratio (R_{OH}) was not reported. The (R_{OH}) can be determined from the known percentage weight of amine 2 or additive 2 per gram of adsorbent (x_2), the molecular weight of the additive (MW_{additive}), the number of OH groups (N_{OH}) per additive molecule and the *N content* of the adsorbent (**Formula 11**). In some instances, the percentage weight loading for amine 2 or additive 2 per gram of adsorbent (x_2) was not reported. This can be estimated based on the organic content and the ratio of amine 2 or additive 2 to amine 1 (R_a) as outlined in **Formula 10**. Notably, R_a is provided in most articles. Detailed calculations are provided in Table S5 (see supplementary xls file).

$$x_2 = \frac{R_a \cdot \text{Organic Content}}{1 + R_a} \quad (10)$$

$$R_{OH} = \frac{x_2 \cdot N_{OH}}{MW_{\text{additive}}} \cdot \frac{1}{\text{N content}} \quad (11)$$

The estimation of average pore diameters. The average pore diameter (V_p) of the pristine solid support was not reported in 15 out of the 52 cases of literature. Those missing value were estimated from the reported BET surface area (S) and the average pore volume (D_p). The formula applies exclusively to porous supports with cylin-

Model	Equation	Rearranged Form	Parameters
Pseudo-second order	$q_t = \frac{k_2 q_e^2 t}{1 + k_2 q_e t}$	$t = \frac{q_t}{k_2 q_e (q_e - q_t)}$	k_2 (mg/g·min) ⁻¹ : the rate constant of second-order adsorption.
Avrami model	$q_t = q_e [1 - \exp(-k_1 t^n)]$	$t = \left(\frac{\ln \frac{q_e}{q_e - q_t}}{k_1} \right)^{\frac{1}{n}}$	k_1 (min) ⁻ⁿ : the rate coefficient n : Avrami constant
Fractional-order model	$q_t = q_e - (q_e^{1-n} + \frac{n-1}{m} k_n t^m)^{\frac{1}{n-1}}$	$t = \left(\frac{m(q_e - q_t)^{1-n} - q_e^{1-n}}{k_n(n-1)} \right)^{\frac{1}{m}}$	n : driving force constant m : diffusion resistance constant
Double Experimental Model	$q_t = q_e - D_1 e^{-k_{D1} t} - D_2 e^{-k_{D2} t}$	Rearrangement of the equation is not applicable. k_{D1} (min ⁻¹) and D_1 (mg/g) are the adsorption rate and diffusion rate for the fast step; k_{D2} (min ⁻¹) and D_2 (mg/g) are the adsorption rate and diffusion constant for the slow step.	

Table 3. The kinetic models used to calculate the time to half saturation and the time to 90% saturation of the adsorbents.

drical pore geometries, such as SBA-15, MCM-41, and MCM-48. Three of the fifteen article didn't report average pore diameters for more complex porous supports, including commercial nano silica, mesoporous silica foam, and zeolites, leading to 47 missing data points.

$$D_p = \frac{4V_p}{S} \quad (12)$$

Calculation of adsorption from kinetic models. The adsorption isotherm is a valuable mathematical representation to describe the kinetics of each type of adsorbent. From the existing studies, exponential curves are typically used to represent the time to half CO₂ saturation and the time to 90% CO₂ saturation of the adsorbents. Thus, both can be easily extracted from the adsorption isotherm. When only graphical data are available and no numerical information is provided in the text, validating these graphical data becomes challenging. Therefore, by calculating some time data using kinetic models that exhibit a strong fit to the experimental results ($R^2 \geq 0.98$), we can assess the accuracy of the data extraction process by comparing the extracted graphical data to the calculated values. However, only 7 out of the 52 selected articles presented adsorption kinetic models with fully specified parameters that demonstrated a coefficient of determination (R^2) from 0.970 to 0.999. The reported kinetic models from all seven articles are summarized in Table 3. Three out of the seven cases reported only kinetic parameters without the full adsorption isotherms^{62–64}. The time to saturation capacity and the time to reach 90% CO₂ saturation were calculated based on the the Avrami model, the fractal-like pseudo-first-order model and the double-exponential model provided in those three studies. The remaining four out of the seven cases reported both the adsorption isotherms and kinetic models. The time parameters were derived from the pseudo-second-order model and the Avrami model and were then compared to the extracted graphic data from the adsorption isotherms in those four cases for data validation. The parameter q_e is the CO₂ capacity at equilibrium and q_t is the adsorbed amount at a given time. The adsorption time (t) can be calculated by rearranging most equations to solve for t or by using the solve function in Python's SciPy package. The detailed calculations are provided in Table S6 (see supplementary xls file).

Extraction of graphical data. The graphical data were extracted using WebplotDigitizer, a widely used data extraction tool in previous research⁶⁵. For example, Aditya *et al.* used WebplotDigitizer to extract critical TGA trace temperatures corresponding to CoRE MOF 2019 structures from 1,886 articles⁶⁶. Yuan *et al.* also use it for data collection of biomass-derived porous carbon adsorbents for CO₂³⁵. In this dataset, the operational parameter features were extracted from approximately 450 figures, including scatter plots and bar charts for amine efficiencies and CO₂ capacity, bar charts for cyclic stabilities, and exponential curves for kinetic studies. The most distant data points on the two axes were selected to calibrate the diagram for the highest possible accuracy. The time to half saturation and time to 90% saturation, derived from the pseudo-equilibrium CO₂ uptake capacity, were identified on the y-axis to determine the corresponding time on the x-axis in each coordination diagram.

Experimental measurement of number averaged molecular weights. Molecular weight is correlated with the viscosity of the polymer and is a straightforward input that can be easily used and understood by scientists and engineers across all chemistry disciplines. For polyamines, the reported molecular weight can be either the weight averaged molecular weight (M_w) or the number averaged molecular weight (M_n). Most studies report the value provided by the vendors. In some cases, when polyamines are synthesized in a specific study, M_n is reported. Those reported M_n values were recorded in the dataset directly. The M_n value is required to obtain the degree of polymerization for amine content estimations. However, some articles only provided the weight-average molecular weight M_w of polyamines as reported by vendors, with the M_n values missing. The polyamines referenced in these studies include branched PEI with molecular weights of 1200, 1800, and 10,000 Da purchased from Alfa Aesar, as well as branched PEI with a molecular weight of 600 Da from Adam and Aladin (Fig. 4). To avoid bias in the dataset, the M_n values of these polyamines were measured using gel permeation chromatography (GPC) in this study. The GPC of the branched polyamines were performed on a Waters 1515 system equipped with a refractive index detector (RI-2414) and PL MIXED 7.5 × 50 mm guard and PL MIXED-C 7.5 × 300 mm column. Each polyamine sample (8 mg) was dissolved in CHCl₃ (4 mL) and filtered through a 0.22 filter membrane into a 1.5 mL vial. The eluent was CHCl₃ at 35 °C, with a flowrate of 1 mL min⁻¹. The system was calibrated with a narrow polystyrene standard.

Column	Column Heading	Type	Number of records	Additional Information
A	Support	string	1336	Abbreviated support names
B	Amine 1 or Additive 1	string	1336	Abbreviated polyamine names
C	1°, 2°, 3° Amine Ratio	integer	1336	Reported as ratios
D	MW or M_n (g/mol)	integer	1336	Reported or measured molecular weight of small amine and polyamines
E	Organic Content on Adsorbent Basis (%)	integer	1336	Obtained from literatures or calculated from the N content
F	N Content on Adsorbent Basis (mmol/g)	integer	1336	Obtained from literature or calculated
G	Amine 2 or Additive 2	string	1336	The designation "0" indicates that there is no amine 2 or additive 2 mixed with amine 1 in the adsorbent.
H	Additive 3	string	1336	The designation "0" indicates that there is no amine 3 or additive 3 mixed with amine 1 in the adsorbent.
I	Amine 2 or Additive 2 to Amine 1 ratio	integer	1336	The designation "0" indicates that there is no amine 2 or additive 2 mixed with amine 1 in the adsorbent. Related calculations are presented in Table S4
J	Additive 3 to Amine 1 ratio	integer	1336	The designation "0" indicates there is no additive 3 added to the adsorbent.
K	OH to N ratio	integer	1336	The designation "0" indicates that there is no OH group present in the adsorbent. Related calculations are presented in Table S5.
L	BET Surface Area of Solid Support (m^2g^{-1})	integer	1336	Obtained from literatures
M	Pore Volume of Solid Support (cm^3g^{-1})	integer	1336	Obtained from literatures
N	Average Pore Diameters (nm)	integer	1289	Obtained from literatures or estimated based on Formula 12. "-" represents missing data
O	Relative Humidity (%)	integer	1336	Obtained from literatures
P	CO_2 Concentration (vol%)	integer	1336	Obtained from literatures
Q	Flow Rate (mL/min)	integer	1158	Obtained from literatures or calculated
R	Adsorption Time (min)	integer	1323	"-" represents missing data
S	Adsorption Temperature ($^{\circ}\text{C}$)	integer	1336	Obtained from literatures
T	CO_2 Capacity (mmol/g)	integer	1335	Obtained from literatures
U	Amine Efficiency (mmol/mmol)	integer	1335	Obtained from literatures or calculated
V	Time to Half Saturation (min)	integer	691	"-" represents missing data. Related calculations are presented in Table S6
W	Time to 90% Saturation (min)	integer	607	"-" represents missing data. Related calculations are presented in Table S6
X	Desorption Temperature ($^{\circ}\text{C}$)	integer	216	"-" represents missing data
Y	Weight Loss Stability (%)	integer	53	"-" represents missing data
Z	Capacity Loss Stability (%)	integer	216	"-" represents missing data
AA	Cyclic Number	integer	216	"-" represents missing data
AB	Heat of adsorption (kJ/mmol)	integer	42	"-" represents missing data
AC	DOI	string	1336	The reference articles corresponding to each data record

Table 4. Descriptions of the dataset of amine-impregnated solid adsorbents containing 28 features and DOI of the data sources.

Data Records

Data storage and data source. The full dataset on amine-impregnated solid adsorbent for CO_2 capture is accessible as an XLSX file at Figshare, a public online repository⁶⁷. This dataset contains 1,336 data points (Number of rows within the Excel sheet) gathered from 52 published articles and supporting information which results in a total of 29,857 records (Number of cells with available data)^{41,43,54,55,59,61–64,68–110}. There are 28 features (Number of columns within the Excel sheet) to describe each data point, however, these features are not equally reported. Some operational parameters and DAC performance indicators such as weight loss stability and heat of adsorptions are often missing. The column name and related information to each column are summarised in Table 4.

These 28 features were mainly recorded from texts, extracted from figures or derived from equations. There are approximately 22% of the M_n data obtained from GPC measurement and less than 5% of M_n data obtained from literature interoperation or estimations. The feature N content is obtained from both literature and calculations due to the different reporting methods used in literature (Fig. 3). Data obtained from text and figures has the highest accuracy followed by data obtained from formulas.

The data source and data processing methods used to construct the database are outlined (Fig. 4). Some chemical features are contextual and inconsistently reported, but they can be added from PubChem, derived from molecular structures, or identified from spectroscopic data. For example, some primary, secondary, and tertiary amine ratios were sourced from peer-reviewed literature¹¹¹. The M_n of LPEI ($M_w = 2500 \text{ g mol}^{-1}$)¹¹² and LPEI ($M_w = 25000 \text{ g mol}^{-1}$)¹¹³ were also sourced from existing Literature. The M_n of LPEI ($M_w = 600 \text{ g mol}^{-1}$)

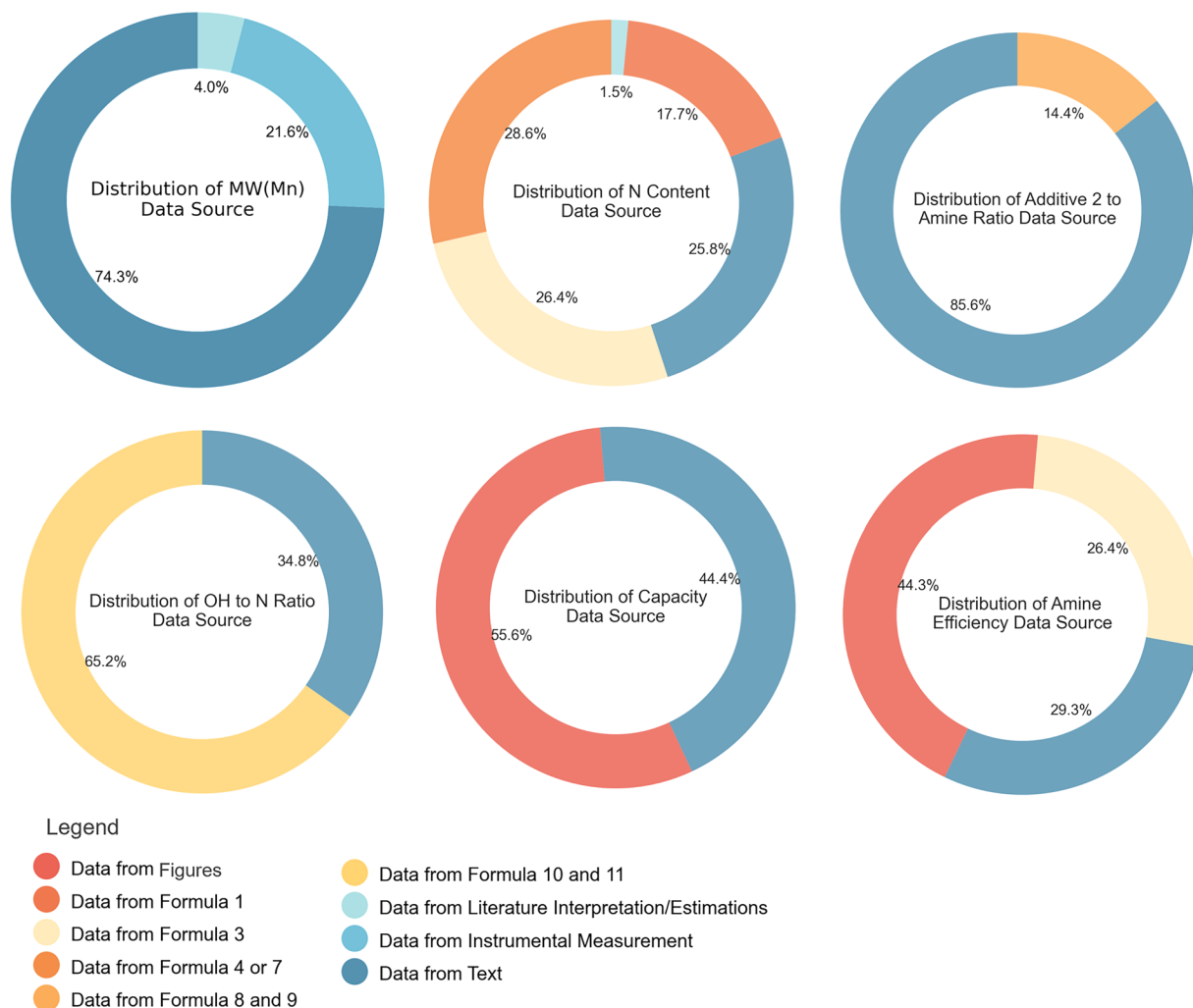


Fig. 3 The data source distribution for selected chemical features. The donut charts suggest that most of the data are sourced from the article texts or extracted from figures. A relatively high percentage of the N content and the OH to N ratio data are derived from formulas.

was estimated based on the reported polydispersity ($PDI = 1.3$) of LPEI. In addition, the M_n of polypropylene guanidine (PPG) was derived from the ion peak of mass spectroscopy⁸⁷. In other cases, the M_n of the pristine polyamine is known, however, the M_n of the product polymer after the epoxidation or guanidinylation reactions are not reported. The M_n value of the modified molecules is therefore estimated based on the provided NMR analysis as well as the number of functional groups added to the pristine polyamine^{86,93}. Even though the process of obtaining these contextual data is not formative, the exceptional cases are rare in the dataset. We present all the derived numerical data of these exceptional cases with their corresponding data processing method for reference in Fig. 4.

Information on chemical features and abbreviated chemical names. Most of the features in the dataset are numeric data with only four strings to describe the various types of porous solid supports, amines and additives. A classification of the porous solid used for amine impregnations is provided to explain the strings listed under the feature “support” (Table 5).

The 31 types of the porous solid supports were classified into 7 categories. Among all the types of solid supports, SBA-15 and MCM-41 showed the two highest frequencies in the dataset, as they are the standard solid supports for DAC research (Fig. 5). The hierarchical silica supports and polyresins have been developed more recently.

There are 31 types of amines (Fig. 6) and 20 types of additives included in the dataset (Fig. 7). The structure information of these molecules is provided for future references.

As indicated by the donut charts and box plots, PEI and TEPA are the most extensively studied amines for CO_2 capture which exhibits wide ranges of CO_2 capacity features (Fig. 8a). Apart from the molecular structure, the variation in the molecular weight of polyamines should also be considered where specific studies were carried to see how M_n of BPEI influence the performance of amine-impregnated solid adsorbents. BPEI with M_n of 600 g/mol from Sigma Aldrich is a commonly used polyamine for CO_2 capture under various conditions, as indicated by the donut chart (Fig. 8b). So far, only PEI and TEAP have been used separately with additives. There

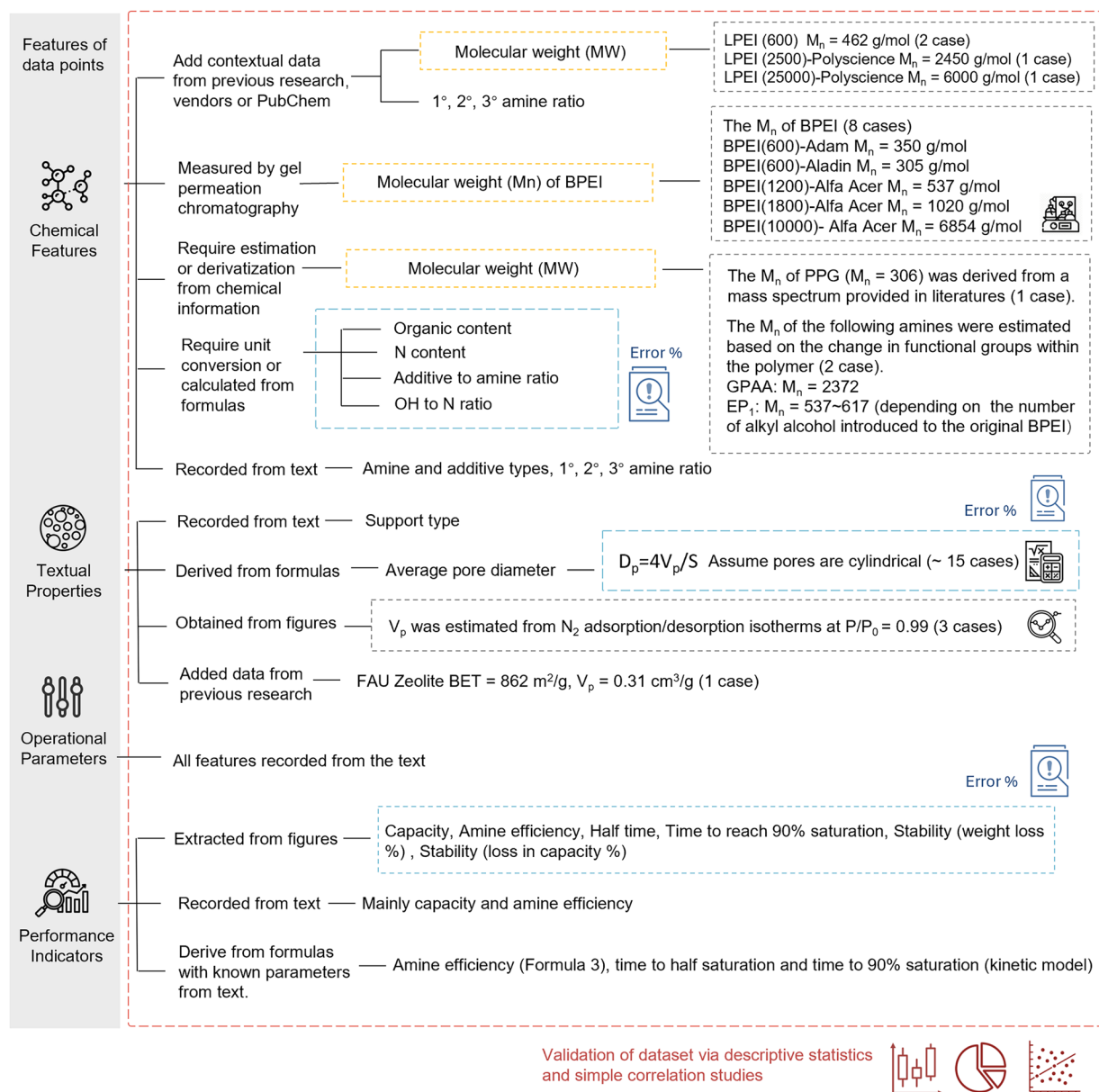


Fig. 4 The scheme for data collection and data derivatization process with selected validation methods and presentation of the exceptional cases.

No	Categories of the Support	Solid Support Name from the Dataset
1	2D siliceous materials	SBA-15, MCM-41 (pore-expanded MCM-41), W-CFA-SBA-15 and R-CFA-SBA-15 (coal fly ash derived wheat like SBA-15 and rod like SBA-15).
2	3D siliceous materials	MCM-48 with 3D-cubic pore structure, MCM-36, MCF (mesocellular silica foam), MMSN (micro-mesoporous silica nanoparticles), MMON (micro-mesoporous organosilica nanoparticles), MPS (commercial mesoporous silica: CARiACT G10 HPV, silica-Sipernat 306 and disordered mesoporous silica), BHMS (hierarchical silica bimodal meso/macroporosity), THMS (hierarchical silica bimodal meso/macroporosity), SA (silica aerogel), FS (fumed silica), NS (nano silica).
3	Carbonaceous materials	AC (activated carbon), CNT (carbon nanotube), CA (carbon aerogel), MPC (mesoporous carbon).
4	3D Aluminium oxides	Mesoporous γ -Al ₂ O ₃
5	Polar Resins	PREXAD7, PREHP2MG, PREDA201, PREHPD450, MF (meso-macroporous melamine formaldehyde).
6	Non-polar Resins	NPREXAD4, NPRED4020, NPREHP20
7	Zeolite and MOF	FAU Zeolite, ZN (zeolite nanotube) and MIL-101(Cr).

Table 5. The categorization and explanation on the strings of porous solid supports.



is one instance where TEPA is impregnated with two types of additives onto the solid support, so the additive “CTAC” is present in a very small portion under the feature “Amine 3 and Additive 3” (Fig. 8c). These box plots are used to describe the distribution of the dataset rather than the analysis of the adsorbent properties, since many of the operational parameters aren’t considered.

The data collection process was conducted by two lab chemists specializing in DAC research. Consequently, the datasets primarily focused on the structure and composition of amine-functionalized adsorbents, as well as the CO₂ capture performances of powder adsorbents at a milligram scale. Interaction with GPT-4 was employed to assess the completeness of the data source. Additionally, the data derivatization process helped identify human errors and typos in the Excel file. The data used for validation, along with a draft dataset from the data collection processes, is presented in Tables S7, S8.

The time to half saturation and the time to reach 90% saturation data extracted from the exponential curves can be validated by comparing them with the calculated values from the adsorption kinetic models (Table 3). Notably the kinetic model can only be used for validation when it is strongly fitted to the experimentally measured adsorption isotherm. Although the coefficient (R^2) of the model exceeds 0.99 which indicates an overall strong fit, it merely signifies the average precision of the kinetic model in characterizing the experimental adsorption process. The discrepancies among specific data points remain possible. Nonetheless, the comparison between the data extracted from exponential curves and the data calculated from kinetic models still provides an indication of the reliability and chemical significance of the extracted data. The extracted graphical data and

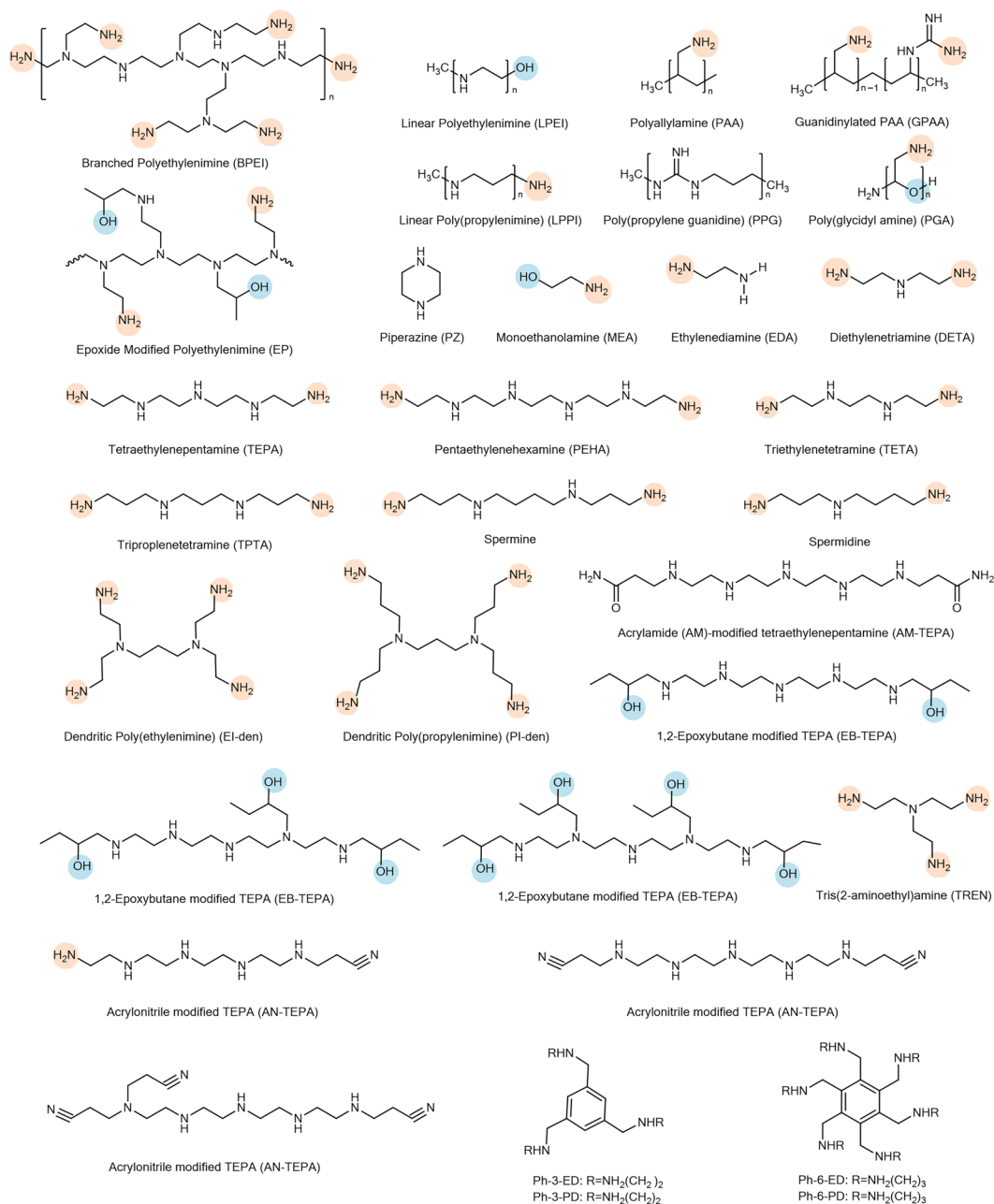


Fig. 6 The molecular structure of amine 1 in the dataset. The primary amines are highlighted with orange and the hydroxy groups are highlighted with blue.

the data calculated from kinetic models from 4 out of the 52 articles were utilized to determine the percentage error associated with data extraction from the line graphs (Table 6)^{74,77,104,106}.

The N content is primarily derived based on the theoretical moles of N atoms from each of the polyamines and small amines. Since the molecular structures of these molecules are known and most of the organic content data are provided, the N content of each data point can be easily obtained with reasonable accuracy (Table 6). The estimated value was validated by randomly selecting a dataset with reported N content or N content calculated from **Formula 3** and comparing it with the N content derived from the molecular structure of amines within the randomly selected dataset (**Formula 4** or **Formula 7**).

During the data collection we noticed several articles didn't report the OH to N ratio when OH group is present, however, the presence of OH group is known to provide positive impact on the amine efficiency, adsorption kinetic and the optimal desorption temperature of adsorbents⁷⁵. Therefore, the OH to N ratio is derived from the original articles. As a drawback, the number of articles in mixed amines and additives is limited. Only three cases reported the N to OH ratio, so the reliability and accuracy of the OH to N ratio is validated with a small dataset (Table 6)^{75,76,81}. The OH to N ratio within those studies were calculated using **Formula 10** and **Formula 11**

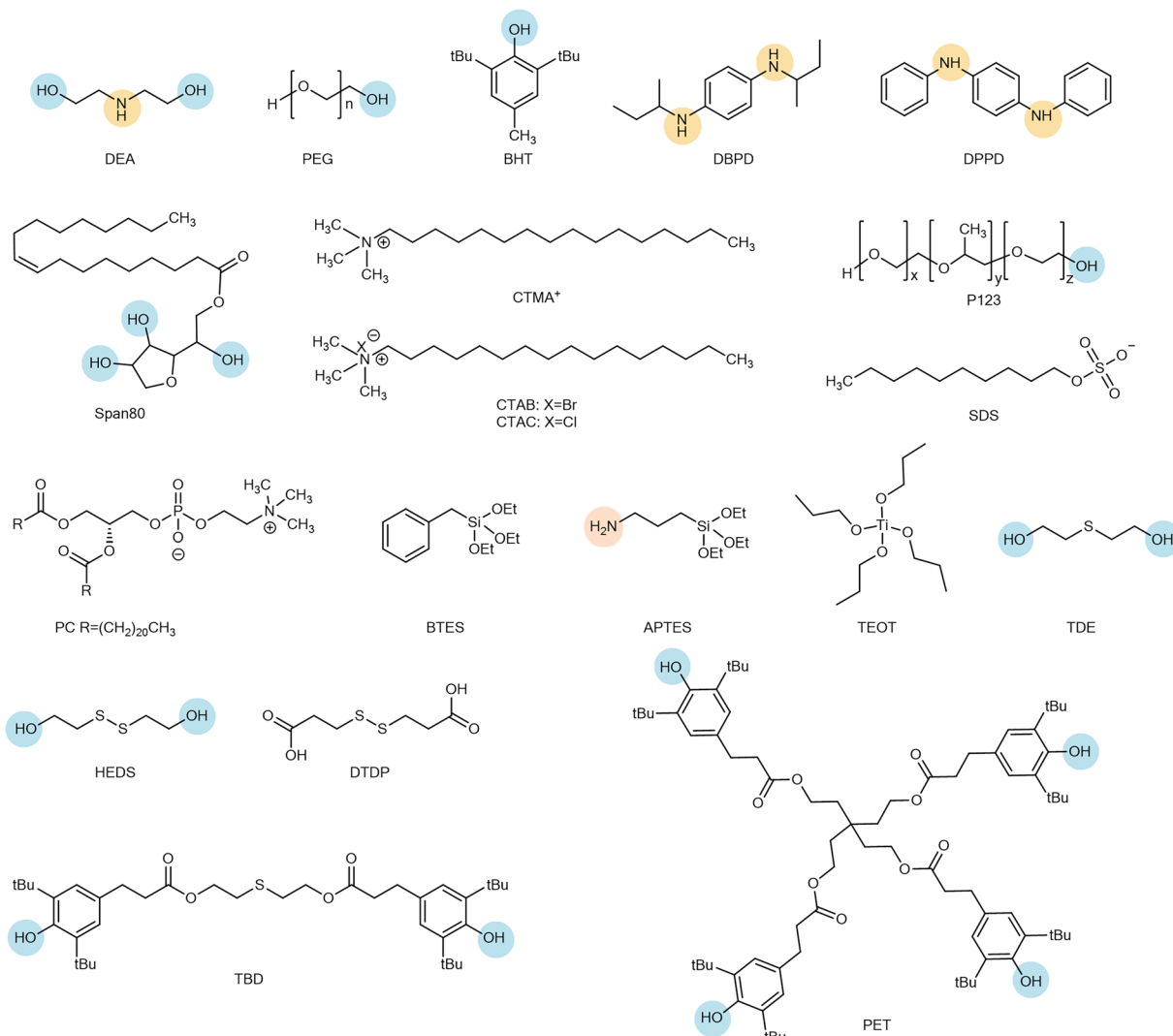


Fig. 7 The molecular structure of the additives (additive 1 to additive 3) in the dataset. The primary amines are highlighted in orange, secondary amines are highlighted in yellow and hydroxy groups are highlighted in blue.

and then compared with the reported data. Even though the size of validation sample is small, the absolute and percentage errors obtained can still suggest the correctness and reliability of the strategy to obtain the OH to N ratio.

The average pore diameters for pristine supports were estimated from 12 articles using **Formula 12**. While the formula is applicable to solid supports with assumed cylindrical pore geometry, the mathematical relationship cannot accurately represent the exact geometries of silica supports from various sources and preparation conditions. To assess the associated percentage error, we identified 7 articles that reported average pore diameters based on assumed cylindrical shapes. The average pore diameters from these articles were calculated using **Formula 12** and compared to the reported values, resulting in an average percentage error of 20.8% (Table 6). Although this percentage error is relatively high, the estimated pore sizes remain within an acceptable range for both commercial and lab-made SBA-15, MCM-41, and MCM-48.

Overall, the percentage errors suggest a reasonable accuracy for the method used to obtain the final dataset. The errors for kinetic data are probably due to the inaccuracies in the kinetic models under different operational parameters. These inaccuracies lead to more deviations when compared to the plotted measurement data, and consequently, higher percentage errors when comparing the model data with extracted data from the plot. When there is a noticeable difference between the extracted experimental data from the plot and the model data, the extracted experimental data from the plot is retained within the dataset. In addition, most of the stability data (percentage loss in adsorbent weight and percentage loss in CO₂ capacity) were obtained from bar graphs. There are limited data from the text to validate the accuracy of extracted numbers from the bar graph, however, many of the stability data were reported in a rounded number or reported as “no significant reduction in the CO₂ capacity of the adsorbent.” It also suggests that inaccuracies in the data extraction process have less influence on the actual stability of the adsorbent.

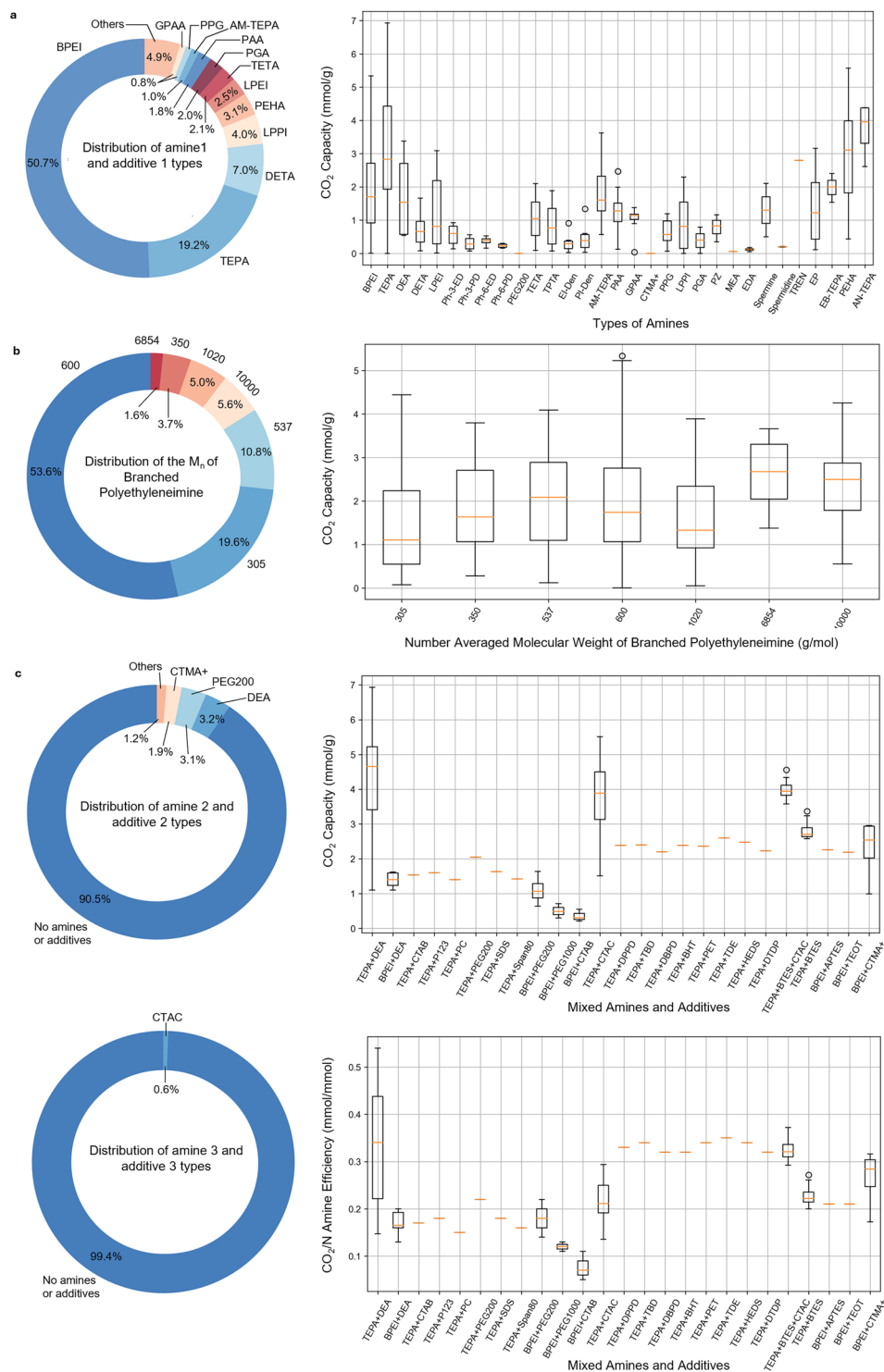


Fig. 8 The donut charts and box plots present the distributions of amines and additives used for the impregnation of porous solid supports in the dataset and their corresponding CO₂ capacities. **(a)** The distribution of amine types (donut chart) and the distribution of CO₂ capacity data for each amine type (box plot). **(b)** The distributions of M_n values for BPEI (donut chart) and the distribution of CO₂ capacity data for each M_n value (box plot). **(c)** The percentage for different types of additive 2 and additive 3 (donut chart) and the distribution of CO₂ capacity and amine efficiency for all combinations of amines and additives within the dataset (box plot).

Validation of the dataset via statistical analysis. The statistical analysis not only demonstrates the general information and relationship within the dataset but also identifies the outliers and human errors. The

Descriptions	Sample Size	Average Percentage error
The data extracted from scatter plot	205	1.3%
The data extracted from line graph	92	6.1%
The estimated N content	167	2.7%
The estimated OH to N ratio	12	3.1%
The estimated pore diameters	12	20.8%

Table 6. The average percentage errors for the comparisons between the extracted graphical data or derived data with the data reported in texts.

data cleaning process and statistical analysis were carried out with Python and generally involved the following steps: 1) Check the empty cells within the dataset. 2) Check the number of distinct data points and their frequency in each feature. 3) Check the presence of data with frequency lower than 10 for each feature. 4) Check for the expected match between data points and features. In this way the outliers can be identified and corrected manually. The statistical analysis was conducted by three data scientists to check the technical reusability of the data. The data cleaning process was also checked several times before the final analysis.

This dataset included the DAC performance of amine-functionalised solid adsorbent for the first time and the performance is expected to be distinct from high-concentration CO₂ capture. Three defined CO₂ concentration ranges were plotted against each of the six DAC performance indicators in a violin plot (Fig. 9a). The extreme values indicated by the boxplots and violin plot were checked manually. Notable, the recently developed amine-functionalised solid adsorbents adopt the highest CO₂ capacity at 5.8 mmol CO₂/g under DAC conditions (0.04 vol% of CO₂) which is comparable to the CO₂ capacity under flue gas conditions (15–25 vol% of CO₂). The distribution and range of CO₂ capacity are generally in agreement with the amine efficiency suggesting the reliability of the database. The violin plot of CO₂ concentrations against the time to half saturation showed an extreme maximum at 75%–100% volume of CO₂ (770 min). In general, the maximum value of the half time should be around one third of the time to 90% saturation. This value was checked, and the original study did not provide the time to 90% saturation of the adsorbent. The violin plot of CO₂ concentrations against the stability of the adsorbent, indicating weight loss stability (%) of the adsorbent under flue gas conditions, has not been reported thus far. While CO₂ capacity loss stability (%) is more frequently reported, the adsorbent under DAC conditions showed less percentage loss compared to other conditions. This result may also be influenced by temperature and the number of adsorption-desorption cycles in the actual experiments.

Following the previous analysis of the private database of amine-functionalised SBA-15 and MCM-41 by Yildiz *et al.*, the CO₂ capacity was plotted against the percentage quantity of primary, secondary and tertiary amines within the solid adsorbents (Fig. 9b)³⁷. Based on the given information, the private datasets contained different ranges of data and data sourced from different time periods. However, the box plots from the previous database showed maximum values that are similar to the violin plots. Both plots suggest that the presence of primary amine contributes to high amine efficiencies. The violin plot also indicated the secondary amines is in comparative quantity to primary amines contributing to the overall amine efficiencies. This statistical result may be influenced by the instrumental method and the actual chemical qualities in use. Nonetheless this comparison has validated the reproducibility of these types of datasets and demonstrated their potential for the analysis of amine-functionalised solid adsorbents. The violin plots also illustrate the data distribution of the textural properties of solid supports against CO₂ capacity (Fig. 9c). It suggested that the solid support with pore volume between 3 to 6 cm³/g and pore size larger than 20 nm provides positive effects to the CO₂ capacity of the adsorbents.

A correlation heatmap was generated to illustrate the relationships among the numerical features within the dataset (Fig. 10a). The heatmap reveals the direction and relative strength of correlations between pairs of numerical features; however, no strong correlations with R² values exceeding 0.99 are observed. Generally, some features with known correlations exhibit R² values above 0.70, such as organic content and CO₂ capacity, as well as CO₂ capacity and amine efficiency. In contrast, some expected correlations, such as those between CO₂ capacity and humidity, are not evident. The weak correlations among the selected numerical features may be influenced by non-linearities and multicollinearity among the parameters, as the heatmap only indicates linear correlation strengths between pairs of variables. Based on the strongest observed correlation, represented by an R² value of 0.79, a scatter density diagram was created to plot amine efficiency against CO₂ capacity for different CO₂ concentrations (Fig. 10b). The dot plot data tend to exhibit a linear relationship, with some exceptions.

The interaction of CO₂ with amine-functionalised adsorbents is affected by multiple chemical and physical processes, including adsorption kinetics, thermodynamics, and specific interactions between CO₂ and functional groups. The strength of correlations in complex chemical properties may not be accurately observed due to the intercorrelation of many factors, complicating the determination of the independent effects between individual variables. More complex correlations should be further explored through modeling studies or explanatory machining learning with lab-based investigations in future research¹¹⁴. For example, Serna-Guerrero *et al.* examined how desorption pressure (PD), desorption temperature (TD), purge gas flow (FD), and their respective interactions affected the performance of the adsorbent through a statistical analysis utilizing a 2³ factorial design. The statistical model used enables objective assessment of the impact of each parameter and allows for the analysis of interactions that may not be apparent from experimental observations.

Furthermore, the weaker correlation strength may be influenced by limitations within the dataset, such as its size and diversity. For instance, the correlation between weight loss percentage and pore size of the pristine solid

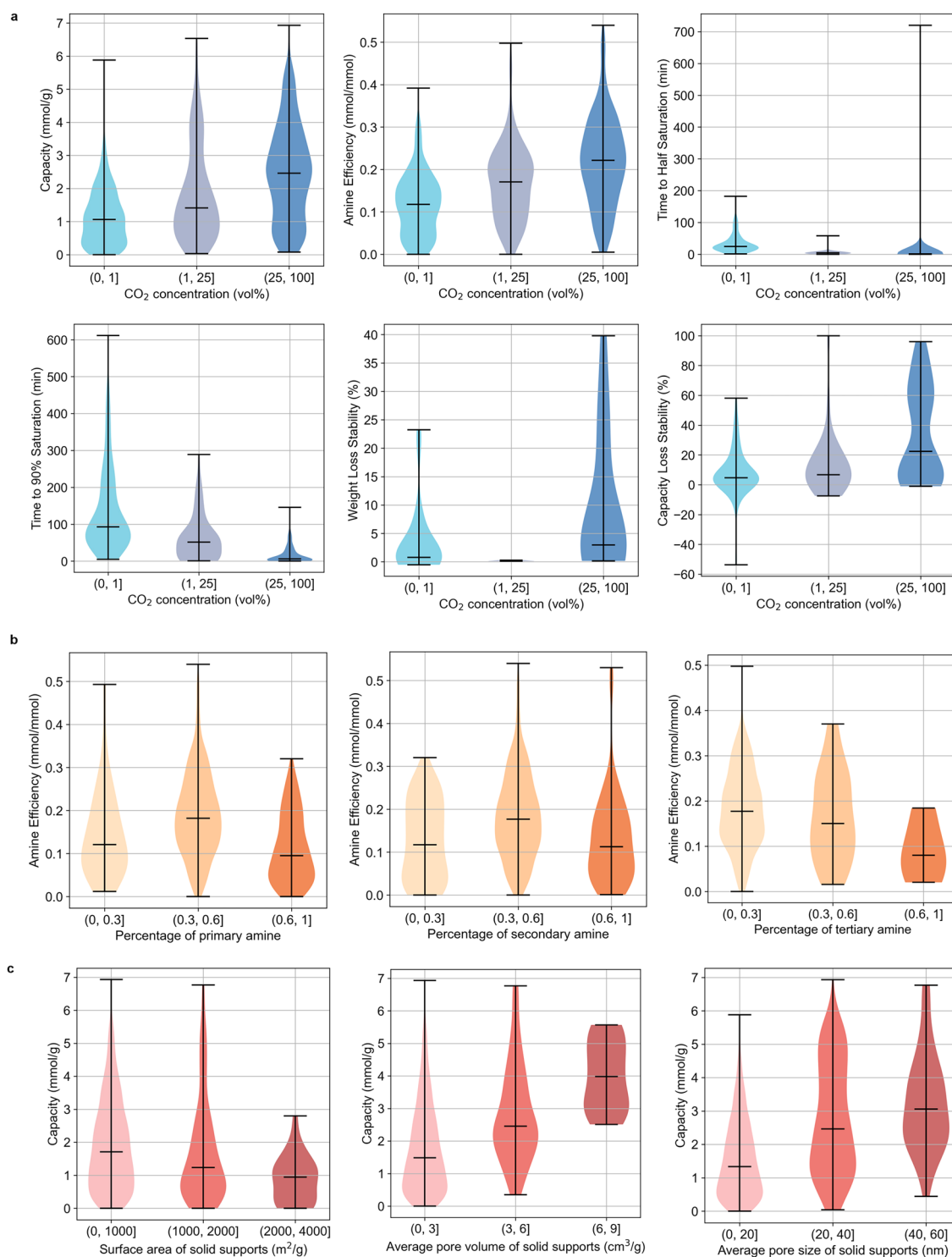


Fig. 9 Violin plots of selected adsorbent characteristics and operational parameters (CO₂ concentrations) against adsorbent performance indicators. **(a)** The ranges of CO₂ concentration against six features of performance indicators (CO₂ capacity, amine efficiency, time to half saturation, time to 90% saturation, weight loss stability (%), capacity loss stability (%)). **(b)** The fraction of primary, secondary and tertiary amines within adsorbents against amine efficiency. **(c)** The textural properties of pristine solid supports against CO₂ capacity.

supports appears relatively strong, which is rarely observed in experimental studies⁷⁹. With only 53 data points for weight loss capacity, the dataset may not adequately capture the full range of variability, leading to potentially misleading correlations. Additionally, since this dataset was manually compiled from various research works, inaccuracies in the measurement or reporting of parameters can introduce noise, obscuring true relationships.

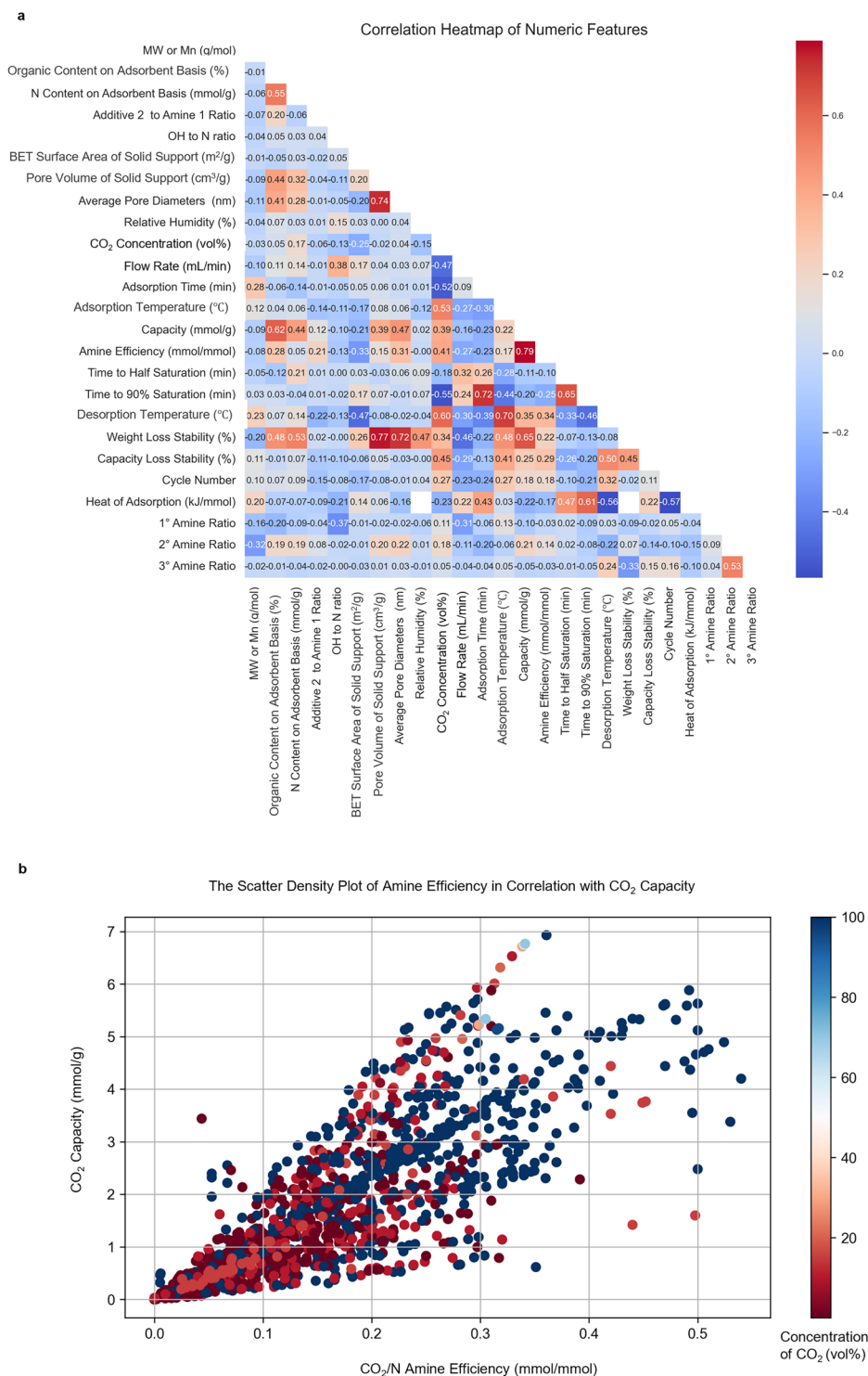


Fig. 10 Correlation heatmap of features within the dataset. **(a)** The heatmap illustrates the general strength of linear correlations among each numerical feature in the dataset. **(b)** The scatter density plot of amine efficiency against CO₂ capacity.

The variables may also be influenced by unobserved factors or interactions not included in the heatmap analysis. These considerations highlight the importance of dataset quality and comprehensiveness for effectively applying AI technology in material science.

Comparisons with a private dataset. There is only one other manually curated amine-functionalized solid adsorbent database containing reported experimental values, as introduced in the 2019 research by Yıldız *et al.*³⁷

Variables of Dataset	Yildiz <i>et al.</i> Dataset	This Dataset
No. of data source	28 publications	52 publications
No. of adsorbent class	2	1
No. of solid support	2	31
No. of amine	17	31
Temperature rang (°C)	0~100	−30~100
CO ₂ partial pressure range (bar)	0.01~1	0.01~1
Relative Humidity	0%~78%	0%~100%
No. of DAC performance indicators	2	6
Excluded features	1) OH to N ratio 2) Organic Content 3) N Content 4) Types of additives 5) BET surface area	1) Amine density 2) Calcination 3) Remaining pore volume

Table 7. The general comparison of our database with the database created by Yıldız *et al.*

Since the dataset is private, it is not possible to make a statistical comparison with the dataset from this work. However, a contextual comparison would further suggest how the database from this study can contribute to the existing one (Table 7). The major difference between the two datasets is that Yıldız *et al.*'s dataset contained amine-functionalised solid adsorbent made from impregnation (Class I) and grafting (Class II) methods. This dataset is an expansion with a specific focus on the adsorbent made from the wet impregnation method which generally shows better performance under DAC conditions¹⁴. The other major differences would be the types of amines and supports involved in the database. Yıldız *et al.*'s dataset only contained two solid supports, the SBA-15 and MCM-41. It allows efficient classification and comparison for the preparation methods of amine-functionalised solid adsorbent. Our dataset incorporated a diverse type of solid support, amine and additive which highlights more on its potential for prediction and selection of ideas from scientists. Finally, our dataset focuses on the performance of amine-impregnated solid adsorbent from 0.0004 bar to 1 bar of CO₂ partial pressure, enabling the exploration and evaluation of adsorbents for the negative emission DAC technology.

Usage Notes

Incorporation into machine learning models. To present a transparent process for the usage of this dataset and to prove its usability, we have trained a Random Forest model (RF) and an Extreme Gradient Boosting model (XGBoost) using this dataset. The code for data cleaning process and training of the ML model is uploaded on Figshare. Following the common practice^{34–36}, the input includes chemical features of amines, textural properties of the pristine support and operational parameters. The CO₂ capacity was used as an example output. We construct a 10-fold cross-validation experiment for each ML model. The values of root mean squared error (RMSE) of RF and XGB are 0.42 mmol/g and 0.53 mmol/g, respectively. These values are comparable with the previously reported RMSE value of 0.68 for an XGB model trained from a database of biomass waste-derived porous carbon for CO₂ capture³⁵. Moreover, the relative root mean squared error (RRMSE) values for RF and XGB are both 0.02, indicating that accurate machine learning models can be trained using our dataset.

It is encouraged to explore the initial ideas for the engineering of any amine-impregnated solid adsorbent using the trained ML model. The input data can be easily entered into the model for a quick discovery of adsorbent performance, ranging from DAC conditions to pure CO₂ conditions. Apart from the provided machine learning models, the dataset still contains statistical value and provides other research opportunities. In addition, the dataset can be expanded and upgraded by experimental results or data generation technologies in the future.

Limitations and outlook. The dataset provided in this study presents several potential limitations that should be kept in mind for future research. Firstly, there are potential bias from the data collection methods, as this database is derived from previous literature. Consequently, some selected chemical features may not fully capture the complexities of real experimental settings, particularly as certain features related to instrument configurations have been excluded. Furthermore, many samples are subject to experimental and measurement errors. Notably, standard errors associated with the measurement of DAC performance indicators are inconsistently reported and are not accounted for within the dataset. The errors inherent in our dataset may be comparable to those encountered under actual experimental conditions. Given both the limited data size and chemical features, accurately predicting specific values from this dataset poses significant challenges. Thus, it is imperative to explore how each feature within the dataset may be influenced by experimental measurements and to develop methodologies that align literature-based datasets with real-world conditions. The applications of AI technologies for information extraction and simulation techniques for data generations would further enrich the datasets, avoid human errors and bridge the gap between lab research and data scientists.

The diversity of chemical descriptors significantly affects the accuracy and capacity of artificial intelligence (AI) models⁴⁰. While this dataset facilitates the exploration of time-saving ideas in laboratory settings, the selected features may limit the discovery of adsorbents with novel textual properties or a wider range of polyamines. As a result, models trained on this dataset may be restricted in generative AI applications, relying primarily on the patterns present in the data. In addition, the representation of solid supports and amines could extend beyond the current textual formats to include Simplified Molecular Input Line Entry System (SMILES) notation or other novel descriptors, promoting innovation and enhancing understanding of amine-functionalized adsorbents.

The selected chemical features may not fully capture the adsorption phenomena of amine-impregnated adsorbents, overlooking key thermodynamic, kinetic, and economic parameters. Including these factors would facilitate a more comprehensive analysis of the adsorbents' performance and viability in practical applications¹¹⁵. The absence of energy-related data limits the dataset's utility for simulation studies and the engineering design of DAC adsorbents. Consequently, the trained machine learning model may not be suitable for selecting scalable adsorbents for DAC applications. Given the complexity of CO₂ adsorption across various scales, the chosen chemical features also impact on the results of correlation studies aimed at elucidating the mechanisms associated with amine-impregnated adsorbents.

Despite these limitations, this dataset highlights the potential for integrating machine learning with experimental research data, underscoring the necessity for uniform data reporting and the significance of small datasets. Interdisciplinary research efforts are essential for effectively applying AI technologies to the development of amine-functionalized adsorbents.

Code availability

The code for statistical analysis and usage guide was uploaded on Figshare together with the dataset.

Received: 9 August 2024; Accepted: 22 April 2025;

Published online: 01 May 2025

References

- United Nations Climate Change. *The Paris Agreement*, <https://unfccc.int/process-and-meetings/the-paris-agreement> (2016).
- Gür, T. M. Carbon dioxide emissions, capture, storage and utilization: Review of materials, processes and technologies. *Prog. Energy Combust. Sci.* **89**, 100965, <https://doi.org/10.1016/j.pecs.2021.100965> (2022).
- Perugini, L. *et al.* Emerging reporting and verification needs under the Paris Agreement: How can the research community effectively contribute? *Environ. Sci. Policy* **122**, 116–126, <https://doi.org/10.1016/j.envsci.2021.04.012> (2021).
- United Nations Framework Conventions on Climate Change. *Report of the Conference of the Parties on its twenty-eighth session, held in the United Arab Emirates from 30 November to 13 December 2023*, https://unfccc.int/sites/default/files/resource/cp2023_11a01_adv.pdf (2024).
- Yan, Y. *et al.* Harnessing the power of machine learning for carbon capture, utilisation, and storage (CCUS)—a state-of-the-art review. *Energy Environ. Sci.* **14**, 6122–6157, <https://doi.org/10.1039/D1EE02395K> (2021).
- Wang, E. *et al.* Reviewing direct air capture startups and emerging technologies. *Cell Rep. Phys. Sci.* **5**, <https://doi.org/10.1016/j.xcrp.2024.101791> (2024).
- Rahimi, M., Moosavi, S. M., Smit, B. & Hatton, T. A. Toward smart carbon capture with machine learning. *Cell Rep. Phys. Sci.* **2**, <https://doi.org/10.1016/j.xcrp.2021.100396> (2021).
- Li, G. *et al.* Integration of Carbon Dioxide Removal (CDR) Technology and Artificial Intelligence (AI) in Energy System Optimization. *Processes* **12**, 402, <https://doi.org/10.3390/pr12020402> (2024).
- Sehnal, D. *et al.* BinaryCIF and CIFTools—Lightweight, efficient and extensible macromolecular data management. *PLoS Comput. Biol.* **16**, e1008247, <https://doi.org/10.1371/journal.pcbi.1008247> (2020).
- Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 1–9, <https://doi.org/10.1038/sdata.2016.18> (2016).
- Sriram, A. *et al.* The Open DAC 2023 dataset and challenges for sorbent discovery in direct air capture. *ACS Cent. Sci.* <https://doi.org/10.1021/acscentsci.3c01629> (2024).
- Xu, X., Song, C., Andresen, J. M., Miller, B. G. & Scaroni, A. W. Novel polyethylenimine-modified mesoporous molecular sieve of MCM-41 type as high-capacity adsorbent for CO₂ capture. *Energy & Fuels* **16**, 1463–1469, <https://doi.org/10.1021/ef020058u> (2002).
- Darunte, L. A., Walton, K. S., Sholl, D. S. & Jones, C. W. CO₂ capture via adsorption in amine-functionalized sorbents. *Curr. Opin. Chem. Eng.* **12**, 82–90, <https://doi.org/10.1016/j.coche.2016.03.002> (2016).
- Zhu, X. *et al.* Recent advances in direct air capture by adsorption. *Chem. Soc. Rev.* **51**, 6574–6651, <https://doi.org/10.1039/d1cs00970b> (2022).
- Yan, H. *et al.* Investigation of CO₂ adsorption performance of amine impregnated adsorbents using amine-support matching strategies. *Sep. Purif. Technol.* **310**, 123178, <https://doi.org/10.1016/j.seppur.2023.123178> (2023).
- Yan, H. *et al.* Investigation of CO₂ adsorption performance of amine impregnated adsorbents using amine-support matching strategies. *Separation and Purification Technology* **310**, 123178, <https://doi.org/10.1016/j.seppur.2023.123178> (2023).
- Harlick, P. J. & Sayari, A. Applications of pore-expanded mesoporous silica. 5. Triamine grafted material with exceptional CO₂ dynamic and equilibrium adsorption performance. *Industrial & Engineering Chemistry Research* **46**, 446–458, <https://doi.org/10.1021/ie060774+> (2007).
- Mafra, L. *et al.* Structure of chemisorbed CO₂ species in amine-functionalized mesoporous silicas studied by solid-state NMR and computer modeling. *Journal of the American Chemical Society* **139**, 389–408, <https://doi.org/10.1021/jacs.6b11081> (2017).
- Builes, S. & Vega, L. F. Understanding CO₂ capture in amine-functionalized MCM-41 by molecular simulation. *The Journal of Physical Chemistry C* **116**, 3017–3024, <https://doi.org/10.1021/jp210494f> (2012).
- Serna-Guerrero, R., Belmabkhout, Y. & Sayari, A. Modeling CO₂ adsorption on amine-functionalized mesoporous silica: 1. A semi-empirical equilibrium model. *Chemical Engineering Journal* **161**, 173–181, <https://doi.org/10.1016/j.cej.2010.04.024> (2010).
- Belmabkhout, Y., Serna-Guerrero, R. & Sayari, A. Amine-bearing mesoporous silica for CO₂ removal from dry and humid air. *Chemical Engineering Science* **65**, 3695–3698, <https://doi.org/10.1016/j.ces.2010.02.044> (2010).
- Cao, Y. *et al.* Molecular dynamic simulations and quantum chemical calculations of adsorption process using amino-functionalized silica. *Journal of Molecular Liquids* **330**, 115544, <https://doi.org/10.1016/j.molliq.2021.115544> (2021).
- Builes, S., Lopez-Aranguren, P., Fraile, J., Vega, L. F. & Domingo, C. Analysis of CO₂ adsorption in amine-functionalized porous silicas by molecular simulations. *Energy & Fuels* **29**, 3855–3862, <https://doi.org/10.1021/acs.energyfuels.5b00781> (2015).
- Moon, H. J., Carrillo, J. M. Y. & Jones, C. W. Distribution and Mobility of Amines Confined in Porous Silica Supports Assessed via Neutron Scattering, NMR, and MD Simulations: Impacts on CO₂ Sorption Kinetics and Capacities. *Acc. Chem. Res.* **56**, 2620–2630, <https://doi.org/10.1021/acs.accounts.3c00363> (2023).
- Heydari-Gorji, A. & Sayari, A. CO₂ capture on polyethylenimine-impregnated hydrophobic mesoporous silica: Experimental and kinetic modeling. *Chemical Engineering Journal* **173**, 72–79, <https://doi.org/10.1016/j.cej.2011.07.038> (2011).
- Jung, W. & Lee, K. S. Isotherm and kinetics modeling of simultaneous CO₂ and H₂O adsorption on an amine-functionalized solid sorbent. *Journal of Natural Gas Science and Engineering* **84**, 103489, <https://doi.org/10.1016/j.jngse.2020.103489> (2020).
- Wu, J., Zhu, X., Yang, F., Wang, R. & Ge, T. Shaping techniques of adsorbents and their applications in gas separation: a review. *Journal of Materials Chemistry A* **10**, 22853–22895, <https://doi.org/10.1039/D2TA04352A> (2022).

28. Serna-Guerrero, R., Da'na, E. & Sayari, A. New insights into the interactions of CO₂ with amine-functionalized silica. *Industrial & Engineering Chemistry Research* **47**, 9406–9412, <https://doi.org/10.1021/ie801186g> (2008).
29. Moon, H. J. *et al.* Underlying Roles of Polyol Additives in Promoting CO₂ Capture in PEI/Silica Adsorbents. *ChemSusChem* **17**, e202400967, <https://doi.org/10.1002/cssc.202400967> (2024).
30. Min, Y. J., Ganesan, A., Realf, M. J. & Jones, C. W. Direct air capture of CO₂ using poly (ethyleneimine)-functionalized expanded poly (tetrafluoroethylene)/silica composite structured sorbents. *ACS Appl. Mater. Interfaces* **14**, 40992–41002, <https://doi.org/10.1021/acsami.2c11143> (2022).
31. Wang, S. *et al.* The amine-functionalized alumina captures CO₂ directly from the ambient air in the rotating adsorption bed. *Chem. Eng. Process. Process Intensif.* **199**, 109735, <https://doi.org/10.1016/j.cep.2024.109735> (2024).
32. Stampi-Bombelli, V., Storione, A., Grossmann, Q. & Mazzotti, M. On comparing packed beds and monoliths for CO₂ capture from air through experiments, theory, and modeling. *Industrial & Engineering Chemistry Research* **63**, 11637–11653, <https://doi.org/10.1021/acs.iecr.4c01392> (2024).
33. Johnston, K. G., Grieco, S. F., Nie, Q., Theis, F. J. & Xu, X. Small data methods in omics: the power of one. *Nature Methods* **21**, 1597–1602, <https://doi.org/10.1038/s41592-024-02390-8> (2024).
34. Fathalian, F., Aarabi, S., Ghaemi, A. & Hemmati, A. Intelligent prediction models based on machine learning for CO₂ capture performance by graphene oxide-based adsorbents. *Sci. Rep.* **12**, 21507, <https://doi.org/10.1038/s41598-022-26138-6> (2022).
35. Yuan, X. *et al.* Applied machine learning for prediction of CO₂ adsorption on biomass waste-derived porous carbons. *Environ. Sci. Technol.* **55**, 11925–11936, <https://doi.org/10.1021/acs.est.1c01849> (2021).
36. Zhang, Z. *et al.* Prediction of carbon dioxide adsorption via deep learning. *Angew. Chem.* **131**, 265–269, <https://doi.org/10.1002/ange.201812363> (2019).
37. Yıldız, M. G., Davran-Candan, T., Günay, M. E. & Yıldırım, R. CO₂ capture over amine-functionalized MCM-41 and SBA-15: Exploratory analysis and decision tree classification of past data. *J. CO₂ Util.* **31**, 27–42, <https://doi.org/10.1016/j.jcou.2019.02.010> (2019).
38. Zhang, S. *et al.* Design and optimization of solid amine CO₂ adsorbents assisted by machine learning. *ACS Sustainable Chem. Eng* **10**, 13185–13193, <https://doi.org/10.1021/acssuschemeng.2c04492> (2022).
39. Li, S., Huang, Z., Li, Y., Deng, S. & Cao, X. E. Methodology for predicting material performance by context-based modeling: A case study on solid amine CO₂ adsorbents. *Energy and AI* **20**, 100477, <https://doi.org/10.1016/j.egyai.2025.100477> (2025).
40. Karande, P., Gallagher, B. & Han, T. Y.-J. A Strategic Approach to Machine Learning for Material Science: How to Tackle Real-World Challenges and Avoid Pitfalls. *Chemistry of Materials* **34**, 7650–7665, <https://doi.org/10.1021/acs.chemmater.2c01333> (2022).
41. Chaikittisilp, W., Kim, H.-J. & Jones, C. W. Mesoporous alumina-supported amines as potential steam-stable adsorbents for capturing CO₂ from simulated flue gas and ambient air. *Energy & Fuels* **25**, 5528–5537, <https://doi.org/10.1021/ef201224v> (2011).
42. Lin, Z. How to write effective prompts for large language models. *Nature Human Behaviour* **8**, 611–615, <https://doi.org/10.1038/s41562-024-01847-2> (2024).
43. Sayari, A., Liu, Q. & Mishra, P. Enhanced adsorption efficiency through materials design for direct air capture over supported polyethylenimine. *ChemSusChem* **9**, 2796–2803, <https://doi.org/10.1002/cssc.201600834> (2016).
44. Romeo, L. M., Minguell, D., Shirmohammadi, R. & Andrés, J. M. Comparative analysis of the efficiency penalty in power plants of different amine-based solvents for CO₂ capture. *Industrial & Engineering Chemistry Research* **59**, 10082–10092, <https://doi.org/10.1021/acs.iecr.0c01483> (2020).
45. Wu, J., Zhu, X., Chen, Y., Wang, R. & Ge, T. The analysis and evaluation of direct air capture adsorbents on the material characterization level. *Chemical Engineering Journal* **450**, 137958, <https://doi.org/10.1016/j.cej.2022.137958> (2022).
46. Serna-Guerrero, R. & Sayari, A. Modeling adsorption of CO₂ on amine-functionalized mesoporous silica. 2: Kinetics and breakthrough curves. *Chemical Engineering Journal* **161**, 182–190, <https://doi.org/10.1016/j.cej.2010.04.042> (2010).
47. Nguyen, C. & Do, D. Heat of desorption measured by means of electrical heat excitation. *Langmuir* **17**, 2287–2290, <https://doi.org/10.1021/la0013291> (2001).
48. Zhang, W. *et al.* Parametric study on the regeneration heat requirement of an amine-based solid adsorbent process for post-combustion carbon capture. *Applied Energy* **168**, 394–405, <https://doi.org/10.1016/j.apenergy.2016.01.049> (2016).
49. Lashaki, M. J., Khiavi, S. & Sayari, A. Stability of amine-functionalized CO₂ adsorbents: a multifaceted puzzle. *Chemical Society Reviews* **48**, 3320–3405, <https://doi.org/10.1039/c8cs00877a> (2019).
50. Wang, Y. *et al.* Cold Temperature Direct Air CO₂ Capture with Amine-Loaded Metal–Organic Framework Monoliths. *ACS Applied Materials & Interfaces* **16**, 1404–1415, <https://doi.org/10.1021/acsami.3c13528> (2023).
51. Bollini, P., Brunelli, N. A., Didas, S. A. & Jones, C. W. Dynamics of CO₂ adsorption on amine adsorbents. 2. Insights into adsorbent design. *Industrial & engineering chemistry research* **51**, 15153–15162, <https://doi.org/10.1021/ie3017913> (2012).
52. Sayari, A., Heydari-Gorji, A. & Yang, Y. CO₂-induced degradation of amine-containing adsorbents: reaction products and pathways. *Journal of the American Chemical Society* **134**, 13834–13842, <https://doi.org/10.1021/ja304888a> (2012).
53. Min, K., Choi, W., Kim, C. & Choi, M. Oxidation-stable amine-containing adsorbents for carbon dioxide capture. *Nature Communications* **9**, 726, <https://doi.org/10.1038/s41467-018-03123-0> (2018).
54. Li, K., Jiang, J., Yan, F., Tian, S. & Chen, X. The influence of polyethyleneimine type and molecular weight on the CO₂ capture performance of PEI-nano silica adsorbents. *Appl. Energy* **136**, 750–755, <https://doi.org/10.1016/j.apenergy.2014.09.057> (2014).
55. Dao, D. S., Yamada, H. & Yogo, K. Response surface optimization of impregnation of blended amines into mesoporous silica for high-performance CO₂ capture. *Energy & Fuels* **29**, 985–992, <https://doi.org/10.1021/ef502656t> (2015).
56. Jahandar Lashaki, M. & Sayari, A. CO₂ capture using triamine-grafted SBA-15: The impact of the support pore structure. *Chemical Engineering Journal* **334**, 1260–1269, <https://doi.org/10.1016/j.cej.2017.10.103> (2018).
57. Serna-Guerrero, R., Belmabkhout, Y. & Sayari, A. Influence of regeneration conditions on the cyclic performance of amine-grafted mesoporous silica for CO₂ capture: An experimental and statistical study. *Chemical Engineering Science* **65**, 4166–4172, <https://doi.org/10.1016/j.ces.2010.04.029> (2010).
58. Serna-Guerrero, R., Belmabkhout, Y. & Sayari, A. Triamine-grafted pore-expanded mesoporous silica for CO₂ capture: effect of moisture and adsorbent regeneration strategies. *Adsorption* **16**, 567–575, <https://doi.org/10.1007/s10450-010-9253-y> (2010).
59. Kumar, D. R. *et al.* Alkyl-aryl amine-rich molecules for CO₂ removal via direct air capture. *ACS Sustain. Chem. Eng.* **8**, 10971–10982, <https://doi.org/10.1021/acssuschemeng.0c03706> (2020).
60. Potter, M. E., Cho, K. M., Lee, J. J. & Jones, C. W. Exploring the Acid Gas Sorption Properties of Oxidatively Degraded Supported Amine Sorbents. *Energy & Fuels* **33**, 1372–1382, <https://doi.org/10.1021/acs.energyfuels.8b03688> (2019).
61. Qi, L. *et al.* Role of brush-like additives in CO₂ adsorbents for the enhancement of amine efficiency. *J. Environ. Chem. Eng.* **9**, 106709, <https://doi.org/10.1016/j.jece.2021.106709> (2021).
62. Chen, Z. *et al.* Polyethylenimine-Impregnated Resin for High CO₂ Adsorption: An Efficient Adsorbent for CO₂ Capture from Simulated Flue Gas and Ambient Air. *ACS Appl. Mater. Interfaces* **5**, 6937–6945, <https://doi.org/10.1021/am400661b> (2013).
63. Liu, Z., Pudasainee, D., Liu, Q. & Gupta, R. Post-combustion CO₂ capture using polyethyleneimine impregnated mesoporous cellular foams. *Sep. Purif. Technol.* **156**, 259–268, <https://doi.org/10.1016/j.seppur.2015.10.009> (2015).
64. S Alivand, M., Habiba, U., Ghasemian, M., Askari, S. & Webley, P. A. Amine-Functionalized Meso-Macroporous Polymers for Efficient CO₂ Capture from Ambient Air. *ACS Appl. Mater. Interfaces* **16**, 17411–17421, <https://doi.org/10.1021/acsami.3c17126> (2024).
65. Rohatgi, A. Webplotdigitizer: Version 4.4 <https://automeris.io/WebPlotDigitizer> (2022).

66. Nandy, A. *et al.* MOFSimplify, machine learning models with extracted stability data of three thousand metal–organic frameworks. *Sci. Data* **9**, 74, <https://doi.org/10.1038/s41597-022-01181-0> (2022).
67. Wang, E. *et al.* A Dataset for Investigations of Amine-Impregnated Solid Adsorbent for Direct Air Capture. *figshare* <https://doi.org/10.6084/m6089.figshare.28761062> (2025).
68. Kwon, H. T. *et al.* Aminopolymer-impregnated hierarchical silica structures: unexpected equivalent CO₂ uptake under simulated air capture and flue gas capture conditions. *Chem. Mater.* **31**, 5229–5237, <https://doi.org/10.1021/acs.chemmater.9b01474> (2019).
69. Kumar, R., Ohtani, S. & Tsunoi, N. *Microporous Mesoporous Mater.* **360**, 112714, <https://doi.org/10.1016/j.micromeso.2022.111956> (2023).
70. Kumar, R., Bandyopadhyay, M., Pandey, M. & Tsunoi, N. Amine-impregnated nanoarchitectonics of mesoporous silica for capturing dry and humid 400 ppm carbon dioxide: A comparative study. *Microporous Mesoporous Mater.* **338**, 111956, <https://doi.org/10.1016/j.micromeso.2022.111956> (2022).
71. Pang, S. H., Lively, R. P. & Jones, C. W. Oxidatively-stable linear poly (propylenimine)-containing adsorbents for CO₂ capture from ultradilute streams. *ChemSusChem* **11**, 2628–2637, <https://doi.org/10.1002/cssc.201800438> (2018).
72. Darunte, L. A., Oetomo, A. D., Walton, K. S., Sholl, D. S. & Jones, C. W. Direct Air Capture of CO₂ Using Amine Functionalized MIL-101(Cr). *ACS Sustain. Chem. Eng.* **4**, 5761–5768, <https://doi.org/10.1021/acssuschemeng.6b01692> (2016).
73. Kulkarni, V., Panda, D. & Singh, S. K. Direct air capture of CO₂ over amine-modified hierarchical silica. *Ind. Eng. Chem. Res.* **62**, 3800–3811, <https://doi.org/10.1021/acs.iecr.2c02268> (2023).
74. Miao, Y., He, Z., Zhu, X., Izikowitz, D. & Li, J. Operating temperatures affect direct air capture of CO₂ in polyamine-loaded mesoporous silica. *J. Chem. Eng.* **426**, 131875, <https://doi.org/10.1016/j.ccej.2021.131875> (2021).
75. Miao, Y. *et al.* Mixed diethanolamine and polyethyleneimine with enhanced CO₂ capture capacity from air. *Adv. Sci.* **10**, 2207253, <https://doi.org/10.1002/advs.202207253> (2023).
76. He, Z. *et al.* Mixed polyamines promotes CO₂ adsorption from air. *J. Environ. Chem. Eng.* **10**, 107239, <https://doi.org/10.1016/j.jece.2022.107239> (2022).
77. Wang, Y. *et al.* Additives enhancing supported amines performance in CO₂ capture from air. *SusMat* **3**, 416–430, <https://doi.org/10.1002/sus2.141> (2023).
78. Lin, L. *et al.* Comparison of characteristics and performance between PEI and DETA impregnated on SBA-15 for CO₂ capture. *Sep. Purif. Technol.* **322**, 124346, <https://doi.org/10.1016/j.seppur.2023.124346> (2023).
79. Lin, L. *et al.* The influence of pore size and pore structure of silica-based material on the amine-modified adsorbent for CO₂ capture. *Sep. Purif. Technol.* **340**, 126735, <https://doi.org/10.1016/j.seppur.2024.126735> (2024).
80. Pang, S. H., Lee, L.-C., Sakwa-Novak, M. A., Lively, R. P. & Jones, C. W. Design of aminopolymer structure to enhance performance and stability of CO₂ sorbents: poly (propylenimine) vs poly (ethylenimine). *J. Am. Chem. Soc.* **139**, 3627–3630, <https://doi.org/10.1021/jacs.7b00235> (2017).
81. Sakwa-Novak, M. A., Tan, S. & Jones, C. W. Role of additives in composite PEI/oxide CO₂ adsorbents: enhancement in the amine efficiency of supported PEI by PEG in CO₂ capture from simulated ambient air. *ACS Appl. Mater. Interfaces.* **7**, 24748–24759, <https://doi.org/10.1021/acsami.5b07545> (2015).
82. Vu, Q. T., Yamada, H. & Yogo, K. Inhibitors of oxidative degradation of polyamine-modified silica sorbents for CO₂ capture. *Ind. Eng. Chem. Res.* **58**, 15598–15605, <https://doi.org/10.1021/acs.iecr.9b02742> (2019).
83. Qi, L. *et al.* Reinforced CO₂ capture on amine-impregnated organosilica with double brush-like additives modified. *Ind. Eng. Chem. Res.* **61**, 14859–14867, <https://doi.org/10.1021/acs.iecr.2c01484> (2022).
84. Choi, S., Gray, M. L. & Jones, C. W. Amine-tethered solid adsorbents coupling high adsorption capacity and regenerability for CO₂ capture from ambient air. *ChemSusChem* **4**, 628–635, <https://doi.org/10.1002/cssc.201000355> (2011).
85. Zhang, X., Zheng, X., Zhang, S., Zhao, B. & Wu, W. AM-TEPA impregnated disordered mesoporous silica as CO₂ capture adsorbent for balanced adsorption–desorption properties. *Ind. Eng. Chem. Res.* **51**, 15163–15169, <https://doi.org/10.1021/ie300180u> (2012).
86. Alkhabbaz, M. A., Khunsupat, R. & Jones, C. W. Guanidinylated poly(allylamine) supported on mesoporous silica for CO₂ capture from flue gas. *Fuel* **121**, 79–85, <https://doi.org/10.1016/j.fuel.2013.12.018> (2014).
87. Park, S. J., Lee, J. J., Hoyt, C. B., Kumar, D. R. & Jones, C. W. Silica supported poly (propylene guanidine) as a CO₂ sorbent in simulated flue gas and direct air capture. *Adsorption* **26**, 89–101, <https://doi.org/10.1007/s10450-019-00171-w> (2020).
88. Sujan, A. R. *et al.* Poly (glycidyl amine)-loaded SBA-15 sorbents for CO₂ capture from dilute and ultradilute gas mixtures. *ACS Appl. Polym. Mater.* **1**, 3137–3147, <https://doi.org/10.1021/acsapm.9b00788> (2019).
89. Liu, L., Wang, Q., Cho, Y.-H., Park, H.-H. & Lee, C.-H. Piperazine-impregnated silica aerogel for direct air capture of CO₂ for prevention of urea formation. *Chem. Eng. Res. Des.* **199**, 74–86, <https://doi.org/10.1016/j.cherd.2023.09.021> (2023).
90. Priyadarshini, P., Rim, G., Rosu, C., Song, M. & Jones, C. W. Direct Air Capture of CO₂ Using Amine/Alumina Sorbents at Cold Temperature. *ACS Environmental Au* **3**, 295–307, <https://doi.org/10.1021/acsenvironau.3c00010> (2023).
91. Rim, G. *et al.* Sub-Ambient Temperature Direct Air Capture of CO₂ using Amine-Impregnated MIL-101(Cr) Enables Ambient Temperature CO₂ Recovery. *JACS Au* **2**, 380–393, <https://doi.org/10.1021/jacsau.1c00414> (2022).
92. Cogswell, C. F. *et al.* Effect of Pore Structure on CO₂ Adsorption Characteristics of Aminopolymer Impregnated MCM-36. *Langmuir* **31**, 4534–4541, <https://doi.org/10.1021/la505037f> (2015).
93. Jo, D. H., Jung, H., Jeon, S. & Kim, S. H. Effect of amine structure on CO₂ adsorption of modified poly (ethyleneimine)-impregnated mesostructured silica sorbents. *BCSJ* **89**, 1462–1469, <https://doi.org/10.1246/bcsj.20160286> (2016).
94. Jia, X. *et al.* Direct air CO₂ capture using coal fly ash derived SBA-15 supported polyethylenimine. *CCST* **10**, 100167, <https://doi.org/10.1016/j.ccst.2023.100167> (2024).
95. Chen, C., Xu, H., Jiang, Q. & Lin, Z. Rational design of silicas with meso-macroporosity as supports for high-performance solid amine CO₂ adsorbents. *Energy* **214**, 119093, <https://doi.org/10.1016/j.energy.2020.119093> (2021).
96. Ahmadian Hosseini, A. & Jahandar Lashaki, M. A comprehensive evaluation of amine-impregnated silica materials for direct air capture of carbon dioxide. *Sep. Purif. Technol.* **325**, 124580, <https://doi.org/10.1016/j.seppur.2023.124580> (2023).
97. Dao, D. S., Yamada, H. & Yogo, K. Enhancement of CO₂ Adsorption/Desorption Properties of Solid Sorbents Using Tetraethylenepentamine/Diethanolamine Blends. *ACS Omega* **5**, 23533–23541, <https://doi.org/10.1021/acsomega.0c01515> (2020).
98. Short, G. N. *et al.* Single-Walled Zeolitic Nanotubes: Advantaged Supports for Poly(ethylenimine) in CO₂ Separation from Simulated Air and Flue Gas. *JACS Au* **3**, 62–69, <https://doi.org/10.1021/jacsau.2c00553> (2023).
99. Meng, Y. *et al.* Insights into the Critical Role of Abundant-Porosity Supports in Polyethylenimine Functionalization as Efficient and Stable CO₂ Adsorbents. *ACS Appl. Mater. Interfaces.* **13**, 54018–54031, <https://doi.org/10.1021/acsami.1c11732> (2021).
100. Meng, Y. *et al.* Comprehensive study of CO₂ capture performance under a wide temperature range using polyethylenimine-modified adsorbents. *Journal of CO₂ Utilization* **27**, 89–98, <https://doi.org/10.1016/j.jcou.2018.07.007> (2018).
101. Park, S. *et al.* Thermal Stability Enhanced Tetraethylenepentamine/Silica Adsorbents for High Performance CO₂ Capture. *Ind. Eng. Chem. Res.* **57**, 4632–4639, <https://doi.org/10.1021/acs.iecr.7b04912> (2018).
102. Huang, K., Liu, F., Fan, J.-P. & Dai, S. Open and Hierarchical Carbon Framework with Ultralarge Pore Volume for Efficient Capture of Carbon Dioxide. *ACS Appl. Mater. Interfaces.* **10**, 36961–36968, <https://doi.org/10.1021/acsami.8b12182> (2018).
103. Chen, L., Cheng, Y., Liu, Y. & Luo, D. Amine-impregnated carbon aerogels with ultra-high pore volume synthesized by pH adjusting for efficient CO₂ Capture. *J. Porous Mater.* **26**, 1313–1323, <https://doi.org/10.1007/s10934-019-00733-x> (2019).
104. Tan, M. *et al.* Fly ash-derived mesoporous silica with large pore volume for augmented CO₂ capture. *Fuel* **351**, 128874, <https://doi.org/10.1016/j.fuel.2023.128874> (2023).

105. Bai, G. *et al.* Polyethylenimine (PEI)-impregnated resin adsorbent with high efficiency and capacity for CO₂ capture from flue gas. *New J. Chem.* **43**, 18345–18354, <https://doi.org/10.1039/c9nj03822a> (2019).
106. Pi, W. *et al.* Carbon dioxide adsorption capacity of porous carbon after polyamine impregnation under different pore structures and temperatures. *J. Chem. Eng.* **485**, 150025, <https://doi.org/10.1016/j.cej.2024.150025> (2024).
107. Jo, D. H., Jung, H., Shin, D. K., Lee, C. H. & Kim, S. H. Effect of amine structure on CO₂ adsorption over tetraethylenepentamine impregnated poly methyl methacrylate supports. *Sep. Purif. Technol.* **125**, 187–193, <https://doi.org/10.1016/j.seppur.2014.01.048> (2014).
108. Peng, H.-L. *et al.* Chitosan-derived mesoporous carbon with ultrahigh pore volume for amine impregnation and highly efficient CO₂ capture. *J. Chem. Eng.* **359**, 1159–1165, <https://doi.org/10.1016/j.cej.2018.11.064> (2019).
109. Gibson, J. A. A., Gromov, A. V., Brandani, S. & Campbell, E. E. B. The effect of pore structure on the CO₂ adsorption efficiency of polyamine impregnated porous carbons. *Microporous Mesoporous Mater.* **208**, 129–139, <https://doi.org/10.1016/j.micromeso.2015.01.044> (2015).
110. Wang, D., Wang, X. & Song, C. Comparative Study of Molecular Basket Sorbents Consisting of Polyallylamine and Polyethylenimine Functionalized SBA-15 for CO₂ Capture from Flue Gas. *ChemPhysChem* **18**, 3163–3173, <https://doi.org/10.1002/cphc.201700828> (2017).
111. Grenda, K. *et al.* An analytical approach to elucidate the architecture of polyethyleneimines. *J. Appl. Polym. Sci.* **139**, 51657, <https://doi.org/10.1002/app.51657> (2022).
112. Wang, W. *et al.* Polydepsipeptide block-stabilized polyplexes for efficient transfection of primary human cells. *Biomacromolecules* **18**, 3819–3833, <https://doi.org/10.1021/acs.biomac.7b01034> (2017).
113. Kadlecova, Z. *et al.* Poly (ethyleneimine)-Mediated Large-Scale Transient Gene Expression: Influence of Molecular Weight, Polydispersity and N-Propionyl Groups. *Macromol. Biosci.* **12**, 628–636, <https://doi.org/10.1002/mabi.201100404> (2012).
114. Anderson, R., Biong, A. & Gómez-Gualdrón, D. A. Adsorption isotherm predictions for multiple molecules in MOFs using the same deep learning model. *Journal of chemical theory and computation* **16**, 1271–1283, <https://doi.org/10.1021/acs.jctc.9b00940> (2020).
115. Mashhadimoslem, H. *et al.* Development of the CO₂ Adsorption Model on Porous Adsorbent Materials Using Machine Learning Algorithms. *ACS Applied Energy Materials* **7**, 8596–8609, <https://doi.org/10.1021/acsaem.4c01465> (2024).
116. Bahmanpour, A. M., Tegeler, E. B., Colbert, T. & Masoudi, M. Data-Driven Analysis of Amine-Based Sorbents for CO₂ Removal from the Atmosphere. *Chem. Eng. Technol.* **47**, 221–232, <https://doi.org/10.1002/ceat.202300297> (2024).
117. Han, Y. *et al.* Facile improvement of amine dispersion in KIT-1 with the alkyl chains template for enhanced CO₂ adsorption capacity. *J. Solid State Chem.* **290**, 121531, <https://doi.org/10.1016/j.jssc.2020.121531> (2020).
118. Hung, C.-T. *et al.* Capture of carbon dioxide by polyamine-immobilized mesostructured silica: A solid-state NMR study. *Microporous Mesoporous Mater.* **238**, 2–13, <https://doi.org/10.1016/j.micromeso.2016.03.006> (2017).
119. Goel, C., Kaur, H., Bhunia, H. & Bajpai, P. K. Carbon dioxide adsorption on nitrogen enriched carbon adsorbents: Experimental, kinetics, isothermal and thermodynamic studies. *Journal of CO₂ Utilization* **16**, 50–63, <https://doi.org/10.1016/j.jcou.2016.06.002> (2016).

Acknowledgements

The authors would like to acknowledge the Science and Technology Commission of Shanghai Municipality (STCSM) for the financial support (no. 21DZ1206200). The authors would also like to express their appreciation for the funding support from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project Reference Number: T32-615/24-R). The authors are deeply grateful to the Consensus project, which received funding from the European Union's Horizon 2020 Research and Innovation program (no. 101022484).

Author contributions

Eryu Wang: Conceptualization, writing, reviewing and editing of the original manuscript, visualizations, collection of the original dataset, data processing, and training of machine learning models; Liping Luo: Reviewing of the manuscript, interactions with the GPT model, data processing, visualizations and statistical analysis. Jiachuan Wang: Reviewing of the manuscript and ideations of using GPT model; Jiaxin Dai: data collection and validations; Shuangyin Li: Supervision; Chen Lei: Supervision; Jia Li: Supervision and reviewing of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-025-05037-1>.

Correspondence and requests for materials should be addressed to S.L. or J.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025