JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JACS
Perspective

Artificial Intelligence Paradigms for Next-Generation Metal−Organic Framework Research

Aydin Ozcan,* François-Xavier Coudert, Sven M. J. Rogge, Greta Heydenrych, Dong Fan, Antonios P. Sarikas, Seda Keskin, Guillaume Maurin, George E. Froudakis,* Stefan Wuttke,* and Ilknur Erucar*

Cite This: J. Am. Chem. Soc. 2025, 147, 23367−23380

ABSTRACT: After the development of the famous "Transformer" network architecture and the meteoric rise of artificial intelligence (AI)-powered chatbots, large language models (LLMs) have become an indispensable part of our daily activities. In this rapidly evolving era, "all we need is attention" as Google's famous transformer paper's title [Vaswani et al., *Adv. Neural Inf. Process. Syst.* **2017**, 30] implies: We need to focus on and give "attention" to what we have at hand, then consider what we can do further. What can LLMs offer for immediate short-term adaptation? Currently, the most common applications in metal−organic framework (MOF) research include automating literature reviews and data extraction to accelerate the material discovery process. In this perspective, we discuss the latest developments in machine-learning and deep-learning research on MOF materials and reflect on how their utilization has evolved within the LLM domain from this standpoint. We finally explore future benefits to accelerate and automate materials development research.

INTRODUCTION

The combinatorial nature of metal−organic frameworks (MOFs) results in a vast chemical toolset and gigantic materials space, offering researchers a theoretically infinite number of candidate materials to choose from for applications spanning from gas storage and separation,[1,2] to drug delivery.[3−5] Given the diversity of this enormous chemical space, it is important to reflect on how we can explore this space efficiently in search of the "top" material for a given application. Data-driven techniques (represented in Figure 1) have emerged in recent years as the primary tool for streamlining the identification of the top MOFs. As shown in Figure 1, machine learning (ML) and deep learning (DL) studies are diverse. On the other hand, large language model (LLM) applications of MOFs are still limited[6−27] but have been increasing rapidly in the last two years.

Additional to the data-centric *"tour de force"* of ML methodologies, the applicability of ML tools in the field of MOF research also appears in the development of machine learning potentials (MLPs), which provide a novel approach to accurately capturing complex interactions with near quantum mechanical precision, while dramatically reducing computational costs for acquiring high-quality data sets, a key ingredient to further train reliable ML-predictive models.[28−39]

In this perspective, we begin our journey by exploring the use of ML methods for predicting structure−property relationships in MOFs. We then discuss the transformative role of DL and LLMs in MOF research, emphasizing their potential to revolutionize the design of novel MOFs with tailored properties on demand. The capability of MLPs to deliver highly accurate predictions, obtained from molecular simulations of MOFs under diverse conditions, is also highlighted. Ensuring easy access to data from diverse material databases, models, and user-friendly tools is crucial for facilitating the widespread adoption of data-driven methods in MOF research and broadening their impact beyond specialized experts to the wider material research community.

INSIGHTS FROM EARLY AI-DRIVEN STRUCTURE−PROPERTY PREDICTIONS FOR MOFS

We are now in an era where data science meets computer simulations. Figure 2 shows the interplay within the overall AI paradigm for the progress of MOF research. High-throughput computational screening (HTCS) approaches based on molecular simulations of MOFs have been important in evaluating large numbers of MOFs (126,800 experimental MOF structures are deposited in the Cambridge Structural Database (CSD),[40] and trillions of hypothetical MOF structures have been created). In addition, these methods provide molecular-level insights into materials' properties by complementing and directing the experimental studies.[41] However, HTCS has become too slow and expensive to explore this materials space effectively and efficiently.

Received: May 15, 2025
Revised: June 17, 2025
Accepted: June 17, 2025
Published: June 24, 2025

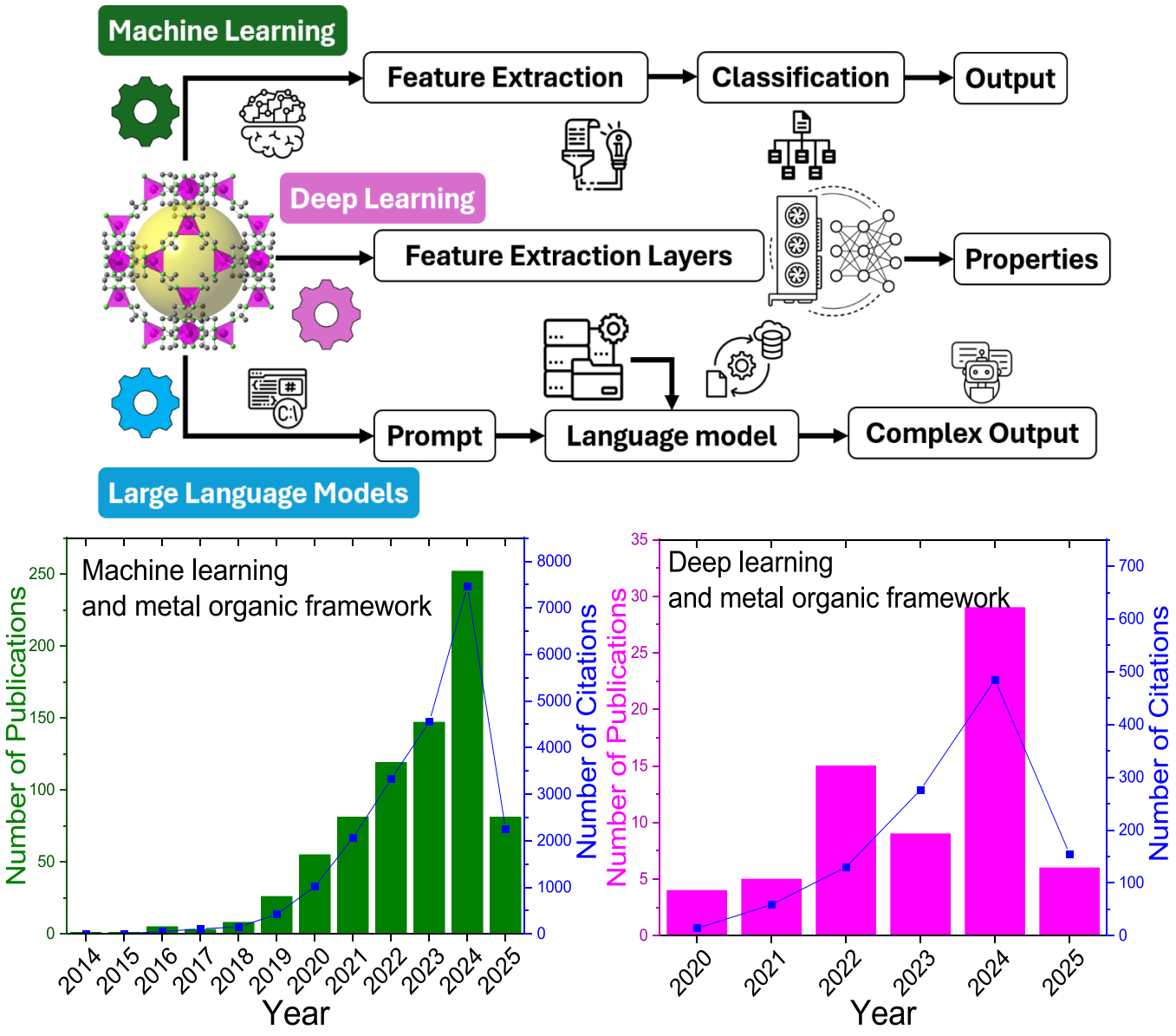

Figure 1. Schematic representation of data-driven methods and their usage in MOF research. Number of publications and their citations featuring the terms "machine learning" and metal organic framework or "deep learning" and metal organic framework in their topics. Accessed: 2025−04−03 from Web of Science.

Data-driven approaches reduce the need to run molecular simulations for every material, and integrating ML into molecular simulations and experiments has significantly accelerated the MOF discovery process in the last years. However, most ML and molecular simulation studies have primarily focused on gas adsorption under moderate to high-pressure conditions. A limitation remains the lack of accuracy in these approaches for more complex energy-related applications, particularly those involving gas capture at low traces, such as direct air capture (DAC) or adsorption of highly volatile compounds. By applying innovative derivative-free optimization methods such as Bayesian optimization[42] and multifidelity methods,[43] and sophisticated techniques such as new neural network (NN) architectures, ML can analyze vast MOF databases to identify key structural patterns associated with desirable properties, such as high gas adsorption, selectivity, or thermal stability. For example, Liu et al.[44] determined the ML hyperparameters via Bayesian optimization and used a crystal graph convolutional NN algorithm to virtually screen MOFs for toluene vapor adsorption. This narrows down the candidates for experimental synthesis, guiding researchers directly to promising compounds and expediting the discovery process. Several reviews[45−51] reflecting aspects of data acquisition, featurization, ML model training, and applications have already been published.

Deep learning (DL),[52] a subfield of ML based on NNs, has revolutionized the AI field with its impressive results in applications such as computer vision, natural language processing and speech recognition. One of the most important factors for the success of DL algorithms is the availability of large data sets like ImageNet,[53] since these algorithms are notorious for being "data hungry". With this in mind and taking into account the development of large MOF databases,[54−61] the appearance of DL techniques within MOF research should not be a surprise. DL algorithms enable researchers to directly process text-, graph- and image-based

https://doi.org/10.1021/jacs.5c08214
J. Am. Chem. Soc. 2025, 147, 23367−23380

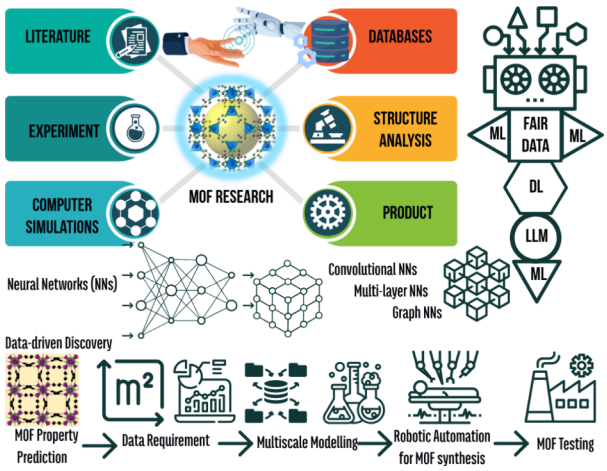

Figure 2. Interplay within the overall AI paradigm for the progress of MOF research.

representations of MOFs or even raw structural information.[62−64] A variety of DL workflows, based on multilayer perceptrons,[31,65−70] recurrent NNs (RNNs),[68,71−73] graph NNs (GNNs),[57,74−80] convolutional NNs (CNNs)[62,81,82] and transformer-based NNs,[83−86] have been developed and successfully applied for predicting various properties of MOFs: gas uptake,[62,83,84,86,87] gas diffusivity,[84] band gap,[83,84] bulk modulus,[65] stability metrics[70] and synthesizability.[71]

Besides uncovering structure−property relationships, predictive DL models can also be applied for accelerating expensive steps in computational workflows.[74,79,87−90] For example, modeling nonbonded interactions in Monte Carlo or molecular dynamics simulations of gas adsorption/diffusion requires accurate partial atomic charges for all MOF atoms. For example, Raza et al.[74] proposed a GNN that takes as input a crystal graph—i.e., a set of nodes and edges, representing atoms and bonds between atoms, respectively—and which generates node-level predictions, corresponding to partial charge predictions for MOF atoms while satisfying the charge neutrality constraint. The GNN was trained with DFT (density functional theory)-derived MOF partial point charges, achieved high-fidelity partial charge assignment, and importantly, with orders of magnitude shorter runtime compared to DFT calculations.

Other than predictive models which have been widely applied for high-throughput screening, generative models such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) are other classes of ML models that can expedite the discovery of high-performing MOFs. Instead of mapping from structure-to-property (as is the case of predictive models), these models adopt an inverse design[73,80,91] approach (i.e., property-to-structure), enabling the targeted design of tailor-made materials.

The long-term vision is to combine these data-driven predictions with real-time feedback loops and autonomous laboratory systems. In the experimental space, DL could integrate with robotic automation for MOF synthesis, enabling fully autonomous laboratories where DL models not only design new MOFs but also control robotic systems to synthesize and test them.[92,93] This would drastically accelerate the pace of discovery. Additionally, robotic and automated chemistry laboratories can produce vast amounts of data, underscoring the need for effective learning methods to process them. At this point, we note a key dilemma in MOF research: the difficulty of automating experimental processes. Many synthesis protocols lack reproducibility, often yielding inconsistent results even when the same procedure is followed. In addition, each research group employs specialized experimental procedures, making integration into a commercially available automated tool challenging. These compatibility issues are not only coming from materials synthesis but also from available software infrastructures. For example, in a *Nature Synthesis* Q&A discussion,[94] Prof. Andrew Cooper highlighted that key barrier to automating material synthesis is not the cost but rather specialized expertise required to implement generalized experimental systems. For example, his laboratory uses the Robot Operation System (ROS), yet compatibility issues persist, as not all robotic platforms are ROS-compatible. Additionally, challenges remain in standardizing software libraries and their interface programming, further complicating widespread adoption. The flexible automation concept[95] may solve these experimental challenges by dividing the workflows into individual tasks such as synthesis, activation, stability testing, and measurement. This task-oriented approach can introduce reconfigurable automated dynamic experiments.

From the materials synthesis perspective, reinforcement learning (RL) adds another layer of precision by dynamically optimizing synthesis parameters—including temperature, solvent choice, and reaction time—to achieve higher yields, enhanced crystallinity, and better phase purity. For example, Yaghi's team[26] developed an integrated AI system to determine the optimal conditions for the synthesis of MOFs and their organic related materials, covalent organic frameworks (COFs), for water harvesting in his laboratory. Microwave-assisted methods required 4 days (6,235 min) to optimize one compound from over 6 million variable combinations. This adaptive optimization could minimize trial and error, especially in challenging syntheses, and can unveil synthesis conditions that may otherwise remain unexplored. In another example, a web-based tool was developed to predict MOF synthesis conditions using ML models.[96] Users can upload the crystallographic files of MOFs and then receive the synthesis conditions of the corresponding MOFs, including synthesis temperature, time, solvent, and additives. These studies introduce a transformative approach in MOF synthesis, moving from experience-driven trial and error toward a systematic inverse design strategy.

The integration of AI tools significantly reduces operational costs by minimizing labor hours, reagent consumption, and equipment usage. To illustrate this, consider the optimization of UiO-66 synthesis. In 2020, Taddei et al.[97] performed 31 experiments to optimize the microwave synthesis conditions, achieving a significant increase in space-time yield (STY) from 23 kg/m^3·day to 2241 kg/m^3·day. The production cost of 1 kg of activated UiO-66, synthesized using dimethylformamide, zirconium chloride, terephthalic acid, and hydrochloric acid, has been reported as approximately 503.9 USD/kg.[98] For 31 reactions, each yielding 360 mg of UiO-66, the total product mass is 11.16 g, resulting in a material cost of 5.62 USD per batch. The power consumption for these 31 experiments was reported as 2438 W. Assuming a total experimental duration of 30 days and an electricity cost of 0.15 USD/kWh, the energy cost totals 263.3 USD. If the experimental work is conducted by 5 PhD students (each with an estimated salary of 2500 USD/month, totaling 2.5 full-time equivalents), the labor cost

https://doi.org/10.1021/jacs.5c08214
J. Am. Chem. Soc. 2025, 147, 23367−23380

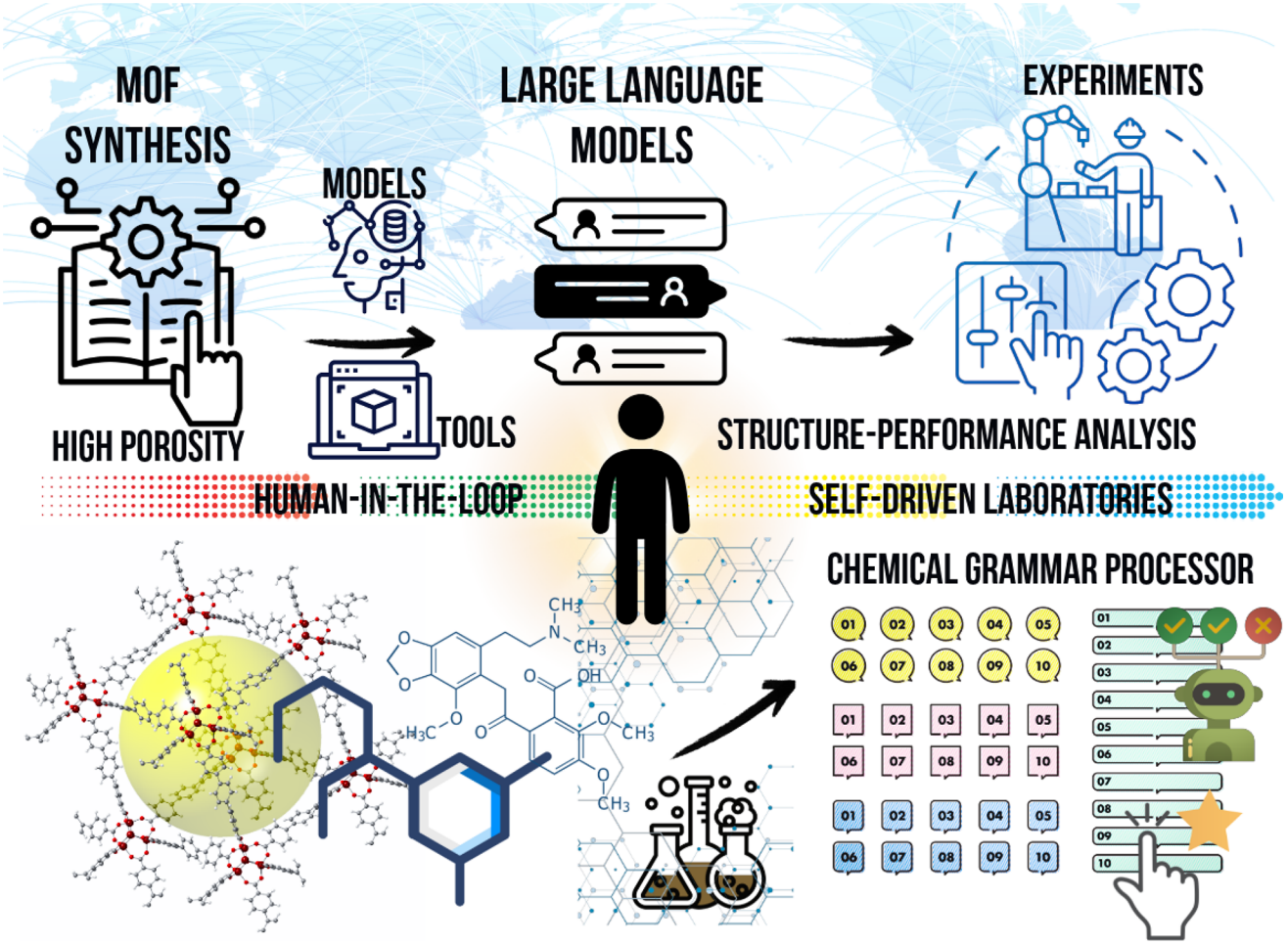

Figure 3. Possible AI-human loop for MOF chemistry.

totals 6250 USD/month. Combining these factors, the total estimated cost for this workflow is 6518.92 USD/month. To reduce costs, AI-driven optimization was explored. By employing AI tools (assumed cost: 2000 USD/month such as subscription fees, computational resources, maintenance, data storage, power consumption etc.) and requiring only 1 PhD student (0.5 full-time equivalent) to perform an optimized synthesis (producing 360 mg per experiment), the total cost is reduced to 3250.18 USD/month (including labor, materials, and AI implementation). This represents an almost 50% reduction in overall cost compared to the conventional approach. Furthermore, multiscale modeling—combining DL models that operate at different scales (from atomic to macroscopic)—could enable researchers to predict how MOFs would behave in real-world conditions, aiding in their deployment in industrial applications.[99] The potential impact is vast, with applications in CO_2 capture, hydrogen storage, water purification, and beyond—ultimately enabling a new era of responsive, high-performance materials tailored to tackle some of the world's most pressing challenges.[100]

Figure 2 highlights the growing concerns about AI potentially replacing human-centric jobs. While such concerns are valid, it is important to recognize that disruptive technologies have consistently brought both challenges and opportunities. By focusing on collaboration and innovation, scientists and society overall can harness the potential of AI for great achievements. Ultimately, as data-sharing platforms expand and interdisciplinary collaborations grow, the synergy between AI and MOF research could revolutionize materials science, enabling breakthroughs in clean energy storage, environmental remediation, and beyond. The future of AI in MOF research is not just about better predictions but about unlocking the ability to design materials with tailored properties on demand. Additionally, future research should not only focus on beating state-of-the-art results but should also provide practical and environmentally sustainable AI solutions. Thus, new researchers are encouraged to report the computational and carbon footprint[101,102] of their proposed approach in addition to standard performance metrics. We believe that AI is a pivotal tool in accelerating the journey toward novel MOFs and the recent breakthroughs are just the tip of the iceberg.

FROM EARLY AI PREDICTIONS TO THE ERA OF LARGE LANGUAGE MODELS

Large Language Models (LLMs) are potential game changers within the rapidly expanding research space of AI. The methodology of LLMs involves training NNs on vast amounts of text data, enabling them to understand, generate, and reason about human language. Figure 3 shows human-interpretable LLMs for MOF research.

LLMs can be used in MOF research in several innovative ways: (a) They can automate literature reviews by scanning large data sets and extracting key insights from research papers,

https://doi.org/10.1021/jacs.5c08214
J. Am. Chem. Soc. 2025, 147, 23367−23380

making it easier for researchers to stay updated on the latest developments. (b) LLMs can generate structured data from textual descriptions, aiding in the discovery of new MOF structures or optimizing synthesis methods. (c) Through integration with ML models, LLMs can predict experimental outcomes, propose new materials, and link disparate sources of information, leading to interdisciplinary breakthroughs. (d) Additionally, they can optimize experimental conditions by analyzing previous experiments and suggesting new approaches, accelerating the discovery of MOFs with desirable properties. Independent of these applications, LLMs hold potential for research automation, such as drafting reports, summarizing results, or hypothesizing new experiments based on available data. Fine-tuned on MOF-specific data and integrated with existing databases, LLMs can enhance information retrieval and facilitate collaboration across different fields, ultimately driving innovation and accelerating progress in MOF research.

To bring them to a wider audience and use them in service of more people, we need to identify the kind of tasks where LLMs are superior: LLMs are particularly good at summarization, sentiment analysis and text classification. For example, LLMs can analyze vast databases of scientific publications to identify trends and extract critical information on material properties, synthesis methods, and experimental results. This capability allows researchers to quickly gather comprehensive insights without manually skimming through thousands of papers. This would be the most obvious short-term adaptation of LLMs directly to MOF research. Implementation with a wider impact and a longer-term view would suggest the need to create tools and methodologies to convert materials synthesis and chemistry into a "language" and to teach the LLMs the "grammar of chemistry". The massive ML research effort of recent years has already created a great portion of the necessary tools, such as molecular representations, reaction descriptors, retrosynthetic analysis, condition optimization and a significant amount of data output.[23] So, what would be the steps of creating and converting an LLM for MOF topology generation and synthesis as a "reasoning engine"?

There are emerging answers in the literature converging on this fundamental question. One example is called ChatMOF, created by Kang and Kim.[7] ChatMOF can create MOFs with user-desired properties from human cognition and predict their properties. For example, ChatMOF can not only answer a text-compatible input, but also generate a MOF structure with user-defined properties. ChatMOF extracts the desired MOF data using a table-search operation from different MOF databases such as CoREMOF,[56] the CSD MOF subset[103] or QMOF[57] and also uses MOFkey[104] and DigiMOF[58] databases to provide topology-based and synthesis information. MOF-Transformer[84] is used as a toolkit to predict the properties of MOFs based on an ML model. Here, it is important to clarify that ChatMOF leverages language to utilize knowledge already curated in its data set, which differs from actual reasoning on a chemical task or question—a capability that remains a challenge for AI.

To dive into the discussion for adapting a LLM for MOF topology generation, the first step is to create a way of generating chemical word embedding and tokenization. Chemical word embedding is a database structure which stores chemical words (ligands and metals in terms of MOFs) according to some proximity rules and locates similar words nearby and dissimilar words far apart. What makes two ligands similar in terms of reactivity, MOF synthesis or output topology, i.e., in their "meaning"? That seems like an open question waiting for a rigorous answer.

This issue was sorted out in Natural Language Processing (NLP) research by word2vec,[105] an analogous methodology and development for MOFs (maybe named: MOF2vec) seem desirable. Additionally, a relevant methodology for tokenization is needed for a MOF reasoning engine. In NLP, tokenizing a given text word-wise overloads the vocabulary dictionary (a sequence assigned to each word). On the other hand, tokenizing a given text character-wise results in a very lightweight vocabulary dictionary (a number assigned to each letter and the list has only 26 items in English). But this time, the context within words is lost. Therefore, in LLM applications, tokenization is a well-tuned process. The question we need to answer is how to fragment a given MOF structure to generate a "sub-word" dictionary that would enable us to speak the language of MOF chemistry.

Word embeddings of chemical elements are used to represent the stoichiometric formula of MOFs, and the chosen embeddings are derived from unsupervised learning on raw text (i.e., natural language texts) to capture implicit knowledge from the corpus (a large and structured collection of texts in a natural language). To construct features based on the composition of each MOF structure (almost 200 embedding dimensions), the ElementProperty featurizer in matminer[106] is utilized. This model facilitates material design by establishing a connection between property predictions and crystal structure to further develop high-performance materials for gas adsorption applications. Geometric entities can be used to predict some chemical properties of MOFs, but given the large variety of MOFs, how can one create a chemistry-based design using a MOF language? We already have some tools and the vision to harness LLMs as a "chemical grammar processor". Although the road to this goal might not be direct, each new tool brings us closer to the rational design of new materials.

The future perspective of LLMs in MOF research holds immense potential to revolutionize how we design, discover, understand, and, more importantly, think about these materials. As LLMs become more sophisticated and fine-tuned for specialized scientific domains, they will likely evolve into essential tools for automating much of the research process. In the coming years, we could see LLMs used for real-time hypothesis generation, where researchers can interact with the model to brainstorm new ideas, identify unexplored research avenues, or propose novel MOF structures with specific properties, such as enhanced gas adsorption or catalytic efficiency. This will significantly shorten the discovery-to-deployment cycle for new materials.

Moreover, as LLMs become better at integrating vast and diverse data sets, including experimental data, computational models, and scientific literature, they could act as highly intelligent systems capable of predicting material behaviors under various conditions. These models may go beyond summarizing existing research to synthesize new knowledge by connecting dots across different scientific domains. This interdisciplinary approach will likely unlock new applications for MOFs in areas like renewable energy, environmental remediation, and drug delivery, as LLMs provide insights that human researchers might overlook.

Looking further ahead, LLMs combined with other AI technologies could enable fully autonomous research pipelines. In these systems, LLMs would not only generate hypotheses

https://doi.org/10.1021/jacs.5c08214
J. Am. Chem. Soc. 2025, 147, 23367−23380

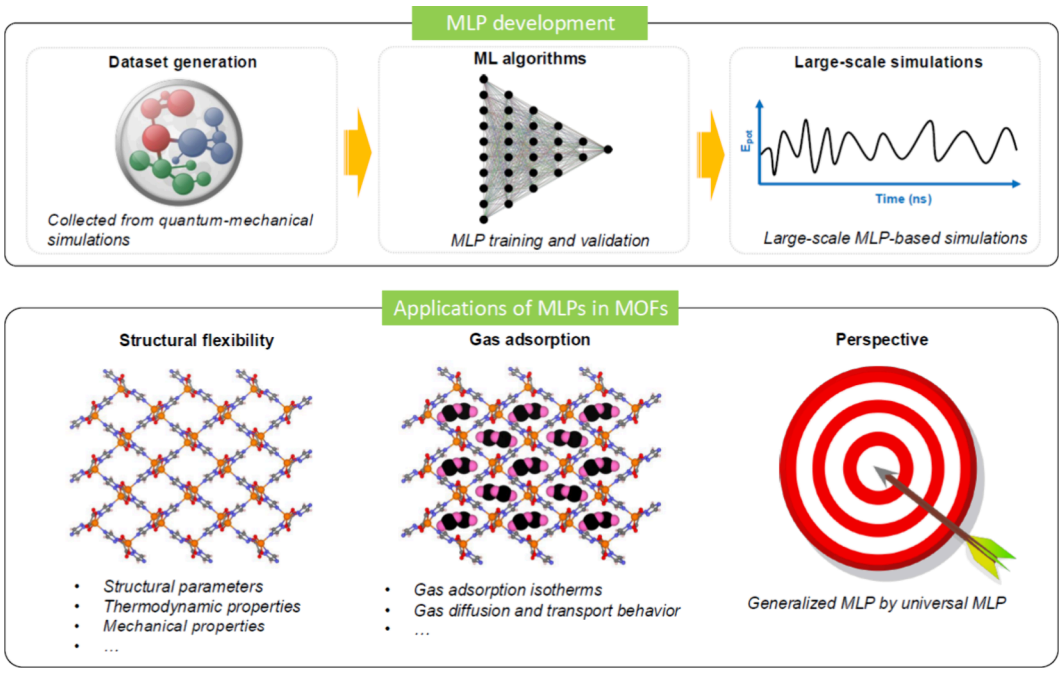

Figure 4. Schematic diagram of the development of MLPs based on quantum mechanical data sets alongside typical applications of MLPs to the MOF field.

but also design and execute simulations, analyze results, and even control robotic systems to perform physical experiments. This could lead to the discovery of completely new classes of materials with tailored properties, optimized for specific industrial or environmental applications. Since industrially relevant conditions are critical for assessing the environmental impact of MOF products,[107] AI can help explore the infinite number of scenarios such as varying reaction conditions, solvent recovery or waste management processes. Ultimately, LLMs could transform MOF research by making it faster, more efficient, and more innovative, pushing the boundaries of what is currently achievable in materials science.

In the past years, the many potential applications and varieties of MOFs have truly come to the fore. One reason for this blossoming of the field is the development of LLMs and generative AI tools that can support scientists in searching the reaction space for this class of materials. What these efforts make clear is that digital standardization of terminology, structural representation, and nomenclature is essential for harnessing the full capabilities of digital tools. Having a standardized nomenclature in the MOF community is important for ensuring interoperability between data management and software systems. These backward-compatible standards are only the first step in creating a common language. As the chemistry community, we must also ensure that there is universal dissemination and uptake. This will enable us all to speak the language of MOFs and the language of chemistry!

Learning algorithms have been proposing solutions not only for data collection, interpretation, and structuring, but also for a long-standing problem of the molecular modeling community called "scale hierarchy". Scale hierarchy problems refer to the difficulty of integrating results from the electronic-level simulation to the atomistic level, from the atomistic level to the mesoscale, and from the mesoscale to the continuum level. To overcome this difficulty and plug the gap between the "electronic level to the atomic level", physical models were commonly used and had limitations. Now, ML-based potentials open up completely new possibilities to tackle this problem.

ML POTENTIALS: A NEW FRONTIER IN ATOMIC-SCALE SIMULATIONS OF MOFS

Interatomic potentials are critical in understanding atomic interactions and predicting the properties of materials at the atomic scale. Traditional quantum mechanical-based methods such as *ab initio* methodologies have delivered essential insights, however they are limited in terms of time and length scales. Alternatively, empirical or semiempirical models, such as Lennard-Jones or embedded atom potentials, have been extensively used in the past two decades to describe the intra- and intermolecular interactions in MOFs. However, these classical potentials often struggle to balance accuracy, computational burden, and generalizability across diverse chemical MOF systems. Machine learning potentials (MLPs) represent a new avenue for capturing the most complex interactions with near quantum mechanical-accuracy but at a fraction of computational cost. This has enabled studying MOF materials with large atomic-scale simulations that were unfeasible at the quantum mechanical level, from exploring phase transitions to predicting their mechanical, thermal and adsorption properties among others. Unlike classical potentials, MLPs are trained on high-quality quantum mechanical data, often derived from DFT or other advanced *ab initio* methods, assembling a large data set of atomic configurations and associated energies/forces of the explored MOF systems.[108,109] As illustrated in Figure 4, by employing ML algorithms ranging from NNs to Gaussian regression or kernel methods to map atomic positions to the potential energy surface, these MLP models can capture the intricate atomic interactions, including bond breaking and formation, with unprecedented precision. By continuously learning from larger data sets and adapting to

https://doi.org/10.1021/jacs.5c08214
J. Am. Chem. Soc. 2025, 147, 23367−23380

various atomic environments, MLPs show a high adaptability because they can be improved by incorporating new data. This makes them highly versatile for simulating MOFs with different compositions, structures, and environmental conditions. MLPs can be easily applied to more complex systems, something that has been a limitation of traditional models. Thus, disordered or amorphous phases can be modeled, providing unprecedented insight into the flexible and dynamic nature of materials.[110,111]

In recent years, the use of MLPs in the field of MOFs has witnessed a profound and rapid expansion, improving the exploration of the physical properties of this class of materials by unlocking deeper insights. Decisively, a precise description of the structural flexibility of MOFs via MLPs offers a unique opportunity to explore how these hybrid frameworks respond to varying temperature and pressure. Some typical illustrations include the exploration of the zeolitic imidazolate frameworks (ZIFs),[112] MOF-5,[113] CALF-20,[114] and 2D MOFs.[115] For example, MLP-based molecular dynamics simulations revealed unique thermodynamic and mechanical properties of CALF-20 at finite temperature, e.g., negative area compressibility, negative thermal expansion and unusual strain-softening behaviors.[7] Another advantage of MLPs is the ability to simulate large systems at experimentally relevant scales. Unlike DFT calculations, which are limited to small systems and short time scales, a recent study demonstrated that a high-quality MLP trained for a 2D MOF can be used to simulate experimental-size MOF membranes (up to 28.2 × 28.2 nm^2) without losing computational accuracy.[115] Thus, the behavior of MOFs can be studied under conditions closer to practical applications.

MLPs have also been applied recently to predict the gas adsorption properties of MOFs, e.g. Al-*soc*-MOF-**1d** and ZIF-8.[87,116] with high precision. Here, one of the most critical challenges is to accurately describe the host/guest interactions. In classical simulations, the van der Waals interactions, typically modeled by Lennard-Jones and Buckingham potentials with parameters taken from generic force fields, e.g. UFF[117] and DREIDING,[118] are augmented by an electrostatic term to account for the interactions between charged MOF atoms. However, there are many examples where force fields must be reparameterized or derived from quantum-mechanical calculations due to a poor agreement with experiments, especially when the MOF framework contains open metal sites (OMSs).[119,120] Decisively, MLPs can accurately model not only the most complex MOF/guest interactions but also the guest-induced dynamics of the MOF framework which is most often overlooked in classical simulations with the use of rigid-lattice models. Typically, a MLP trained on a relatively large data set of MOF/guest configurations generated by *ab initio* molecular dynamics was demonstrated to accurately capture the H_2/Al-*soc*-MOF-**1d** interactions as well as the H_2-triggered MOF scaffold dynamics leading to a predicted adsorption isotherm at 77 K via grand Canonical Monte Carlo simulations in excellent agreement with the experimental data.[87] This strategy could be generalized to provide a more accurate and efficient assessment of the adsorption behavior of the most complex MOFs.

Typically, an MLP is mostly trained on a single MOF phase and therefore struggles with transferability across different chemical and structural environments. This would lead to inaccurate predictions for other MOFs. The breakthrough in this field would therefore be the development of universal MLPs to model the vast structural diversity of MOFs with quantum mechanical-level accuracy. The integration of such highly accurate MLPs with HTCS tools would enable researchers to anticipate a myriad of MOF properties in real-time and with unprecedented precision. In this context, the ongoing development of universal MLPs, e.g., CHGNet,[121] M3GNet,[122] MACE,[123] holds immense promise. Another key evolution lies in the development of self-improving MLPs, where models continuously learn and adapt from new data, expanding their predictive power across an ever-wider range of MOFs. This could drastically accelerate the discovery of novel MOFs for next-generation applications by exploring vast MOF chemical/structure spaces.

To be able to use all these fascinating methods, we initially need high-quality data. This is a tedious problem since the collection and precollection are still not very well-defined to reach "high-quality" data. That is why we would like to insist again that the major problem here boils down to, inspired by Sherlock Holmes, the famous fictional detective created by Sir Arthur Conan Doyle:[124]*Data, data, data! I cannot make bricks without clay.* In this context, we would like to identify some key entry barriers for data-centric design as well as suggest some systematic precollection and collection methods to reach high-quality data for MOF research.

LOWERING THE BARRIER TO ENTRY FOR DATA-BASED MATERIALS DESIGN

As highlighted above, we are witnessing a rapid explosion of the number of proposed scientific methods based on data for the discovery of novel materials, the identification of known materials with specific desirable properties, the optimization of chemical engineering processes and, more broadly speaking, multiobjective optimization and decision making in the field of applied materials sciences. Accompanying this fast pace of theoretical development, there is an important demand from the wider research community — and not only experts in data science — to be able to use these methods that they hear so much about (both in scientific publications and in the general press), with legitimate questions such as *"If AI and/or ML is going to change the way of chemistry, how can I leverage it in my own research projects?"*. We see in the field a growing recognition of the need to democratize access to data — and beyond data, to models.

The drive to lower the barrier to entry for the use of data-based methods is 3-fold, in our view: (i) it concerns the access to data, (ii) the access to models, (iii) and their ease of use to the wider community. All three aspects are necessary to drive the adoption of data-based methods in the materials science research community, beyond data scientists and specialists in theoretical and numerical methods.

The first aspect that we highlight is the sharing of data, to maximize the reuse of research data. This need is driven not only by research ethos but also by an increased recognition that there is a clear economic cost associated with dark data or unshared data − which was costed at €10.2bn per year at the scale of the EU economy.[125] We note again that probably the best-known set of guiding principles for sharing scientific data in the modern age are the FAIR principles.[126] The FAIR data standards are Findability, Accessibility, Interoperability, and Reusability. While many research groups nowadays make efforts to make their data findable and accessible, we emphasize that interoperability and reusability are sometimes more difficult to achieve, especially in a field where the nature of the data (and the materials that are described) are incredibly

https://doi.org/10.1021/jacs.5c08214
J. Am. Chem. Soc. 2025, 147, 23367−23380

diverse. However, interoperability and reusability are of absolutely fundamental importance. Progress in this area will require work to improve the metadata associated with the data itself, their representation and standardization, as well as the use of shared, accessible vocabularies and ontologies. This touches on the core issue of "What defines a material?", a question that different research communities (and families or classes of materials) would have different answers to.

The second aspect necessary to democratize data-based methods is that of access to models, i.e., to the trained ML models, to the code that was used to train them, and to the code that is necessary to deploy them. This is crucially important for the reproducibility of scientific research in our field, and a cornerstone of the scientific method. Without full sharing of models, it is impossible (for groups other than the original authors) to benchmark published methods on new data, or to compare the merits of different algorithms — and therefore an obstacle on the road to progress. This aspect is also of vital importance to bolstering accountability in AI research, something that is necessary to build the trust of the broader community. Furthermore, it is also linked, to some level, to a necessary homogenization of the reporting standards for new data-based studies in our field. We hope that in the future the community in materials sciences and chemistry will propose and enact coherent reporting requirements. We note that such efforts have been proposed before, either in the form of best practices by experts in the field,[127] or through funder mandates (similar to open access mandates).[128]

The third prong of this push toward democratization of data-based methods may appear, on first view, as less "scientific" or technological than the previous two. However, we argue that there is an important demand to make published data and models easier to use for the wider community, through initiatives like centralized databases with online portals, user-friendly data visualization tools, training programs, and data analytics programs. For example, while many MOF databases have been proposed, they are often hosted on different platforms (some on Zenodo, some on GitHub, some on specific websites, etc.). One example of the push toward unification (and therefore, greater interoperability) is that of the Materials Project web portal.[129] The Materials Project portal is an open web-based resource of computed properties of materials. It is centralized but it also allows for the upload of user-created data in the form of external "contributions", allowing better sharing of data through a common platform with a well-defined user interface. Future efforts should be encouraged for such platforms, extending to a wider range of chemical space (beyond crystalline materials, for example) and suitably integrating both experimental and computational data of both physical and chemical nature, further enabling interdisciplinary collaborations. Similarly, published models can be integrated into such online platforms, making it easy for nonexperts to perform simple property prediction (or other ML tasks) simply by uploading one or several structure files. This would allow large-scale screening of both experimental and hypothetical structures to identify promising candidates by pushing the Pareto front for specific applications, in a multiobjective optimization strategy — possibly, in the long term, including questions that are deemed too difficult at the moment, such as generative models for materials design and realistic estimation of synthesizability (or feasibility) of hypothetical structures.

This whole task is an aspirational but necessary attempt to democratize the landscape of data and AI methods for the new generation materials development endeavor. How can we adapt all these suggestions and guiding principles into a vibrant ecosystem of MOFs, COFs, and similar materials?

CREATING AN INTEGRATED MATERIAL DATABASE

Responding to this challenge, we suggest to integrate existing material databases into an overarching, curated platform and diversifying its coverage through three distinct steps. This suggestion is based on recent evaluations of existing databases—both experimental and computational in nature—that uncovered important shortcomings, as discussed by Gibaldi et al.[130] and De Vos et al.,[131] as well as on personal experiences of developing and using these databases.

Step 1. Data Collection and Integration. In the past decade, a wide variety of MOF and COF databases based on experimental structures,[56−58,56−58,132−134] hypothetical structures,[54,55,131,135−137] or a combination thereof[59] have been developed, often as a starting point for high-throughput screening studies. These MOF and COF structures, once properly represented using, e.g., the MOFid format[104] or the Weisfeiler-Lehman kernel in the graph2vec algorithm,[138] form a promising starting point to establish the integrated database platform envisioned here. In this platform, the tokenized reticular structures, properties reported in the original database, and, additionally, properties such as pore size distribution,[139] persistence diagrams,[140] and revised autocorrelation (RAC) descriptors[141] that are calculated separately from the original database, would form individual 'documents', as in the MOF recommendation system established by Zhang et al.[142] Such standardized document-structured genomes[140] ensure that the database can afterward be leveraged to recommend candidate materials for specific applications based on the similarity of their material embedding vectors with embedding vectors of known well-performing materials generated through a Doc2Vec model. In addition, once the materials are tokenized, this database can be hugely expanded and enriched through text-mining the existing MOF and COF literature, similar to the DigiMOF database.[58] This algorithm could cover new literature continuously, ensuring the platform remains up-to-date.[58] For both the original data points and those obtained by text mining, sufficient metadata must be present, including the source of the data, whether it was experimentally verified, computationally calculated, or predicted by a model, and the context in which the property was reported.[141] This would help adhere to the FAIR principles and identify and correct the "fuzzy" context omnipresent in materials science.[142]

Step 2. Data Correction and Curation. At this point, duplicate structures may be present in our database, which would bias subsequent model training.[143] The database would likely contain errors, either because they were present in the literature, or because they were introduced during text mining. Recently, MOSAEC-DB (Metal Oxidation State Automated Error Checker Database),[130] a database containing over 124,000 MOF structures, was shared within the MOF community to address structural problems arising from computational processing, such as those caused by solvent removal or unreasonable assigned oxidation states. These contributions are highly appreciated, as they establish a unified platform to support MOF research. While it is straightforward

https://doi.org/10.1021/jacs.5c08214
J. Am. Chem. Soc. 2025, 147, 23367−23380

to remove (near-)duplicates based on their overlapping embedding vectors, correcting and enriching data requires more attention. A possible path forward here is based on the observation that the collective knowledge in large databases can help correct mistakes in individual entries, as demonstrated by Jablonka et al.[144] To do so, various competing ML models would be trained and tested on our database's existing {material, property} pairs and then used to predict these properties for the whole database. In most cases, this is new information that enriches the database, while a limited amount of material predictions can be compared with actual experimental or simulation outcomes. This step is vital to test the accuracy of the ML models and to help identify possible errors in the database, as in ref.[144], especially when the predictions of multiple ML models agree with one another. In cases where these models disagree, domain experts would still need to identify the most likely result.[25]

Step 3. Data Diversification. Most hypothetical databases are biased due to the limited amount of building blocks and topologies to generate hypothetical structures, while both literature studies and experimental databases tend to be skewed toward easy-to-synthesize materials.[131,141] To ensure our platform is sufficiently diverse and can be adopted to identify promising materials beyond the limited chemical space explored until now,[143,145] it is essential to assess its variety, balance, and disparity in terms of the properties calculated in Step 1.[141] By identifying unexplored and weakly explored regions in chemical space, hypothetical MOF and COF structures could be generated through, e.g., LLM-based genetic algorithms targeting these regions, as demonstrated recently.[7,23,146,147] Besides the excellent integration between LLMs and genetic algorithms for this inverse design task,[7,23,146,147] this also allows for the efficient exploration of the structure space of crystalline materials by varying the temperature in the final softmax layer.[23,148] Such hypothetical structures can be straightforwardly generated using the reticular principle by extracting and recombining building blocks, as demonstrated before,[145,147] paying specific attention to defective and disordered structures, given their profound impact on the resulting material properties.

CONCLUSIONS

The 2024 Nobel Prizes in Physics and Chemistry both celebrated groundbreaking advancements in AI, physics, chemistry, and computational methods that deepen our understanding of complex structures. If we turn to MOF chemistry, the integration of ML, DL, and LLM has shown remarkable potential in advancing material research. By accelerating the discovery process, predicting material properties, and enabling efficient data analysis, these tools address complex challenges in MOF design and optimization. The use of ML and DL algorithms can significantly reduce experimental time and resource costs, while LLMs support researchers by quickly summarizing relevant literature, hypothesizing new MOF structures, and even generating new pathways for synthesis. Together, these approaches are making MOF research more vibrant, insightful and data-driven.

Looking ahead, the application of ML, DL, and LLMs in MOF research is expected to expand significantly. Figure 5 summarizes the extraordinary impetus of AI for new materials development in the field of porous materials, specifically MOFs. Future efforts will likely focus on creating more specialized models tailored for complex MOF systems,

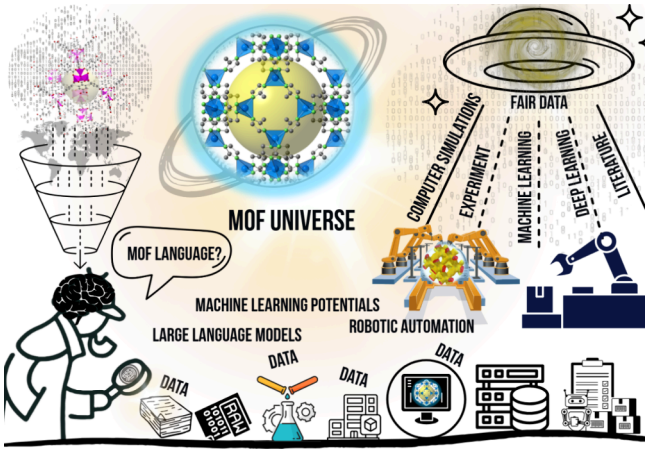

Figure 5. Future progress of MOF research.

integrating multimodal data for enhanced predictive capabilities, and improving model interpretability to deepen our understanding of MOF behavior at the atomic and molecular levels. As these tools become increasingly accessible and sophisticated, they may eventually support real-time, AI-driven experimental design, fostering an era where MOF discovery and application reach unprecedented heights. This synergy between AI and materials science promises to catalyze transformative advancements across fields such as energy storage, catalysis, and environmental remediation by using MOFs.

There is still much room for exploring the vast chemical spaces of materials and their flexible/stable configurations. Predictive modeling remains in its infancy for applications like healthcare and energy storage. Integrating AI with automated MOF synthesis for scale-up technologies, especially using green chemistry principles can enable tailored solutions, such as high-performance batteries or personalized cancer treatments. These are only a few examples, and the possibilities are endless.

Overall, with the increasing importance of data-driven decision making and the proliferation of (AI-driven) chatbots, there is a growing recognition of the need to democratize access to data. The MOF community should find ways to remove the barrier to data access and use by providing user-friendly databases, data visualization tools, training programs, and data analytics programs. Cross-functional collaboration should be encouraged by supporting data-driven initiatives. These provide vital goalposts to be reached in the area of material design with AI.

We believe that with the potential of human creativity and the power of experimental design and modeling tools, the sky is the limit!

AUTHOR INFORMATION

Corresponding Authors

Aydin Ozcan − *TUBİTAK Marmara Research Center, Materials Technologies, Gebze, Kocaeli 41470, Türkiye; Gebze Technical University, Kocaeli, Gebze 41400, Türkiye*; Email: aydin.ozcan@tubitak.gov.tr

George E. Froudakis − *Department of Chemistry, University of Crete, Voutes Campus, Heraklion 70013, Greece*; Email: frudakis@uoc.gr

Stefan Wuttke − *Academic Centre for Materials and Nanotechnology, AGH University of Krakow, Krakow 30-*

https://doi.org/10.1021/jacs.5c08214
J. Am. Chem. Soc. 2025, 147, 23367−23380

059, Poland; Department of Chemistry, United Arab Emirates University, Al-Ain 15551, United Arab Emirates; orcid.org/0000-0002-6344-5782; Email: swuttke@agh.edu.pl

Ilknur Erucar − *Faculty of Engineering, Ozyegin University, Istanbul 34794, Türkiye;* orcid.org/0000-0002-6059-6067; Email: ilknur.erucar@ozyegin.edu.tr

Authors

François-Xavier Coudert − *Chimie ParisTech, PSL University, CNRS, Institut de Recherche de Chimie Paris, Paris 75005, France;* orcid.org/0000-0001-5318-3910

Sven M. J. Rogge − *Center for Molecular Modeling (CMM), Ghent University, Ghent 9052, Belgium;* orcid.org/0000-0003-4493-5708

Greta Heydenrych − *Research Commons Building 4501, Research Triangle Park, North Carolina 27709, United States;* Present Address: Springer Nature

Dong Fan − *ICGM, Univ. Montpellier, CNRS, ENSCM, Montpellier F-34293, France*

Antonios P. Sarikas − *Department of Chemistry, University of Crete, Voutes Campus, Heraklion 70013, Greece*

Seda Keskin − *Department of Chemical and Biological Engineering, Koç University, Istanbul 34450, Türkiye;* orcid.org/0000-0001-5968-0336

Guillaume Maurin − *ICGM, Univ. Montpellier, CNRS, ENSCM, Montpellier F-34293, France;* orcid.org/0000-0002-2096-0450

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.5c08214

Notes

Figures were created by Adobe Express. Accessed April 15, 2025. https://www.adobe.com/express/.

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank the European Union (European Cooperation in Science and Technology) for the COST Action EU4MOFs (CA22147). F.X.C. acknowledges funding under the France 2030 framework by Agence Nationale de la Recherche (project ANR-22-PEXD-0009 "MOFs Learning" as part of PEPR DIADEME). S.M.J.R. acknowledges funding by the Research Board of Ghent University (BOF, grant no. BOF/STA/202309/008) and the European Union (ERC-StG grant no. 101115787 − STRAINSWITCH). S.K. acknowledges funding by the European Union (ERC, STARLET, 101124002). Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. G.M. acknowledges the Institut Universitaire de France for the Senior Chair.

■ REFERENCES

(1) Li, B.; Wen, H.-M.; Zhou, W.; Chen, B. Porous metal−organic frameworks for gas storage and separation: what, how, and why? *J. Phys. Chem. Lett.* **2014**, *5* (20), 3468−3479.

(2) Freund, R.; Zaremba, O.; Arnauts, G.; Ameloot, R.; Skorupskii, G.; Dincă, M.; Bavykina, A.; Gascon, J.; Ejsmont, A.; Goscianska, J.; Kalmutzki, M.; Lächelt, U.; Ploetz, E.; Diercks, C. S.; Wuttke, S. The current status of MOF and COF applications. *Angew. Chem., Int. Ed.* **2021**, *60* (45), 23975−24001.

(3) Wang, A.; Walden, M.; Ettlinger, R.; Kiessling, F.; Gassensmith, J. J.; Lammers, T.; Wuttke, S.; Peña, Q. Biomedical metal−organic framework materials: perspectives and challenges. *Adv. Funct. Mater.* **2023**, *34*, No. 2308589.

(4) Andreo, J.; Ettlinger, R.; Zaremba, O.; Peña, Q.; Lächelt, U.; de Luis, R. F.; Freund, R.; Canossa, S.; Ploetz, E.; Zhu, W.; Diercks, C. S.; Gröger, H.; Wuttke, S. Reticular nanoscience: bottom-up assembly nanotechnology. *J. Am. Chem. Soc.* **2022**, *144* (17), 7531−7550.

(5) Barsoum, M. L.; Fahy, K. M.; Morris, W.; Dravid, V. P.; Hernandez, B.; Farha, O. K. The road ahead for metal−organic frameworks: current landscape, challenges and future prospects. *ACS Nano* **2025**, *19* (1), 13−20.

(6) Bai, X.; He, S.; Li, Y.; Xie, Y.; Zhang, X.; Du, W.; Li, J. R. Construction of a knowledge graph for framework material enabled by large language models and its application. *npj Comput. Mater.* **2025**, *11* (1), 51.

(7) Kang, Y.; Kim, J. ChatMOF: an artificial intelligence system for predicting and generating metal-organic frameworks using large language models. *Nat. Commun.* **2024**, *15*, 4705.

(8) Schilling-Wilhelmi, M.; Ríos-García, M.; Shabih, S.; Gil, M. V.; Miret, S.; Koch, C. T.; Márquez, J. A.; Jablonka, K. M. From text to insight: large language models for chemical data extraction. *Chem. Soc. Rev.* **2025**, *54*, 1125−1150.

(9) Zheng, Z.; Rampal, N.; Inizan, T. J.; Borgs, C.; Chayes, J. T.; Yaghi, O. M. Large language models for reticular chemistry. *Nat. Rev. Mater.* **2025**, 1−13.

(10) Bai, X.; Xie, Y.; Zhang, X.; Han, H.; Li, J.-R. Evaluation of open-source large language models for metal−organic frameworks research. *J. Chem. Inf. Model.* **2024**, *64* (13), 4958−4965.

(11) Wellawatte, G. P.; Schwaller, P. Human interpretable structure-property relationships in chemistry using explainable machine learning and large language models. *Commun. Chem.* **2025**, *8* (1), 11.

(12) Kang, Y.; Lee, W.; Bae, T.; Han, S.; Jang, H.; Kim, J. Harnessing large language models to collect and analyze metal−prganic framework property data set. *J. Am. Chem. Soc.* **2025**, *147* (5), 3943−3958.

(13) Dagdelen, J.; Dunn, A.; Lee, S.; Walker, N.; Rosen, A. S.; Ceder, G.; Persson, K. A.; Jain, A. Structured information extraction from scientific text with large language models. *Nat. Commun.* **2024**, *15* (1), 1418.

(14) An, Y.; Greenberg, J.; Uribe-Romo, F. J.; Gómez-Gualdrón, D. A.; Langlois, K.; Furst, J.; Kalinowski, A.; Zhao, X.; Hu, X.Knowledge Graph Question Answering for Materials Science (KGQA4MAT). In *Metadata and Semantic Research. MTSR 2023*; Communications in Computer and Information Science; Garoufallou, E., Sartori, F. (eds), Springer: Cham, **2024**, vol. *2048*.

(15) Zhao, Y.; Zhao, Y.; Wang, J.; Wang, Z. Artificial Intelligence Meets Laboratory Automation in Discovery and Synthesis of Metal-Organic Frameworks: A Review. *Ind. Eng. Chem. Res.* **2025**, *64* (9), 4637−4668.

(16) Vu, V.-H.; Bui, K.-H.; Dang, K. D. D.; Duong-Tuan, M.; Le, D. D.; Nguyen-Dang, T. Finding environmental-friendly chemical synthesis with AI and high-throughput robotics. *J. Sci.:Adv. Mater. Devices* **2025**, *10* (1), No. 100818.

(17) Zhang, W.; Wang, Q.; Kong, X.; Xiong, J.; Ni, S.; Cao, D.; Niu, B.; Chen, M.; Li, Y.; Zhang, R.; et al. Fine-tuning large language models for chemical text mining. *Chem. Sci.* **2024**, *15* (27), 10600−10611.

(18) Zheng, Z.; Zhang, O.; Borgs, C.; Chayes, J. T.; Yaghi, O. M. ChatGPT chemistry assistant for text mining and the prediction of MOF synthesis. *J. Am. Chem. Soc.* **2023**, *145* (32), 18048−18062.

(19) Kalhor, P.; Jung, N.; Bräse, S.; Wöll, C.; Tsotsalas, M.; Friederich, P. Functional material systems enabled by automated data extraction and machine learning. *Adv. Funct. Mater.* **2024**, *34* (20), No. 2302630.

(20) Zheng, Z.; Rong, Z.; Rampal, N.; Borgs, C.; Chayes, J. T.; Yaghi, O. M. A GPT-4 Reticular Chemist for Guiding MOF Discovery. *Angew. Chem., Int. Ed.* **2023**, *62* (46), No. e202311983.

(21) Ansari, M.; Moosavi, S. M. Agent-based learning of materials datasets from the scientific literature. *Digit. Discovery* **2024**, *3* (12), 2607−2617.
(22) Yin, J.; Bose, A.; Cong, G.; Lyngaas, I.; Anthony, Q.Comparative study of large language model architectures on frontier. In *2024 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*; IEEE: **2024**, (pp 556−569).
(23) Jablonka, K. M.; Schwaller, P.; Ortega-Guerrero, A.; Smit, B. Leveraging large language models for predictive chemistry. *Nat. Mach. Intell.* **2024**, *6* (2), 161−169.
(24) Van Herck, J.; Gil, M. V.; Jablonka, K. M.; Abrudan, A.; Anker, A. S.; Asgari, M.; Blaiszik, B.; Buffo, A.; Choudhury, L.; Corminboeuf, C.; Daglar, H.; Elahi, A. M.; Foster, I. T.; Garcia, S.; Garvin, M.; Godin, G.; Good, L. L.; Gu, J.; Xiao Hu, N.; Jin, X.; Junkers, T.; Keskin, S.; Knowles, T. P. J.; Laplaza, R.; Lessona, M.; Majumdar, S.; Mashhadimoslem, H.; McIntosh, R. D.; Moosavi, S. M.; Mouriño, B.; Nerli, F.; Pevida, C.; Poudineh, N.; Rajabi-Kochi, M.; Saar, K. L.; Hooriabad Saboor, F.; Sagharichiha, M.; Schmidt, K. J.; Shi, J.; Simone, E.; Svatunek, D.; Taddei, M.; Tetko, I.; Tolnai, D.; Vahdatifar, S.; Whitmer, J.; Wieland, D. C. F.; Willumeit-Römer, R.; Züttel, A.; Smit, B. Assessment of fine-tuned large language models for real-world chemistry and material science applications. *Chem. Sci.* **2025**, *16* (2), 670−684.
(25) Jablonka, K. M.; Ai, Q.; Al-Feghali, A.; Badhwar, S.; Bocarsly, J. D.; Bran, A. M.; Bringuier, S.; Brinson, L. C.; Choudhary, K.; Circi, D.; et al. 14 examples of how LLMs can transform materials science and chemistry: a reflection on a large language model hackathon. *Digit. Discovery* **2023**, *2* (5), 1233−1250.
(26) Zheng, Z.; Zhang, O.; Nguyen, H. L.; Rampal, N.; Alawadhi, A. H.; Rong, Z.; Head-Gordon, T.; Borgs, C.; Chayes, J. T.; Yaghi, O. M. ChatGPT research group for optimizing the crystallinity of MOFs and COFs. *ACS Cent. Sci.* **2023**, *9* (11), 2161−2170.
(27) Zheng, Z.; He, Z.; Khattab, O.; Rampal, N.; Zaharia, M. A.; Borgs, C.; Chayes, J. T.; Yaghi, O. M. Image and data mining in reticular chemistry powered by GPT-4V. *Digit. Discovery* **2024**, *3* (3), 491−501.
(28) Fanourgakis, G. S.; Gkagkas, K.; Tylianakis, E.; Froudakis, G. Fast screening of large databases for top performing nanomaterials using a self-consistent, machine learning based approach. *J. Mater. Chem. C* **2020**, *124* (36), 19639−19648.
(29) Ahmed, A.; Siegel, D. J. Predicting hydrogen storage in MOFs via machine learning. *Patterns* **2021**, *2* (7), No. 100291.
(30) Borboudakis, G.; Stergiannakos, T.; Frysali, M.; Klontzas, E.; Tsamardinos, I.; Froudakis, G. E. Chemically intuited, large-scale screening of MOFs by machine learning techniques. *npj Comput. Mater.* **2017**, *3* (1), 40.
(31) Anderson, G.; Schweitzer, B.; Anderson, R.; Gómez-Gualdrón, D. A. Attainable volumetric targets for adsorption-based hydrogen storage in porous crystals: molecular simulation and machine learning. *J. Phys. Chem. C* **2019**, *123* (1), 120−130.
(32) Fanourgakis, G. S.; Gkagkas, K.; Tylianakis, E.; Froudakis, G. E. A universal machine learning algorithm for large-scale screening of materials. *J. Am. Chem. Soc.* **2020**, *142* (8), 3814−3822.
(33) Fernandez, M.; Trefiak, N. R.; Woo, T. K. Atomic property weighted radial distribution functions descriptors of metal−organic frameworks for the prediction of gas uptake capacity. *J. Phys. Chem. C* **2013**, *117* (27), 14095−14105.
(34) Pardakhti, M.; Moharreri, E.; Wanik, D.; Suib, S. L.; Srivastava, R. Machine learning using combined structural and chemical descriptors for prediction of methane adsorption performance of metal organic frameworks (MOFs). *ACS Comb. Sci.* **2017**, *19* (10), 640−645.
(35) Fanourgakis, G. S.; Gkagkas, K.; Tylianakis, E.; Klontzas, E.; Froudakis, G. A robust machine learning algorithm for the prediction of methane adsorption in nanoporous materials. *J. Mater. Chem. A* **2019**, *123* (28), 6080−6087.
(36) Wu, Y.; Duan, H.; Xi, H. Machine learning-driven insights into defects of zirconium metal−organic frameworks for enhanced ethane−ethylene separation. *Chem. Mater.* **2020**, *32* (7), 2986−2997.
(37) Batra, R.; Chen, C.; Evans, T. G.; Walton, K. S.; Ramprasad, R. Prediction of water stability of metal−organic frameworks using machine learning. *Nat. Mach. Intell.* **2020**, *2* (11), 704−710.
(38) Ren, E.; Coudert, F.-X. Prediction of the Diffusion Coefficient through Machine Learning Based on Transition-State Theory Descriptors. *J. Phys. Chem. C* **2024**, *128* (16), 6917−6926.
(39) Menon, D.; Fairen-Jimenez, D. Guiding the rational design of biocompatible metal-organic frameworks for drug delivery. *Matter* **2025**, *8* (3), No. 101958.
(40) Allen, F. H. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, *58* (3), 380−388.
(41) Daglar, H.; Gulbalkan, H. C.; Aksu, G. O.; Keskin, S. Computational simulations of metal−organic frameworks to enhance adsorption applications. *Adv. Mater.* **2024**, No. 2405532.
(42) Liu, T.-W.; Nguyen, Q.; Dieng, A. B.; Gomez-Gualdron, D. Diversity-driven, efficient exploration of a MOF design space to optimize MOF properties. *Chem. Sci.* **2024**, *15*, 18903−18919.
(43) Gantzler, N.; Deshwal, A.; Doppa, J. R.; Simon, C. M. Multi-fidelity Bayesian optimization of covalent organic frameworks for xenon/krypton separations. *Digit. Discovery* **2023**, *2* (6), 1937−1956.
(44) Liu, X.; Wang, R.; Wang, X.; Xu, D. High-throughput computational screening and machine learning model for accelerated metal−organic frameworks discovery in toluene vapor adsorption. *J. Phys. Chem. C* **2023**, *127* (23), 11268−11282.
(45) Altintas, C.; Altundal, O. F.; Keskin, S.; Yildirim, R. Machine learning meets with metal organic frameworks for gas storage and separation. *J. Chem. Inf. Model.* **2021**, *61* (5), 2131−2146.
(46) Du, R.; Xin, R.; Wang, H.; Zhu, W.; Li, R.; Liu, W. Machine learning: An accelerator for the exploration and application of advanced metal-organic frameworks. *Chem. Eng. J.* **2024**, *490*, No. 151828.
(47) Li, C.; Bao, L.; Ji, Y.; Tian, Z.; Cui, M.; Shi, Y.; Zhao, Z.; Wang, X. Combining machine learning and metal−organic frameworks research: Novel modeling, performance prediction, and materials discovery. *Coord. Chem. Rev.* **2024**, *514*, No. 215888.
(48) Demir, H.; Keskin, S. A new era of modeling MOF-based membranes: cooperation of theory and data science. *Macromol. Mater. Eng.* **2024**, *309* (1), No. 2300225.
(49) Moghadam, P. Z.; Chung, Y. G.; Snurr, R. Q. Progress toward the computational discovery of new metal−organic framework adsorbents for energy applications. *Nat. Energy* **2024**, *9* (2), 121−133.
(50) Neikha, K.; Puzari, A. Metal−organic frameworks through the lens of artificial intelligence: a comprehensive review. *Langmuir* **2024**, *40* (42), 21957−21975.
(51) Park, J.; Kim, H.; Kang, Y.; Lim, Y.; Kim, J. From data to discovery: recent trends of machine learning in metal−organic frameworks. *JACS Au* **2024**, *4* (10), 3727−3743.
(52) LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521* (7553), 436−444.
(53) Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; Fei-Fei, L.Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*; IEEE: **2009**, (pp 248−255).
(54) Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. Large-scale screening of hypothetical metal−organic frameworks. *Nat. Chem.* **2012**, *4* (2), 83−89.
(55) Boyd, P. G.; Chidambaram, A.; García-Díez, E.; Ireland, C. P.; Daff, T. D.; Bounds, R.; Gładysiak, A.; Schouwink, P.; Moosavi, S. M.; Maroto-Valer, M. M.; Reimer, J. A.; Navarro, J. A. R.; Woo, T. K.; Garcia, S.; Stylianou, K. C.; Smit, B. Data-driven design of metal−organic frameworks for wet flue gas CO_2 capture. *Nature* **2019**, *576* (7786), 253−256.
(56) Chung, Y. G.; Haldoupis, E.; Bucior, B. J.; Haranczyk, M.; Lee, S.; Zhang, H.; Vogiatzis, K. D.; Milisavljevic, M.; Ling, S.; Camp, J. S.; Slater, B.; Siepmann, J. I.; Sholl, D. S.; Snurr, R. Q. Advances, updates, and analytics for the computation-ready, experimental metal−organic framework database: CoRE MOF 2019. *J. Chem. Eng. Data* **2019**, *64* (12), 5985−5998.

(57) Rosen, A. S.; Iyer, S. M.; Ray, D.; Yao, Z.; Aspuru-Guzik, A.; Gagliardi, L.; Notestein, J. M.; Snurr, R. Q. Machine learning the quantum-chemical properties of metal−organic frameworks for accelerated materials discovery. *Matter* **2021**, *4* (5), 1578−1597.
(58) Glasby, L. T.; Gubsch, K.; Bence, R.; Oktavian, R.; Isoko, K.; Moosavi, S. M.; Cordiner, J. L.; Cole, J. C.; Moghadam, P. Z. DigiMOF: a database of metal−organic framework synthesis information generated via text mining. *Chem. Mater.* **2023**, *35* (11), 4510−4524.
(59) Burner, J.; Luo, J.; White, A.; Mirmiran, A.; Kwon, O.; Boyd, P. G.; Maley, S.; Gibaldi, M.; Simrod, S.; Ogden, V.; Woo, T. K. ARC−MOF: a diverse database of metal-organic frameworks with DFT-derived partial atomic charges and descriptors for machine learning. *Chem. Mater.* **2023**, *35* (3), 900−916.
(60) Bobbitt, N. S.; Shi, K.; Bucior, B. J.; Chen, H.; Tracy-Amoroso, N.; Li, Z.; Sun, Y.; Merlin, J. H.; Siepmann, J. I.; Siderius, D. W.; Snurr, R. Q. MOFX-DB: An online database of computational adsorption data for nanoporous materials. *J. Chem. Eng. Data* **2023**, *68* (2), 483−498.
(61) Sriram, A.; Choi, S.; Yu, X.; Brabson, L. M.; Das, A.; Ulissi, Z.; Uyttendaele, M.; Medford, A. J.; Sholl, D. S. The Open DAC 2023 dataset and challenges for sorbent discovery in direct air capture. *ACS Cent. Sci.* **2024**, *10* (5), 923−941.
(62) Sarikas, A. P.; Gkagkas, K.; Froudakis, G. E. Gas adsorption meets deep learning: voxelizing the potential energy surface of metal-organic frameworks. *Sci. Rep.* **2024**, *14* (1), 2242.
(63) Huang, H.; Magar, R.; Xu, C.; Farimani, A. B.Materials informatics transformer: A language model for interpretable materials properties prediction. *arXiv***2023**.
(64) Wang, J.; Liu, J.; Wang, H.; Ke, G.; Zhang, L.; Wu, J.; Gao, Z.; Lu, D.Metal-organic frameworks meet Uni-MOF: a revolutionary gas adsorption detector. *ChemRxiv***2023** DOI: 10.26434/chemrxiv-2023-v9jwh-v2
(65) Moghadam, P. Z.; Rogge, S. M. J.; Li, A.; Chow, C. M.; Wieme, J.; Moharrami, N.; Aragones-Anglada, M.; Conduit, G.; Gomez-Gualdron, D. A.; Van Speybroeck, V.; Fairen-Jimenez, D. Structure-mechanical stability relations of metal-organic frameworks via machine learning. *Matter* **2019**, *1* (1), 219−234.
(66) Thornton, A. W.; Simon, C. M.; Kim, J.; Kwon, O.; Deeg, K. S.; Konstas, K.; Pas, S. J.; Hill, M. R.; Winkler, D. A.; Haranczyk, M.; Smit, B. Materials genome in action: identifying the performance limits of physical hydrogen storage. *Chem. Mater.* **2017**, *29* (7), 2844−2854.
(67) Sun, Y.; DeJaco, R. F.; Li, Z.; Tang, D.; Glante, S.; Sholl, D. S.; Colina, C. M.; Snurr, R. Q.; Thommes, M.; Hartmann, M.; Siepmann, J. I. Fingerprinting diverse nanoporous materials for optimal hydrogen storage conditions using meta-learning. *Sci. Adv.* **2021**, *7* (30), No. eabg3983.
(68) Zhang, X.; Zhang, K.; Yoo, H.; Lee, Y. Machine learning-driven discovery of metal−organic frameworks for efficient CO_2 capture in humid condition. *ACS Sustain. Chem. Eng.* **2021**, *9* (7), 2872−2879.
(69) Burner, J.; Schwiedrzik, L.; Krykunov, M.; Luo, J.; Boyd, P. G.; Woo, T. K. High-performing deep learning regression models for predicting low-pressure CO_2 adsorption properties of metal−organic frameworks. *J. Phys. Chem. C* **2020**, *124* (51), 27996−28005.
(70) Nandy, A.; Terrones, G.; Arunachalam, N.; Duan, C.; Kastner, D. W.; Kulik, H. J. MOFSimplify, machine learning models with extracted stability data of three thousand metal−organic frameworks. *Sci. Data* **2022**, *9* (1), 74.
(71) Park, H.; Kang, Y.; Choe, W.; Kim, J. Mining insights on metal−organic framework synthesis from scientific literature texts. *J. Chem. Inf. Model.* **2022**, *62* (5), 1190−1198.
(72) Li, W.; Situ, Y.; Ding, L.; Chen, Y.; Yang, Q. MOF-GRU: A MOFid-aided deep learning model for predicting the gas separation performance of metal-organic frameworks. *ACS Appl. Mater. Interfaces* **2023**, *15* (51), 59887−59894.
(73) Cleeton, C.; Sarkisov, L.Design of metal-organic frameworks using deep dreaming approaches. *ChemRxiv***2024**. DOI: DOI: 10.26434/chemrxiv-2024-9q39w-v2.
(74) Raza, A.; Sturluson, A.; Simon, C. M.; Fern, X. Message passing neural networks for partial charge assignment to metal−organic frameworks. *J. Phys. Chem. C* **2020**, *124* (35), 19070−19082.
(75) Wang, R.; Zhong, Y.; Bi, L.; Yang, M.; Xu, D. Accelerating discovery of metal−organic frameworks for methane adsorption with hierarchical screening and deep learning. *ACS Appl. Mater. Interfaces* **2020**, *12* (47), 52797−52807.
(76) Rosen, A. S.; Fung, V.; Huck, P.; O'Donnell, C. T.; Horton, M. K.; Truhlar, D. G.; Persson, K. A.; Notestein, J. M.; Snurr, R. Q. High-throughput predictions of metal−organic framework electronic properties: theoretical challenges, graph neural networks, and data exploration. *npj Comput. Mater.* **2022**, *8* (1), 112.
(77) Chen, P.; Jiao, R.; Liu, J.; Liu, Y.; Lu, Y. Interpretable graph transformer network for predicting adsorption isotherms of metal−organic frameworks. *J. Chem. Inf. Model.* **2022**, *62* (22), 5446−5456.
(78) Jalali, M.; Wonanke, A. D. D.; Wöll, C. MOFGalaxyNet: a social network analysis for predicting guest accessibility in metal−organic frameworks utilizing graph convolutional networks. *J. Cheminf.* **2023**, *15* (1), 94.
(79) Zhao, G.; Chung, Y. G. PACMAN: A Robust Partial Atomic Charge Predicter for Nanoporous Materials Based on Crystal Graph Convolution Networks. *J. Chem. Theory Comput.* **2024**, *20* (12), 5368−5380.
(80) Korolev, V.; Mitrofanov, A. Coarse-grained crystal graph neural networks for reticular materials design. *J. Chem. Inf. Model.* **2024**, *64* (6), 1919−1931.
(81) Cho, E. H.; Lin, L.-C. Nanoporous material recognition via 3D convolutional neural networks: Prediction of adsorption properties. *J. Phys. Chem. Lett.* **2021**, *12* (9), 2279−2285.
(82) Hung, T.-H.; Xu, Z.-X.; Kang, D.-Y.; Lin, L.-C. Chemistry-encoded convolutional neural networks for predicting gaseous adsorption in porous materials. *J. Mater. Chem. C* **2022**, *126* (5), 2813−2822.
(83) Cao, Z.; Magar, R.; Wang, Y.; Barati Farimani, A. MOFormer: self-supervised transformer model for metal−organic framework property prediction. *J. Am. Chem. Soc.* **2023**, *145* (5), 2958−2967.
(84) Kang, Y.; Park, H.; Smit, B.; Kim, J. A multi-modal pre-training transformer for universal transfer learning in metal−organic frameworks. *Nat. Mach. Intell.* **2023**, *5* (3), 309−318.
(85) Park, H.; Kang, Y.; Kim, J. Enhancing structure−property relationships in porous materials through transfer learning and cross-material few-shot learning. *ACS Appl. Mater. Interfaces* **2023**, *15* (48), 56375−56385.
(86) Cui, J.; Wu, F.; Zhang, W.; Yang, L.; Hu, J.; Fang, Y.; Ye, P.; Zhang, Q.; Suo, X.; Mo, Y.; Cui, X.; Chen, H.; Xing, H. Direct prediction of gas adsorption via spatial atom interaction learning. *Nat. Commun.* **2023**, *14* (1), 7043.
(87) Wang, J.; Liu, J.; Wang, H.; Zhou, M.; Ke, G.; Zhang, L.; Wu, J.; Gao, Z.; Lu, D. A comprehensive transformer-based approach for high-accuracy gas adsorption predictions in metal-organic frameworks. *Nat. Commun.* **2024**, *15* (1), 1904.
(88) Vandenhaute, S.; Cools-Ceuppens, M.; DeKeyser, S.; Verstraelen, T.; Van Speybroeck, V. Machine learning potentials for metal-organic frameworks using an incremental learning approach. *npj Comp. Mater.* **2023**, *9* (1), 19.
(89) Sharma, A.; Sanvito, S. Quantum-Accurate Machine Learning Potentials for Metal-Organic Frameworks using Temperature Driven Active Learning. *npj Comput. Mater.* **2024**, *10*, 237.
(90) Wieser, S.; Zojer, E. Machine learned force-fields for an Ab-initio quality description of metal-organic frameworks. *npj Comput. Mater.* **2024**, *10* (1), 18.
(91) Park, J.; Lim, Y.; Lee, S.; Kim, J. Computational design of metal−organic frameworks with unprecedented high hydrogen working capacity and high synthesizability. *Chem. Mater.* **2023**, *35* (1), 9−16.
(92) Dai, T.; Vijayakrishnan, S.; Szczypiński, F. T.; Ayme, J. F.; Simaei, E.; Fellowes, T.; Clowes, R.; Kotopanov, L.; Shields, C. E.; Zhou, Z.; Ward, J. W.; Cooper, A. I.; et al. Autonomous mobile robots for exploratory synthetic chemistry. *Nature* **2024**, *635*, 890−897.

(93) Tom, G.; Schmid, S. P.; Baird, S. G.; Cao, Y.; Darvish, K.; Hao, H.; Lo, S.; Pablo-García, S.; Rajaonson, E. M.; Skreta, M.; Yoshikawa, N.; Corapi, S.; Akkoc, G. D.; Strieth-Kalthoff, F.; Seifrid, M.; Aspuru-Guzik, A. Self-driving laboratories for chemistry and materials science. *Chem. Rev.* **2024**, *124* (16), 9633−9732.
(94) Seavill, P. Research with robotics. *Nat. Synth.* **2023**, *2* (6), 467−468.
(95) MacLeod, B. P.; Parlane, F. G. L.; Brown, A. K.; Hein, J. E.; Berlinguette, C. P. Flexible automation accelerates materials discovery. *Nat. Mater.* **2022**, *21* (7), 722−726.
(96) Luo, Y.; Bag, S.; Zaremba, O.; Cierpka, A.; Andreo, J.; Wuttke, S.; Friederich, P.; Tsotsalas, M. MOF synthesis prediction enabled by automatic data mining and machine learning. *Angew. Chem., Int. Ed.* **2022**, *61* (19), No. e202200242.
(97) Taddei, M.; Dau, P. V.; Cohen, S. M.; Ranocchiari, M.; van Bokhoven, J. A.; Costantino, F.; Sabatini, S.; Vivani, R. Efficient microwave assisted synthesis of metal−organic framework UiO-66: optimization and scale up. *Dalton Trans.* **2015**, *44* (31), 14019−14026.
(98) Ryu, U.; Jee, S.; Rao, P. C.; Shin, J.; Ko, C.; Yoon, M.; Park, K. S.; Choi, K. M. Recent advances in process engineering and upcoming applications of metal−organic frameworks. *Coord. Chem. Rev.* **2021**, *426*, No. 213544.
(99) Wright, A. M.; Kapelewski, M. T.; Marx, S.; Farha, O. K.; Morris, W. Transitioning metal−organic frameworks from the laboratory to market through applied research. *Nat. Mater.* **2025**, *24* (2), 178−187.
(100) De Vos, J. S.; Ravichandran, S.; Borgmans, S.; Vanduyfhuys, L.; Van Der Voort, P.; Rogge, S. M. J.; Van Speybroeck, V. High-throughput screening of covalent organic frameworks for carbon capture using machine learning. *Chem. Mater.* **2024**, *36* (9), 4315−4330.
(101) Budennyy, S. A.; Lazarev, V. D.; Zakharenko, N. N.; Korovin, A. N.; Plosskaya, O. A.; Dimitrov, D. V.; Akhripkin, V. S.; Pavlov, I. V.; Oseledets, I. V.; Barsola, I. S.; Egorov, I. V.; Kosterina, A. A.; Zhukov, L. E. ECO2AI: carbon emissions tracking of machine learning models as the first step towards sustainable AI. *Dokl. Math.* **2022**, *106* (Suppl 1), S118−S128.
(102) Korolev, V.; Mitrofanov, A. The carbon footprint of predicting CO_2 storage capacity in metal-organic frameworks within neural networks. *iScience* **2024**, *27* (5), No. 109644.
(103) Moghadam, P. Z.; Li, A.; Wiggin, S. B.; Tao, A.; Maloney, A. G. P.; Wood, P. A.; Ward, S. C.; Fairen-Jimenez, D. Development of a Cambridge Structural Database subset: a collection of metal−organic frameworks for past, present, and future. *Chem. Mater.* **2017**, *29* (7), 2618−2625.
(104) Bucior, B. J.; Rosen, A. S.; Haranczyk, M.; Yao, Z.; Ziebel, M. E.; Farha, O. K.; Hupp, J. T.; Siepmann, J. I.; Aspuru-Guzik, A.; Snurr, R. Q. Identification schemes for metal−organic frameworks to enable rapid search and cheminformatics analysis. *Cryst. Growth Des.* **2019**, *19* (11), 6682−6697.
(105) Tshitoyan, V.; Dagdelen, J.; Weston, L.; Dunn, A.; Rong, Z.; Kononova, O.; Persson, K. A.; Ceder, G.; Jain, A. Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature* **2019**, *571* (7763), 95−98.
(106) Ward, L.; Dunn, A.; Faghaninia, A.; Zimmermann, N. E. R.; Bajaj, S.; Wang, Q.; Montoya, J.; Chen, J.; Bystrom, K.; Dylla, M.; Chard, K.; Asta, M.; Persson, K. A.; Snyder, G. J.; Foster, I.; Jain, A. Matminer: An open source toolkit for materials data mining. *Comput. Mater. Sci.* **2018**, *152*, 60−69.
(107) Dutta, S.; Walden, M.; Sinelshchikova, A.; Ettlinger, R.; Lizundia, E.; Wuttke, S. Cradle-to-Gate Environmental Impact Assessment of Commercially Available Metal-Organic Frameworks Manufacturing. *Adv. Funct. Mater.* **2024**, *34* (52), No. 2410751.
(108) Wang, H.; Zhang, L.; Han, J.; Weinan, E. DeePMD-kit: A deep learning package for many-body potential energy representation and molecular dynamics. *Comput. Phys. Commun.* **2018**, *228*, 178−184.
(109) Pinheiro, M.; Ge, F.; Ferré, N.; Dral, P. O.; Barbatti, M. Choosing the right molecular machine learning potential. *Chem. Sci.* **2021**, *12* (43), 14396−14413.
(110) Deringer, V. L.; Caro, M. A.; Csányi, G. A general-purpose machine-learning force field for bulk and nanostructured phosphorus. *Nat. Commun.* **2020**, *11* (1), 5461.
(111) Schran, C.; Thiemann, F. L.; Rowe, P.; Müller, E. A.; Marsalek, O.; Michaelides, A. Machine learning potentials for complex aqueous systems made simple. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118* (38), No. e2110077118.
(112) Castel, N.; André, D.; Edwards, C.; Evans, J. D.; Coudert, F.-X. Machine learning interatomic potentials for amorphous zeolitic imidazolate frameworks. *Digit. Discovery* **2024**, *3* (2), 355−368.
(113) Ying, P.; Liang, T.; Xu, K.; Zhang, J.; Xu, J.; Zhong, Z.; Fan, Z. Sub-micrometer phonon mean free paths in metal−organic frameworks revealed by machine learning molecular dynamics simulations. *ACS Appl. Mater. Interfaces* **2023**, *15* (30), 36412−36422.
(114) Fan, D.; Naskar, S.; Maurin, G. Unconventional mechanical and thermal behaviours of MOF CALF-20. *Nat. Commun.* **2024**, *15* (1), 3251.
(115) Fan, D.; Ozcan, A.; Lyu, P.; Maurin, G. Unravelling abnormal in-plane stretchability of two-dimensional metal−organic frameworks by machine learning potential molecular dynamics. *Nanoscale* **2024**, *16* (7), 3438−3447.
(116) Goeminne, R.; Vanduyfhuys, L.; Van Speybroeck, V.; Verstraelen, T. DFT-Quality adsorption simulations in metal−organic frameworks enabled by machine learning Potentials. *J. Chem. Theory Comput.* **2023**, *19* (18), 6313−6325.
(117) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **1992**, *114* (25), 10024−10035.
(118) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. DREIDING: a generic force field for molecular simulations. *J. Phys. Chem.* **1990**, *94* (26), 8897−8909.
(119) Kökçam-Demir, Ü.; Goldman, A.; Esrafili, L.; Gharib, M.; Morsali, A.; Weingart, O.; Janiak, C. Coordinatively unsaturated metal sites (open metal sites) in metal−organic frameworks: design and applications. *Chem. Soc. Rev.* **2020**, *49* (9), 2751−2798.
(120) Formalik, F.; Shi, K.; Joodaki, F.; Wang, X.; Snurr, R. Q. Exploring the structural, dynamic, and functional properties of metal-organic frameworks through molecular modeling. *Adv. Funct. Mater.* **2023**, *34*, No. 2308130.
(121) Deng, B.; Zhong, P.; Jun, K.; Riebesell, J.; Han, K.; Bartel, C. J.; Ceder, G. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nat. Mach. Intell.* **2023**, *5* (9), 1031−1041.
(122) Chen, C.; Ong, S. P. A universal graph deep learning interatomic potential for the periodic table. *Nat. Comput. Sci.* **2022**, *2* (11), 718−728.
(123) Batatia, I.; Benner, P.; Chiang, Y.; Elena, A. M.; Kovács, D. P.; Riebesell, J.; Advincula, X. R.; Asta, M.; Avaylon, M.; Baldwin, W. J.A foundation model for atomistic materials chemistry. *arXiv***2023**
(124) Doyle, A. C.*The adventures of Sherlock Holmes*; Wordsworth Editions: 1992.
(125) European Commission: Directorate-General for Research and Innovation*Cost-benefit analysis for FAIR research data − Cost of not having FAIR research data*. Publications Office: 2018. https://data.europa.eu/doi/10.2777/02999.
(126) Wilkinson, M. D.; Dumontier, M.; Aalbersberg, I. J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J. W.; da Silva Santos, L. B.; Bourne, P. E.; Bouwman, J.; Brookes, A. J.; Clark, T.; Crosas, M.; Dillo, I.; Dumon, O.; Edmunds, S.; Evelo, C. T.; Finkers, R.; Gonzalez-Beltran, A.; Gray, A. J. G.; Groth, P.; Goble, C.; Grethe, J. S.; Heringa, J.; 't Hoen, P. A. C.; Hooft, R.; Kuhn, T.; Kok, R.; Kok, J.; Lusher, S. J.; Martone, M. E.; Mons, A.; Packer, A. L.; Persson, B.; Rocca-Serra, P.; Roos, M.; van Schaik, R.; Sansone, S. A.; Schultes, E.; Sengstag, T.; Slater, T.; Strawn, G.; Swertz, M. A.; Thompson, M.; van der Lei, J.; van Mulligen, E.; Velterop, J.; Waagmeester, A.;

https://doi.org/10.1021/jacs.5c08214
J. Am. Chem. Soc. 2025, 147, 23367−23380

Wittenburg, P.; Wolstencroft, K.; Zhao, J.; Mons, B. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016**, *3* (1), 160018.
(127) Artrith, N.; Butler, K. T.; Coudert, F.-X.; Han, S.; Isayev, O.; Jain, A.; Walsh, A. Best practices in machine learning for chemistry. *Nat. Chem.* **2021**, *13* (6), 505−508.
(128) Gardner, A.; Smith, A. L.; Steventon, A.; Coughlan, E.; Oldfield, M. Ethical funding for trustworthy AI: proposals to address the responsibilities of funders to ensure that projects adhere to trustworthy AI practice. *AI Ethics* **2022**, *2*, 277−291.
(129) Horton, M. K.; Dwaraknath, S.; Persson, K. A. Promises and perils of computational materials databases. *Nat. Comput. Sci.* **2021**, *1* (1), 3−5.
(130) Gibaldi, M.; Kapeliukha, A.; White, A.; Luo, J.; Mayo, R. A.; Burner, J.; Woo, T. K. MOSAEC-DB: a comprehensive database of experimental metal−organic frameworks with verified chemical accuracy suitable for molecular simulations. *Chem. Sci.* **2025**, *16*, 4085−4100.
(131) De Vos, J. S.; Borgmans, S.; Van Der Voort, P.; Rogge, S. M. J.; Van Speybroeck, V. ReDD-COFFEE: a ready-to-use database of covalent organic framework structures and accurate force fields to enable high-throughput screenings. *J. Mater. Chem. A* **2023**, *11* (14), 7468−7487.
(132) Ongari, D.; Yakutovich, A. V.; Talirz, L.; Smit, B. Building a consistent and reproducible database for adsorption evaluation in covalent−organic frameworks. *ACS Cent. Sci.* **2019**, *5* (10), 1663−1675.
(133) Chung, Y. G.; Camp, J.; Haranczyk, M.; Sikora, B. J.; Bury, W.; Krungleviciute, V.; Yildirim, T.; Farha, O. K.; Sholl, D. S.; Snurr, R. Q. Computation-ready, experimental metal−organic frameworks: A tool to enable high-throughput screening of nanoporous crystals. *Chem. Mater.* **2014**, *26* (21), 6185−6192.
(134) Tong, M.; Lan, Y.; Qin, Z.; Zhong, C. Computation-ready, experimental covalent organic framework for methane delivery: screening and material design. *J. Phys. Chem. C* **2018**, *122* (24), 13009−13016.
(135) Gómez-Gualdrón, D. A.; Colón, Y. J.; Zhang, X.; Wang, T. C.; Chen, Y.-S.; Hupp, J. T.; Yildirim, T.; Farha, O. K.; Zhang, J.; Snurr, R. Q. Evaluating topologically diverse metal−organic frameworks for cryo-adsorbed hydrogen storage. *Energy Environ. Sci.* **2016**, *9* (10), 3279−3289.
(136) Martin, R. L.; Simon, C. M.; Medasani, B.; Britt, D. K.; Smit, B.; Haranczyk, M. In silico design of three-dimensional porous covalent organic frameworks via known synthesis routes and commercially available species. *J. Phys. Chem. C* **2014**, *118* (41), 23790−23802.
(137) Mercado, R.; Fu, R.-S.; Yakutovich, A. V.; Talirz, L.; Haranczyk, M.; Smit, B. In silico design of 2D and 3D covalent organic frameworks for methane storage applications. *Chem. Mater.* **2018**, *30* (15), 5069−5086.
(138) Narayanan, A.; Chandramohan, M.; Venkatesan, R.; Chen, L.; Liu, Y.; Jaiswal, S.graph2vec: Learning distributed representations of graphs. *arXiv***2017** DOI: 10.48550/arXiv.1707.05005.
(139) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. *Microporous Mesoporous Mater.* **2012**, *149* (1), 134−141.
(140) Krishnapriyan, A. S.; Montoya, J.; Haranczyk, M.; Hummelshøj, J.; Morozov, D. Machine learning with persistent homology and chemical word embeddings improves prediction accuracy and interpretability in metal-organic frameworks. *Sci. Rep.* **2021**, *11* (1), 8888.
(141) Moosavi, S. M.; Nandy, A.; Jablonka, K. M.; Ongari, D.; Janet, J. P.; Boyd, P. G.; Lee, Y.; Smit, B.; Kulik, H. J. Understanding the diversity of the metal-organic framework ecosystem. *Nat. Commun.* **2020**, *11* (1), 4068.
(142) Zhang, X.; Jablonka, K. M.; Smit, B. Deep Learning-Based Recommendation System for Metal-Organic Frameworks (MOFs). *Digit. Discovery* **2024**, *3*, 1410−1420.
(143) Tang, H.; Duan, L.; Jiang, J. Leveraging Machine Learning for Metal-Organic Frameworks: A Perspective. *Langmuir* **2023**, *39* (45), 15849−15863.
(144) Jablonka, K. M.; Ongari, D.; Moosavi, S. M.; Smit, B. Using collective knowledge to assign oxidation states of metal cations in metal−organic frameworks. *Nat. Chem.* **2021**, *13*, 771−777.
(145) Nandy, A.; Yue, S.; Oh, C.; Duan, C.; Terrones, G. G.; Chung, Y. G.; Kulik, H. J. A database of ultrastable MOFs reassembled from stable fragments with machine learning models. *Matter* **2023**, *6*, 1585−1603.
(146) White, A. D.; Hocky, G. M.; Gandhi, H. A.; Ansari, M.; Cox, S.; Wellawatte, G. P.; Sasmal, S.; Yang, Z.; Liu, K.; Singh, Y.; et al. Assessment of chemistry knowledge in large language models that generate code. *Digit. Discovery* **2023**, *2*, 368−376.
(147) Pollice, R.; dos Passos Gomes, G.; Aldeghi, M.; Hickman, R. J.; Krenn, M.; Lavigne, C.; Lindner-D'Addario, M.; Nigam, A.; Ser, C. T.; Yao, Z.; Aspuru-Guzik, A. Data-Driven Strategies for Accelerated Materials Design. *Acc. Chem. Res.* **2021**, *54*, 849−860.
(148) Park, H.; Majumdar, S.; Zhang, X.; Kim, J.; Smit, B. Inverse design of metal−organic frameworks for direct air capture of CO_2 via deep reinforcement learning. *Digit. Discovery* **2024**, *3*, 728−741.

https://doi.org/10.1021/jacs.5c08214
J. Am. Chem. Soc. 2025, 147, 23367−23380