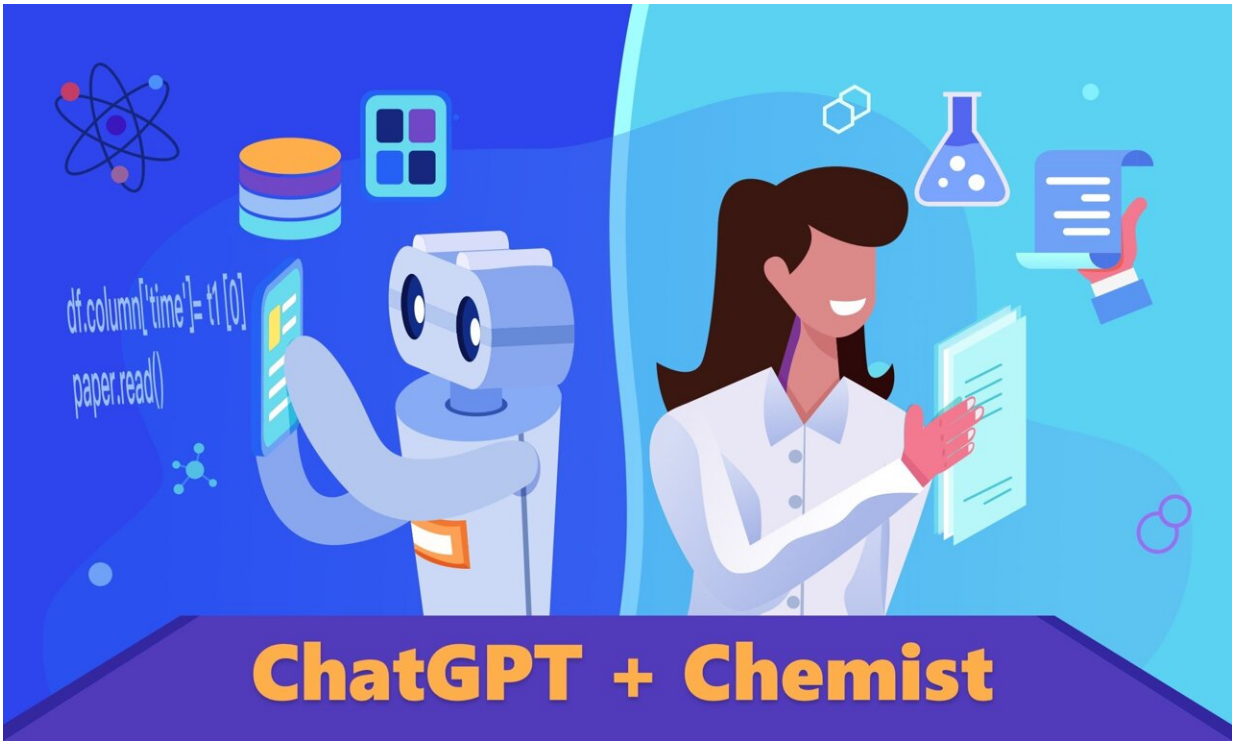
Turning ChatGPT into a 'chemistry assistant'

August 10 2023



This illustration shows ChatGPT and chemists teaming up to glean new insights on how to make MOFs, which could have applications in clean energy. Credit: Adapted from the *Journal of the American Chemical Society*, 2023, DOI: 10.1021/jacs.3c05819

Developing new materials requires significant time and labor, but some chemists are now hopeful that artificial intelligence (AI) could one day

shoulder much of this burden. In a new study in the *Journal of the American Chemical Society*, a team prompted a popular AI model, ChatGPT, to perform one particularly time-consuming task: searching scientific literature. With that data, they built a second tool, a model to predict experimental results.

Reports from previous studies offer a vast trove of information that chemists need, but finding and parsing the most relevant details can be laborious. For example, those interested in designing highly porous, crystalline metal-organic frameworks (MOFs)—which have potential applications in areas such as clean energy—must sort through hundreds of scientific papers describing a variety of experimental conditions.

Researchers have previously attempted to coax AI to take over this task; however, the language processing models they used required significant technical expertise, and applying them to new topics meant changing the program. Omar Yaghi and colleagues wanted to see if the next generation of language models, which includes ChatGPT, could offer a more accessible, flexible way to extract information.

To analyze text from scientific papers, the team gave ChatGPT prompts, or instructions, guiding it through three processes intended to identify and summarize the experimental information the manuscripts contained. The researchers carefully constructed these prompts to minimize the model's tendency to make up responses, a phenomenon known as hallucination, and to ensure the best responses possible.

When tested on 228 papers describing MOF syntheses, this system extracted more than 26,000 factors relevant for making roughly 800 of these compounds. With these data, the team trained a separate AI model to predict the crystalline state of MOFs based on these conditions.

And finally, to make the data more user friendly, they built a chatbot to

answer questions about it. The team notes that, unlike previous AI-based efforts, this one does not require expertise in coding. What's more, scientists can shift its focus simply by adjusting the narrative language in the prompts. This new system, which they dub the "ChatGPT Chemistry Assistant," could also be useful in other fields of chemistry, according to the researchers.

More information: Zhiling Zheng et al, ChatGPT Chemistry Assistant for Text Mining and the Prediction of MOF Synthesis, *Journal of the American Chemical Society* (2023). DOI: 10.1021/jacs.3c05819

Provided by American Chemical Society

METAL-ORGANIC FRAMEWORKS

ChatGPT lab assistant accelerates MOF synthesis

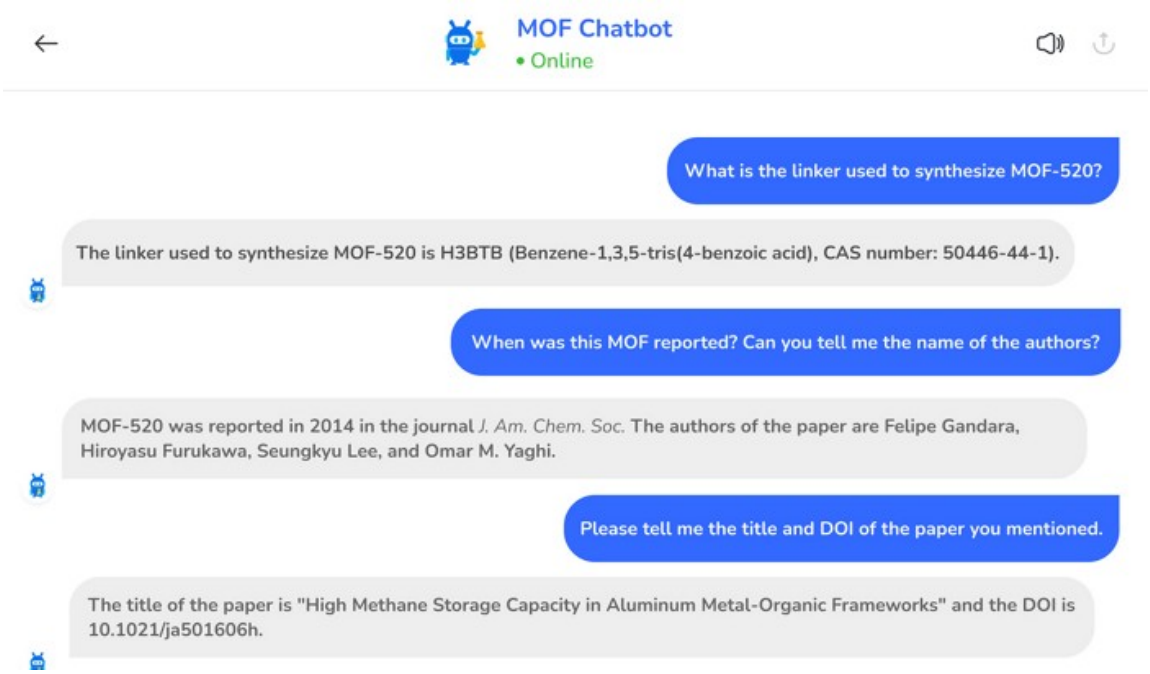
An algorithm analyzes synthesis datasets to provide precise predictions and instructions to lab chemists

by ***Fernando Gomollón-Bel, special to C&EN***

August 11, 2023

c&en CHEMICAL & ENGINEERING NEWS

Researchers have trained ChatGPT to create a chemistry lab assistant that summarizes information about synthesis from papers with high accuracy (*J. Am. Chem. Soc.* 2023, DOI: **10.1021/jacs.3c05819**). In particular, this program extracts over 26,000 parameters from peer-reviewed articles and supporting information about **metal-organic frameworks (MOFs)**. Once trained, the interactive chatbot is able to answer questions about the preparation of MOFs quickly and accurately.



Credit: ***J. Am. Chem. Soc.***

The chemistry chatbot based on ChatGPT creates an interface for assisted literature searches.

"We've always been interested in simplifying and speeding up chemical synthesis," says Omar Yaghi from the University of California, Berkeley, lead author of the study. The ChatGPT models mined the supporting information of hundreds of MOF papers, where information on synthesis is unstructured and sparse, often extended over hundreds of pages. Thus, "we developed a filtering strategy that excludes the least relevant sections—like references, crystal coordinates, acknowledgments—increasing the efficiency," adds Yaghi.

Large language models, like ChatGPT, can be prone to what are called hallucinations. These are responses that seem correct, but aren't. The team minimized the emergence of misleading affirmations with careful prompt engineering. "It's a means of training ChatGPT," says Yaghi. "We ensure the prompt contains information . . . to help improve the response." This approach advises

MOST POPULAR IN MATERIALS

Flow batteries, the forgotten energy storage device

Chemistry in Pictures: Odd MOF

Why glass recycling in the US is brok

Chemists roll out rings made from stacked sandwich complexes

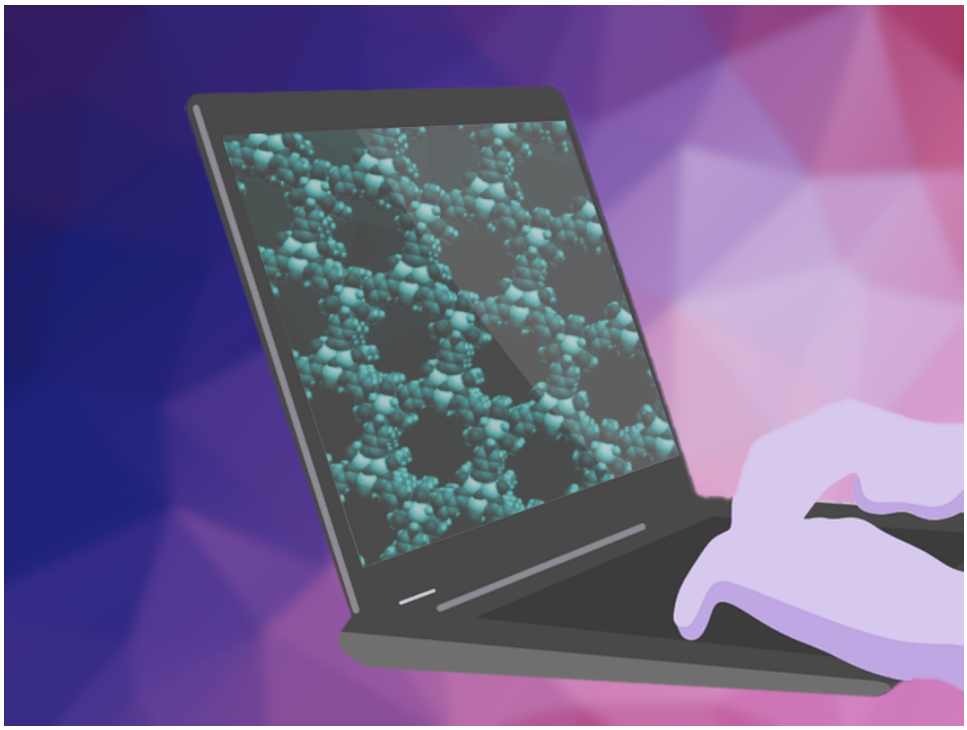
It's time to get serious about recyclin lithium-ion batteries

the algorithm to recognize uncertainty, rather than fabricating fake answers. For example, when asked about a MOF not present in the training database, the program will simply say: “I do not know.” Additionally, this iterative approach helps researchers get structured responses such as bulleted lists and step-by-step synthesis of many MOFs, which reference the correct sources.

This process of curation could have taken a chemist months, but ChatGPT scans and registers synthetic procedures in a fraction of the time, says Yaghi. “It only takes one minute per paper.”

The team envisions researchers applying their publicly available ChatGPT model to other fields of chemistry, after **training it with the relevant papers and datasets**. Eventually, the chatbot could predict the outcome of chemical reactions, or propose potential synthetic routes leveraging its knowledge and understanding of the chemical space.

Chemical & Engineering News
ISSN 0009-2347



Author: ChemistryViews
Published On: August 9, 2023
Copyright: Wiley-VCH GmbH

Category: NEWS
Subject: INORGANIC | MATERIALS | ORGANIC
Related society: AMERICAN CHEMICAL SOCIETY (ACS), USA

NEWS

ChatGPT Used for Text Mining of MOF Syntheses in the Literature

Large language models, such as the GPT series of models used in ChatGPT, are trained using large amounts of text and can predict the probabilities of series of words in a given language. This can be used for a variety of applications, e.g., to generate a probable text output based on a user input. The chemical literature also contains vast amounts of and performing a comprehensive literature review and extracting

useful data and insights for a specific application quickly can be challenging. Large language models could help with this issue.

Omar M. Yaghi, University of California, Berkeley, USA, and King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia, and colleagues have used ChatGPT to automate text mining and quickly create datasets on difficult-to-aggregate research about metal-organic frameworks (MOFs). The team curated 228 relevant peer-reviewed research papers, and then used ChatGPT to process the relevant sections in the papers and to extract, clean up, and organize the data. ChatGPT successfully extracted 26,257 distinct synthesis parameters for ca. 800 MOFs reported in the selected research articles. It mined the synthetic conditions of the MOFs with high accuracy and very quickly.

The extracted datasets can then be used to inform predictive models, which might help chemists to develop new MOFs. Using the data gathered by text mining, the team created a machine-learning model that achieved 87 % accuracy in predicting MOF experimental crystallization outcomes. According to the researchers, the text-mining approach can be easily transferred to other contexts with minimal coding knowledge. Further exploration of large language models for AI-assisted chemistry could, thus, be useful for accelerating research.

- ChatGPT Chemistry Assistant for Text Mining and the Prediction of MOF Synthesis,
  Zhiling Zheng, Oufan Zhang, Christian Borgs, Jennifer T. Chayes, Omar M. Yaghi,
  J. Am. Chem. Soc. **2023**.
  https://doi.org/10.1021/jacs.3c05819

Leave a Reply

Kindly review our community guidelines before leaving a comment.

Your email address will not be published. Required fields are marked *

Comment *

Name *

Email *

POST COMMENT

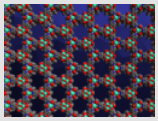
SHARE

RELATED CONTENT

ARTICLES

Angewandte Chemie 32/2023: Persistence

August 5, 2023



NEWS

Conductive Lanthanide-Based Metal–Organic Frameworks

August 1, 2023

EVENTS

Virtual Conference on Chemistry and its Applications (VCCA-2023)

August 7, 2023 to August 11, 2023

ARTICLES

Most Accessed Articles: June 2023

July 16, 2023

ChatGPT accelerates chemistry discovery for climate response, study shows

By Rachel Leven | August 7, 2023

UC Berkeley experts taught ChatGPT how to quickly create datasets on difficult-to-aggregate research about certain materials that can be used to fight climate change, according to a new paper published in the *Journal of the American Chemical Society.*

These datasets on the synergy of the highly-porous materials known as metal-organic frameworks (MOFs) will inform predictive models. The models will accelerate chemists' ability to create or optimize MOFs, including ones that alleviate water scarcity and capture air pollution. All chemists – not just coders – can build these databases due to the use of AI-fueled chatbots.

Zhiling Zheng and the MOF-powered water harvester, which helps address water stresses, pictured in Death Valley National Park. (Photo/ Zhiling Zheng)

“In a world where you have sparse data, now you can build large datasets,” said Omar Yaghi, the Berkeley chemistry professor who invented MOFs and an author of the study. “There are hundreds of thousands of MOFs that have been reported, but nobody has been able to mine that information. Now we can mine it, tabulate it and build large datasets.”

This breakthrough by experts at the College of Computing, Data Science, and Society’s Bakar Institute of Digital Materials for the Planet (BIDMaP) will lead to efficient and cost-effective MOFs more quickly, an urgent need as the planet warms. It can also be applied to other areas of chemistry. It is one example of how AI can augment and democratize scientific research.

“We show that ChatGPT can be a very helpful assistant,” said Zhiling Zheng, lead author of the study and a chemistry Ph.D. student at Berkeley. “Our ultimate goal is to make [research] much easier.”

Other authors of the study, “ChatGPT Chemistry Assistant for Text Mining and Prediction of MOF Synthesis,” include the Department of Chemistry ’s Oufan Zhang and the Department of Electrical Engineering and Computer Sciences ’s Christian Borgs and Jennifer Chayes. All are affiliated with BIDMaP, except Zhang.

Certain authors are also affiliated with the Kavli Energy Nanoscience Institute , the Department of Mathematics , the Department of Statistics , the School of Information and KACST-UC Berkeley Center of Excellence for Nanomaterials for Clean Energy Applications .

‘A substantial jump’ in AI for science

The team guided ChatGPT to quickly conduct a literature review. They curated 228 relevant papers. Then they enabled ChatGPT to process the relevant sections in those papers and to extract, clean and organize that data.

To help them teach ChatGPT to generate accurate and relevant information, they modified an approach called “prompt engineering” into “ChemPrompt Engineering.” They developed prompts that avoided asking ChatGPT for made up or misleading content; laid out detailed directions that explained to the chatbot the context and format for the response; and provided the large language model a template or instructions for extracting data.

The chatbot’s literature review – and the experts’ approach – was successful. ChatGPT finished in a fraction of an hour what would have taken a student years to complete, said Borgs, BIDMaP’s director. It mined the synthetic conditions of MOFs with 95% accuracy, Yaghi said.

> ″AI has transformed many other sectors of our society,″ said Omar Yaghi, BIDMaP’s co-director and chief scientist. ″Why not transform science?″

“One big area of how you do ‘AI for science’ is probing literature more effectively. This is really a substantial jump in doing natural language processing in chemistry,” said Chayes, dean of the College of Computing, Data Science, and Society. “And to use it, you can just be a chemist, not a computer scientist.”

This development will speed up MOF-related science work, including those efforts aimed at combating climate change, said Borgs. With natural disasters becoming more severe and frequent, we need that saved time, he said.

Yaghi noted that using AI in this way is still new. Like any new tool, experts will need time to identify its shortcomings and address them. But it’s worth investing the effort, he said.

“If we don't use it, then we can't make it better. If we can't make it better, then we will have missed a whole area that society is already using,” Yaghi said. “AI has transformed many other sectors of our society – commerce, banking, travel. Why not transform science?”

For more information

- *Journal of the American Chemical Society:* ChatGPT Chemistry Assistant for Text Mining and Prediction of MOF Synthesis
- *CDSS News:* New institute brings together chemistry and machine learning to tackle climate change
- *CDSS News:* Hand-held water harvester powered by sunlight could combat water scarcity

For Media Inquiries

CDSS-COMMS@BERKELEY.EDU

Berkeley College of Computing, Data Science, and Society

University of California, Berkeley
Berkeley, CA 94720-1786

Email: CDSSinfo@berkeley.edu

GIVE TO CDSS TODAY