Supplementary Information (SI) for Digital Discovery.
This journal is © The Royal Society of Chemistry 2025

Supporting information

for

Comparison of LLMs in Extracting Synthesis Conditions and Generating Q&A Datasets for Metal-Organic Frameworks

Yuang Shi[1,2,‡], Nakul Rampal[1,2,3,‡], Chengbin Zhao[1,2], Dongrong Joe Fu[3], Christian Borgs[3,4], Jennifer T. Chayes[3,4,5,6], Omar M. Yaghi[1,2,3,7,*]

[1] Department of Chemistry, University of California, Berkeley, California 94720, United States

[2] Kavli Energy Nanoscience Institute, University of California, Berkeley, California 94720, United States

[3] Bakar Institute of Digital Materials for the Planet, College of Computing, Data Science, and Society, University of California, Berkeley, California 94720, United States

[4] Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, California 94720, United States

[5] Department of Mathematics, University of California, Berkeley, California 94720, United States

[6] Department of Statistics, University of California, Berkeley, California 94720, United States

[7] KACST−UC Berkeley Center of Excellence for Nanomaterials for Clean Energy Applications, King Abdulaziz City for Science and Technology, Riyadh 11442, Saudi Arabia

[‡]indicates equal contribution

(a) **Synthesis conditions**

Prompt

```
You are a synthesis condition classification agent. You are required
to go through the given text and identify the synthesis conditions
for each and every material given in the paper (both MS and SI, if
available). There may be information about the synthesis condition
of more than one material. Please make sure to separate these
materials when generating the .json file. For each material try to
classify the conditions under different labels such as temperature,
solvents, the amount of each solvent (this is important), equipment,
chemicals used, time, washing method, drying method, yield, etc.
This is not at an exhaustive list of labels, please feel free to add
more labels as required. Some synthesis conditions may involve
multiple steps, please take that into account. Please do not include
any experimental characterization data such as those from Powder X-
Ray Diffraction (PXRD), Infrared (IR) spectra, , adsorption
isotherms, thermogravimetric analysis (TGA), nuclear magnetic
resonance (NMR) experiments etc (this is not an exhaustive list and
there may be other characterization techniques) including
information about its properties. I reiterate, please do not include
any experimental characterization data. Generate a machine readable
.json file containing the synthesis conditions for the following
text: {combined_text}.
```

(b) **Single-hop Q&A**

Prompt

```
You are a single hop Question and Answering (Q&A) dataset generation
agent. A single hop question and answer set is one that requires a
single step of reasoning. You are required to go through the given
text and identify the synthesis conditions and based on those
synthesis conditions develop a set of 20 Q&As. There may be
information about the synthesis conditions of more than one material
in the text. For example, you may come across a series of different
materials such as ZIF-1, ZIF-2, .... ZIF-12. Please try to diversify
the types of questions that you include. Please also try to include
a question for each material you come across in the paper. Please
feel free to include labels that are also used in some of the most
widely used Q&A datasets e.g., the question, the answer, the
difficulty level, and the type of question. the different types of
questions are factual, reasoning (single step reasoning), and True
or False. Please generate 6 ‘factual’ type questions, 7 ‘reasoning’
type questions, and 7 True or False type questions. Generate a
single hop .json file for the following text. Please include
questions of different types including factual (6 questions),
single-step reasoning (7 questions), and True or False (7
questions): {combined_text}.
```

(c) **Multi-hop Q&A**

Prompt

```
You are a multi-hop Question and Answering (Q&A) dataset generation
agent. A multi-hop Q&A is one that requires multi step reasoning to
come to an answer (this information can come from any part of the
paper, both MS and SI). To give you more details: A multi-hop Q&A
will always involve going through multiple parts of the paper to
come to an answer. This may include different paragraphs, different
pages, and also different documents (i.e. the manuscript and the
supplementary information). You are required to go through the given
text and identify the synthesis conditions and based on those
synthesis conditions develop a set of multi-hop (questions that
require multiple steps of reasoning) 20 Q&As for each DOI. There may
be information about the synthesis conditions of more than one
material in the text. For example, you may come across a series of
different materials such as ZIF-1, ZIF-2, .... ZIF-12. Please
diversify the type of questions to encompass different ideas and
materials. Please feel free to include labels that are also used in
some of the most widely used Q&A dataset for e.g., the question, the
answer, the difficulty level, and the type of question. the
different types of questions are factual, reasoning (single step
reasoning), and True or False. Please generate 6 'factual' type
questions, 7 'reasoning' type questions, and 7 True or False type
questions. For factual questions, please try to be creative with the
questions as it should require information from different parts of
the text to answer: {combined_text}.
```

Figure S1. Prompts for (a) synthesis conditions extraction, (b) single-hop and (c) multi-hop Q&A generation tasks.

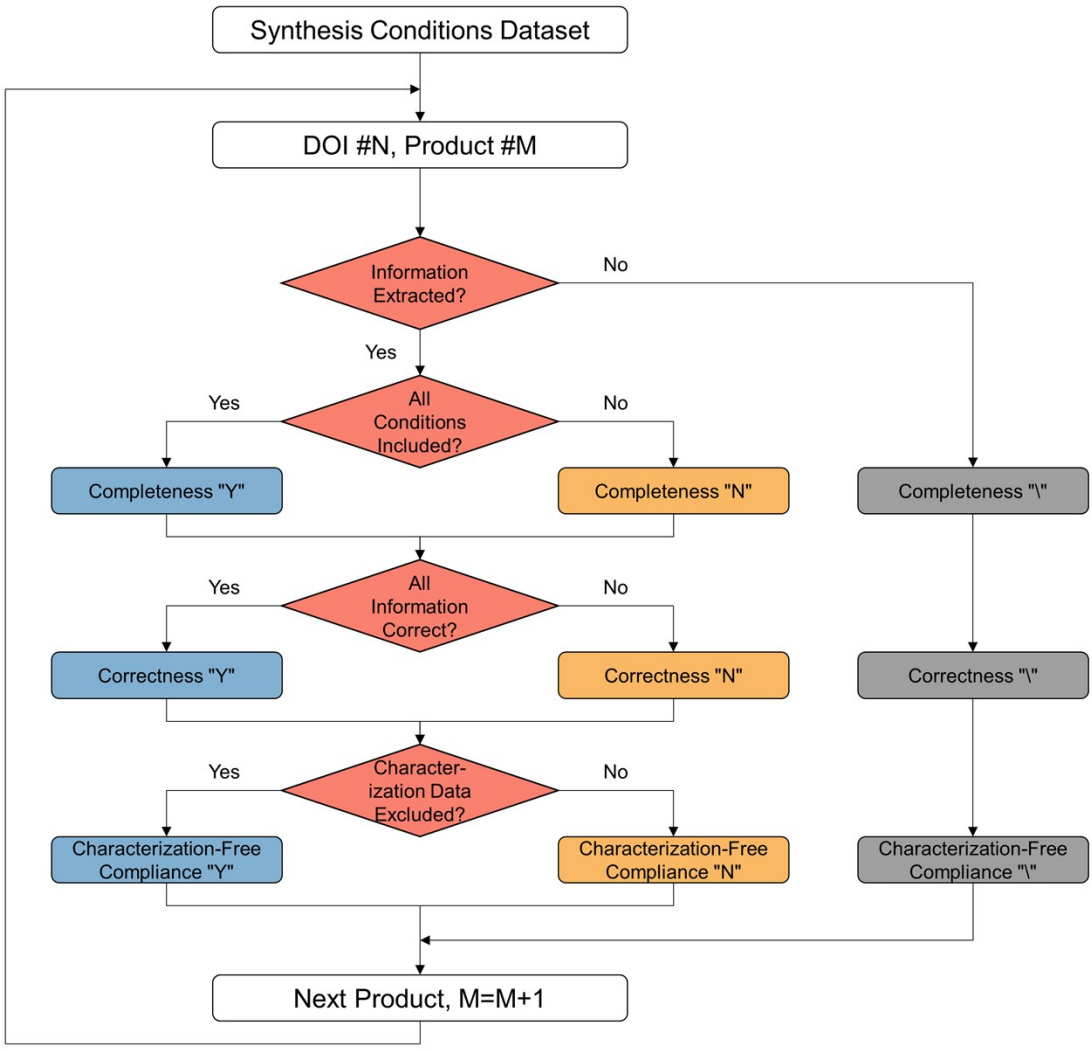

Figure S2. Evaluation flowchart for each product in the synthesis condition dataset. For a product from a DOI, we firstly check if the LLM has successfully extracted the corresponding information. If not, all three criteria will be marked as "\". Otherwise, we check whether the response has met three criteria and choose "Y" or "N" for each criterion. Then we move to the next product, or the next DOI when synthesis conditions of every product in the paper have been evaluated.

Output

1. **Question:** What is the role of the central aromatic core in the octacarboxylate linkers used in the synthesis of MFM materials?
    * **Answer:** The central aromatic core dictates the distance (l2) between the metal-organic cages (B and C) and thus influences the overall porosity of the resulting MFM material.
    * **Difficulty:** Hard
2. **Question:** What alternative activation method was employed for MFM-183a and MFM-185a compared to MFM-180a and MFM-181a?
    * **Answer:** MFM-183a and MFM-185a were activated using supercritical CO2 drying before undergoing heating under dynamic vacuum, while MFM-180a and MFM-181a were directly subjected to heating under vacuum.
    * **Difficulty:** Medium
3. **Question:** Which analytical technique revealed the pore size distribution for the activated MFM materials?
    * **Answer:** N_2 adsorption isotherms at 77 K, analyzed using a non-local density-functional theory (NLDFT) model, revealed the pore size distribution.
    * **Difficulty:** Medium
4. **Question:** What type of nodes are formed by the combination of the octacarboxylate linkers and copper ions in the MFM frameworks?
    * **Answer:** The combination forms square planar $[Cu_2(O_2CR)_4]$ nodes bridged by the octacarboxylate linkers.
    * **Difficulty:** Medium
5. **Question:** What specific structural feature allows MFM-185a to achieve a higher gravimetric CO_2 uptake at 20 bar compared to MFM-180a?
    * **Answer:** The larger pore volume of MFM-185a, due to its extended linker structure, enables a higher gravimetric CO_2 uptake at 20 bar compared to the less porous MFM-180a.
    * **Difficulty:** Hard
6. **Question:** What is the highest volumetric CO_2 uptake achieved among the studied MFM materials at 20 bar and 298 K, and which MOF exhibits it?
    * **Answer:** The highest volumetric CO_2 uptake is 292.4 vol/vol, achieved by MFM-181a due to its large surface area and moderately low density.
    * **Difficulty:** Medium

Factual

A question for which the answer is a stated fact.

Figure S3. Formatted Gemini response of multi-hop Q&A generation - six factual questions.

Output

1. **Question:** Why does increasing the length of the linker (l2) not always lead to an increase in porosity in octacarboxylate MOFs?
    * **Answer:** Simply increasing the length in all directions can lead to flexibility issues, highly strained frameworks, and/or interpenetration, which can reduce porosity.
    * **Difficulty:** Hard
2. **Question:** Why was the synthesis of MFM-185 conducted with different solvents and copper salt compared to the other MFM materials?
    * **Answer:** H8L5, the linker used for MFM-185, showed poor solubility in the N,N'-diethylformamide/ethanol/aqueous HCl mixture used for other MOFs. Therefore, a different solvent system (N,N'-dimethylformamide/DMSO/aqueous HCl) and copper salt ($Cu(NO_3)_2 \cdot 2.5H_2O$) were employed.
    * **Difficulty:** Medium
3. **Question:** Based on the reported BET surface areas and pore volumes, how does the structural design strategy of selective cage elongation impact the porosity of the MFM series?
    * **Answer:** The increasing BET surface areas and pore volumes from MFM-180a to MFM-185a demonstrate that selectively extending the metal-organic cages along one direction effectively increases the porosity of the MFM materials.
    * **Difficulty:** Medium
4. **Question:** How does the high connectivity of the MFM frameworks contribute to their CH_4 storage properties?
    * **Answer:** The high connectivity leads to relatively high crystal density. This minimizes the trade-off between gravimetric and volumetric capacity, allowing for simultaneously high values of both.
    * **Difficulty:** Hard
5. **Question:** Considering the CO_2 adsorption behavior at 1 bar, what is the significance of a high density of open metal sites in these MOFs?
    * **Answer:** The higher CO_2 uptake at 1 bar for the less porous MOFs with a higher density of open metal sites suggests that these sites are crucial for strong host-guest interactions with CO_2 at low pressure.
    * **Difficulty:** Medium
6. **Question:** Explain the rationale behind choosing to deuterate the aromatic rings of the linkers for 2H NMR studies.
    * **Answer:** Deuterating the aromatic rings provides a sensitive probe to investigate the molecular dynamics of these rings within the MOF structure. The 2H NMR spectra provide information about their rotational freedom and dynamic behavior.
    * **Difficulty:** Hard
7. **Question:** How does the change in the central core of the linker from MFM-180a-d16 to MFM-181a-d16 affect the rotational dynamics of the phenyl rings as observed by 2H NMR?
    * **Answer:** The change in the central core primarily affects the flipping mode (k1) of the phenyl rings. MFM-181a-d16 exhibits a faster flipping motion compared to MFM-180a-d16, indicated by the lower transition temperature (T1) for MFM-181a-d16.
    * **Difficulty:** Hard

Reasoning

A question that requires an 'understanding' of the text to answer.

Figure S4. Formatted Gemini response of multi-hop Q&A generation – seven reasoning questions.

Output

1. **Question:** True or False: MFM-180a exhibits a higher BET surface area than MFM-185a.
    * **Answer:** False
    * **Difficulty:** Easy
2. **Question:** True or False: The isosteric heat of CO_2 adsorption is significantly different across the MFM series.
    * **Answer:** False
    * **Difficulty:** Easy
3. **Question:** True or False: Elongating the metal-organic cages in the MFM series leads to a decrease in the volumetric CH_4 uptake at 65 bar.
    * **Answer:** False
    * **Difficulty:** Medium
4. **Question:** True or False: The 2H NMR studies revealed that the phenyl rings in both MFM-180a-d16 and MFM-181a-d16 exhibit free, unrestricted rotation at all temperatures.
    * **Answer:** False
    * **Difficulty:** Medium
5. **Question:** True or False: The MFM materials demonstrate a higher selectivity for CO_2 adsorption over CH4.
    * **Answer:** True
    * **Difficulty:** Easy
6. **Question:** True or False: The synthesis of all MFM materials was achieved under identical solvent and temperature conditions.
    * **Answer:** False
    * **Difficulty:** Easy
7. **Question:** True or False: The presence of open metal sites in the MFM series contributes to their CO_2 adsorption capabilities.
    * **Answer:** True
    * **Difficulty:** Easy

True/False

A question that for which the answer is either True or False.

Figure S5. Formatted Gemini response of multi-hop Q&A generation – seven true or false questions.

GPT-4 | **Output**


```
"question": "What is the main goal of the high-throughput protocol developed for the
synthesis of zeolitic imidazolate frameworks (ZIFs)?",
"answer": "The main goal of the high-throughput protocol developed for the synthesis of
ZIFs was to efficiently discover and synthesize new crystalline solid-state compounds,
specifically ZIFs, by systematically exploring a wide range of reaction conditions and
compositions.", "difficulty": "medium", "type": "factual"

"question": "How many different ZIF crystals were synthesized from the high-throughput
experiments?",
"answer": "Twenty-five different ZIF crystals were synthesized from the high-throughput
experiments.", "difficulty": "easy", "type": "factual"

"question": "What are the key differences between ZIFs and traditional zeolites?",
"answer": "The key differences between ZIFs and traditional zeolites are that in ZIFs,
transition metal atoms (such as Zn and Co) replace the tetrahedral linkers (such as Si, Al,
and P) found in zeolites, and imidazolates or imidazolate-type linkers replace the bridging
oxides in zeolites, resulting in structures with tetrahedral frameworks.", "difficulty":
"medium", "type": "factual"

"question": "What was the outcome of using a high-throughput method in the synthesis of
ZIFs?",
"answer": "The outcome of using a high-throughput method in the synthesis of ZIFs was the
discovery of 25 different ZIF structures, including 16 previously unreported compositions
and structures, and 5 with topologies not observed in zeolites before.", "difficulty":
"medium", "type": "factual"

"question": "What makes ZIF-69 and ZIF-70 particularly interesting for CO2 capture?",
"answer": "ZIF-69 and ZIF-70 are particularly interesting for CO2 capture due to their
exceptional stability, high porosity, and the presence of heterolinks in their structure,
which contribute to their extraordinary capacity for storing CO2 and their unusual
selectivity for CO2 capture from CO2/CO mixtures.", "difficulty": "hard", "type":
"reasoning"

"question": "How does the thermal stability of ZIF-68, ZIF-69, and ZIF-70 compare to other
materials?",
"answer": "ZIF-68, ZIF-69, and ZIF-70 exhibit high thermal stability, withstanding
temperatures up to 390°C, which is extraordinary compared to other materials and indicates
their potential for industrial applications.", "difficulty": "medium", "type": "reasoning"

"question": "What was the method used to confirm the permanent porosity of ZIF-68, ZIF-69,
and ZIF-70?",
"answer": "The permanent porosity of ZIF-68, ZIF-69, and ZIF-70 was confirmed through N2
adsorption measurements, which showed type I adsorption isotherm behavior typical of
materials with permanent porosity.", "difficulty": "medium", "type": "factual"

"question": "Are ZIFs chemically stable in boiling organic and aqueous media?",
"answer": "True", "difficulty": "easy", "type": "True or False"

"question": "Did the high-throughput synthesis method allow for the discovery of ZIFs with
heterolinks?",
"answer": "True", "difficulty": "easy", "type": "True or False"

"question": "Is the high-throughput synthesis method scalable to multigram quantities for
ZIF production?",
"answer": "True", "difficulty": "easy", "type": "True or False"
```


DOI: 10.1126/science.1152516

Figure S6. Unformatted GPT-4 response of single-hop Q&A generation. In this response, GPT-4 only generated 5 factual, 2 reasoning and 3 T/F questions, while our prompt asked for 6, 7, 7 respectively. Different question types were also mixed, which we did not want.

Table S1. Number of DOIs included in the datasets.

Publisher	Synthesis Conditions	Single-Hop Q&As	Multi-Hop Q&As
American Association for the Advancement of Science	1	5	3
American Chemical Society			7
American Institute of Physics		1	1
American Physical Society		1	1
Chinese Chemical Society	1	2	1
Elsevier		8	6
National Academy of Sciences	1	2	1
Springer Nature	18	10	10
Royal Society of Chemistry	21	10	10
Wiley	8	11	10

Table S2. Cost analysis for the LLMs

	LLM		
	GPT-4 Turbo	**Claude 3 Opus**	**Gemini 1.5 Pro**
Pricing per 1M tokens	Input **$10.00** Output **$30.00**	Input **$15.00** Output **$75.00**	Input **$1.25** (prompts <= 128k) or **$2.50** (prompts > 128k) Output **$5** (prompts <= 128k) or **$10** (prompts > 128k)
Total tokens	337.6M[1]	15.3M	17.1M
Total costs	$3600[1]	$257.14	$85.40
Number of requests	~7590[1]	328	469
Cost per DOI per task	**$0.47**	**$0.78**	**$0.18**

References

(1) Rampal, N.; Wang, K.; Burigana, M.; Hou, L.; Al-Johani, J.; Sackmann, A.; Murayshid, H. S.; AlSumari, W. A.; AlAbdulkarim, A. M.; Alhazmi, N. E.; Alawad, M. O.; Borgs, C.; Chayes, J. T.; Yaghi, O. M. Single and Multi-Hop Question-Answering Datasets for Reticular Chemistry with GPT-4-Turbo. *J. Chem. Theory Comput.* 2024, 20, 9128−9137.