

Finding Environmental-friendly Chemical Synthesis with AI and High-throughput Robotics

Anonymous Submission

Abstract—Recent challenges with the environment have resulted in tremendous interest in Green Chemistry, which includes the design of chemical products and processes that reduce the use of environmentally-harmful substances. Until now, finding new environmental chemical synthesis is largely a trial-and-error process, requiring trained expertise and a large amount of work. Here, we report a high-throughput process, combining AI techniques and robotic synthesis, allowing us to find a more environmentally friendly way to synthesize an existing material. The model materials in this study are to replace nitrate salts (NO_3^-), which might be responsible for algae bloom if leaked into open water, by a chloride salt (Cl^-), a natural abundant ion, in the synthesis of a metal-organic framework, Zn-HKUST-1. Our high-throughput process starts with using large language models (LLM)-based literature summary to create a database on the synthesis of Zn-HKUST-1 with NO_3^- , so that optimized concentrations of Cl^- can be suggested. Subsequently, these suggestions are tested with automatic robotic processes, increasing the speed and precision of the experiment, to find the optimal synthesis condition. Using this process, we successfully obtained MOFs (Metal Organic Frameworks) crystals from ZnCl_2 precursors. An AI-based learning process is developed, comparing the structural information of our synthesized materials with that of the database to confirm indeed the obtained MOFs are Zn-HKUST-1. This success proves that our process holds the promise to accelerate the discovery of new environmental-friendly materials in the near future.

Keywords—LLM literature synthesis, Metal Organic Frameworks, LLM experimental verification, Green Chemistry, Robotic synthesis

I. INTRODUCTION

Metal-organic frameworks (MOFs) are a class of material composed of metal clusters coordinated to organic ligand.[1], [2], [3] These materials combine the best of both organic and inorganic chemistry, offering remarkable properties and diverse applications. MOFs can be highly porous, with uniform pore structures and tunable porosity. The multitude of structures of MOFs also allow for flexibility in network topology. High porosity of MOFs enables several interesting applications in storage, supercapacitors, and sensing. Their hollow structure also gives rise to an extraordinarily large internal surface area, reaching even 780 square meters per gram and used for catalyst purposes.[4] In particular, MOFs are promising materials to mitigate climate change for their potential of CO_2 capturing.[5]

The synthesis of MOFs is relatively simple, consisting of mixing metal ions to organic linker and activating using thermal treatment. In this process, several non-

environmentally friendly solvents and compounds can be used, which eliminate the overall environmental impact of the MOFs applications. For instance, to synthesize the MOF named Zn-HKUST-1, it is common to use $\text{Zn}(\text{NO}_3)_2$, instead of a more environmental friendly ZnCl_2 . [6], [7], [8] The main reason is due to the high solubility of ZnCl_2 compared to that of $\text{Zn}(\text{NO}_3)_2$, which leads to faster but less controlled crystal growth during the synthesis conditions thus difficult to optimize the synthesis conditions.[8] Nevertheless, considering the minimal environmental impact of Cl^- as well as its several benefits such as their abundance and low cost, synthesizing Zn-HKUST-1 from ZnCl_2 as a green chemistry process would be much preferred if optimized condition is found. The transition from an established and well-study system to a new and more environmentally friendly one would require a large amount of work, including long training processes and extending hours using conventional trial/error approaches. Recent advances in Artificial Intelligence (AI) techniques and automation enable a possibility to minimize the resource and time of improving environmental impact on the environment.[9], [10], [11], [11], [12], [13] Specially, it will allow to the optimize conditions in short times and allows to find the right conditions rapidly.[14], [15]

In this work, we propose a new AI-based process for materials discovery integrating high-throughput literature synthesis with Large Language Model, high-throughput materials synthesis with automation robotics. We show that using this process, we can synthesize a target material Zn-HKUST-1, using ZnCl_2 instead of a less environmental friend but more common compound $\text{Zn}(\text{NO}_3)_2$. We first use a large language model to extract relevant information of HKUST-1 from $\text{Zn}(\text{NO}_3)_2$, and identify a few experimentally accessible conditions for Cl^- ions. Then we use a high throughput pipetting robot in testing those conditions of $\text{Zn}(\text{NO}_3)_2$ on ZnCl_2 .

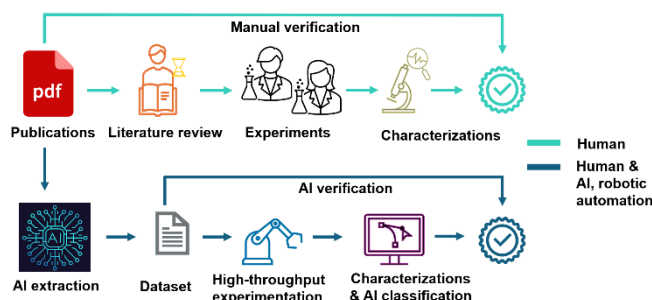


Figure 1: Schematic showing the process for a human, and our new approach with AI in all three steps extraction, characterization, and verification steps.

We focus on the concentration of ZnCl_2 in dimethylformamide (DMF) as the varying parameters, keeping the linkers, temperatures, and annealing time the same. Robotic automation allows to range of concentrations to be used in the synthesis of Zn-HKUST-1 from ZnCl_2 . The result is scanned first with characterization with AI classification to identify successful conditions, which then will be confirmed with advanced and conventional characterization techniques confirmed with data extracted from AI.

II. GENERAL FRAMEWORK

Figure 1 illustrates the general pipeline of this work. Conventionally, to replace a chemical in an establish synthesis process with a “greener” one, researchers would embark on a comprehensive process involving literature analysis, understanding fundamental principles, and gathering data from literature reviews. Those data typically contain crucial information such as chemical substances, reaction conditions (such as reaction time or temperature if required) and expected outcomes. Subsequently, experiments are conducted to synthesize materials, followed by the identification of necessary material characterization methods. Once material characterization results are obtained, they are cross-referenced with existing publications to validate the synthesis process. The whole process could take weeks or months of time to obtain an optimized synthesis process.

Here we propose a more efficient approach for accelerating material synthesis based on robotic automation and AI capabilities. Initially, we employ a Large Language Models platform to obtain a dataset from published literature. This platform facilitates the extraction and aggregation of pertinent information, which is then exported into a structured format such as a CSV file. This dataset provides valuable insights into experimental conditions pertaining to chemical substances and reaction parameters, enabling the selection of optimized conditions for subsequent experiments. Upon identifying the optimal experimental conditions, we utilize a pipetting robot for solution handler to simultaneously generate various concentrations of reaction mixtures, thereby accelerating the synthesis of materials. Different characterization techniques such optical microscopy, X-ray diffraction (XRD), Fourier-transformed Infrared spectroscopy (FTIR) are used to examine crystals under different concentrations. Once after obtaining crystals, to streamline the analysis process, we developed an AI-based classification method to verify that the obtained crystals are the expected one, for example by comparing the XRD curve of the obtained literature with the database.

III. RESULTS AND DISCUSSION

A. Materials

Zinc Chloride powder (>94% purity) was purchased from Viet Quang Chemicals Co., Ltd that plays a role as metal ion. Benzene-1,3,5-tricarboxylic acid (BTC) as an organic linker and Dimethylformamide (DMF, organic solvent) were supplied by Shanghai Haohong Biomedical Technology Co., Ltd and Shanghai Zhanyun Chemical Co., Ltd, respectively.

B. AI-literature synthesis

Manual extraction of synthesis conditions from extensive literature poses a bottleneck in material discovery, demanding significant time from experienced chemists. Previously, specialized natural language processing (NLP) models have

been adopted to enable efficient mining, which, however, are “**labor-intensive and necessitate expertise in coding, computer science, and data science**”. Recently, the emergence of large language models like GPT-3, GPT-3.5, and GPT-4 has the potential to revolutionize chemistry research in the coming decade. [15], [16], [17]

While there have been endeavors in utilizing the capability of these LLMs to aid researchers in text mining and data analysis, these approaches suffer from the high-cost usage of OpenAI’s API. This is due to the adoption of few-shot prompting for classification of synthesis paragraphs, which inherently involves a large number of input tokens to the API, thus further increasing the cost.

Scientific knowledge is predominantly stored in books and scientific journals, often in the form of PDFs, which present a challenge due to their diverse layouts. Traditional PDF parsers that rely solely on text extraction may struggle with this variability, making the combination of OCR (Optical Character Recognition) and Document Layout models essential for optimal extraction of information from such documents. **Figure 2A** shows a schematic representation of a data-driven workflow utilizing capabilities of current LLM(s) and Machine Learning models for the extraction and characterization of scientific data from chemistry literature. Papers (in PDF format) are converted into JPG/PNG images and transformed into structured representation via a Document Layout analysis model. The texts from the files are indexed for searching by ColBERT – a fast and accurate retrieval model, enabling scalable BERT-based search over large text collections in tens of milliseconds. Sections of the files which might contain synthesis conditions are then retrieved and fed into ChatGPT, which acts as a versatile extraction tool.

To achieve this, it should be recognized that creating prompts with expertise is a major factor in how effective LLM(s) is when used, with important challenge lies in writing clear instructions that can help LLMs. [16] To produce precise and accurate outcomes in a desired format, rapid design precision is crucial, highlighting the need of well-articulated, precise, and complete instructions. Furthermore, tiny and unique prompt signals can affect LLMs without altering the underlying meaning; as a result, occasionally similarly looking prompts have quite diverse qualities. When compared to human-written suggestions, the newly developed tool DSPy (Declarative Self-Improving Python) [18] offers a major leap in this regard. By integrating methods for prompting and optimizing language models, DSPy simplifies the prompt building process. It is necessary to have a small development set and some data annotation for DSPy to do the optimization automatically in order to apply DSPy for rapid optimization. From a conceptual standpoint, this novel system might allow researchers to start their investigations with a simple scaffolding of the high-level design steps and utilize DSPy to automate the training of LLMs. This could be viewed as a way to further strengthen our workflow and enhance the effectiveness of LLMs. The result of this step is shown in **Figure 2B**, with the example of a table containing the synthesis condition of HKUST-1.

We manually verify the table obtained with LLMs for data processing. To facilitate the verification synthesis, we

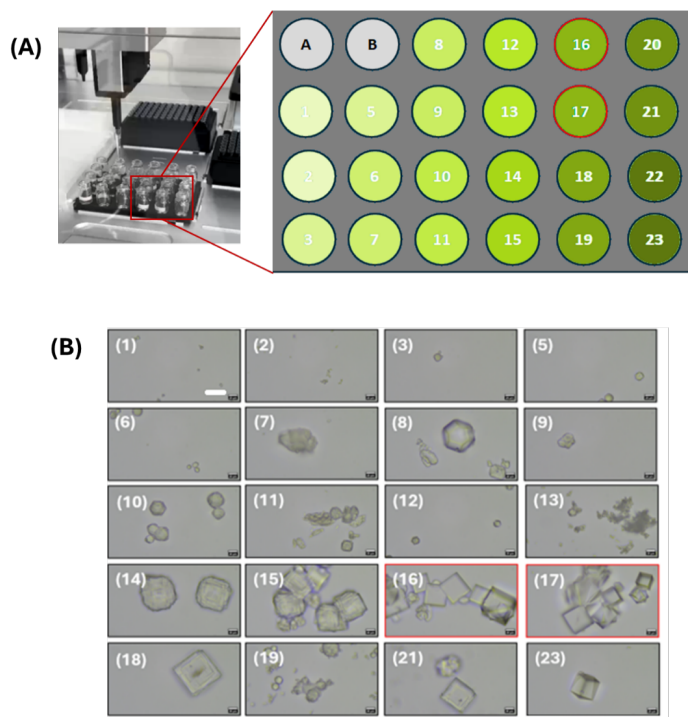


Figure 3: The process of automation robotic synthesis. (A) real-life picture of the hand manipulating pipette tips with custom-made tray containing 24 vials, comprising two precursors and twenty-two concentrations, colored from light green to dark green, corresponding to an increase in concentration; vials number 16 and number 17 contains the suggested concentration (B) the obtained optical microscopy images of the synthesis results, highlighting crystals in vials number 16 and 17.

between metal ions and organic linkers after dispensing by pipetting robot, the detail positions are illustrated in **Figure 3A**. When the dispensing progress finished, the solvothermal method was carried out for reaction happening.[20]

After obtaining 22 vials with different concentrations of two components, for the post-processing, those vials were all baked in an oven at 85°C in 24 hours. When annealing time finished, the crystal structure shapes were characterized with optical microscopy for each distinct concentration. Examples

of the obtained images are shown in **Figure 3B**, in which the same scale-bar of 10 μm is applied to all images.

It can be seen that different concentrations gave the different results of the experiment, from success to non-success, and among the successful experiments that resulted in crystal formation, the size and shape of crystals varied. The optimal concentration showing the full-formed cubic shape were found in vial number 16 and 17, corresponding to 60 mM of ZnCl_2 , while other concentrations either did not form crystals or resulted in defect non-cubical crystals. For instance, at nearly right concentration at around 57.6 mM (vial 14) to 64 mM (vial 23) formed plenty of square sheets and overlapped to shape the cubic (**Figure 3B**). With vial 16 and 17, the crystals showed cubic shape, which is expected for Zn-HKUST-1. To confirm the obtained result, the material in vial 16 and 17 are dried out of solution to become power samples for further characterizations.

D. Structural characterization and verification of MOFs

To confirm that the obtained material is indeed Zn-HKUST-1, we conducted a series of morphological, structural, and chemical characterizations. **Figure 4A** shows a scanning electron microscopy (SEM) image of the obtained structure, clearly indicating a simple cubic structure of HKUST-1. Furthermore, the morphology of the sample is homogeneous morphology, proving that the optimized process could yield good crystals with Cl^- as the counter ion.

To further investigate the crystal structure of the obtained material, powder X-ray diffraction (XRD), with the 2θ -intensity plot shown in **Figure 4B**. Five highest peaks were identified: at around 6.5° , 13.30° , 11.45° , 17.18° and 18.78° , respectively with the highest 2θ peak being at 11.45° . Using ChatGPT-4, we compared the similarity of our PXRD results with ones conducted on Zn-HKUST-1 and found a strong similarity.

Furthermore, we could identify the crystal plan associated with each peak. Using Bragg's equations, $n\lambda = 2d \sin \theta$ and $d_{hkl} = a/(\sqrt{h^2 + k^2 + l^2})$, the lattice constant of the structure is obtained $a = 26.79 \pm 0.19 \text{ \AA}$ (**Table 2**), close to the value found in the literature of reported in the literature.[7][8]

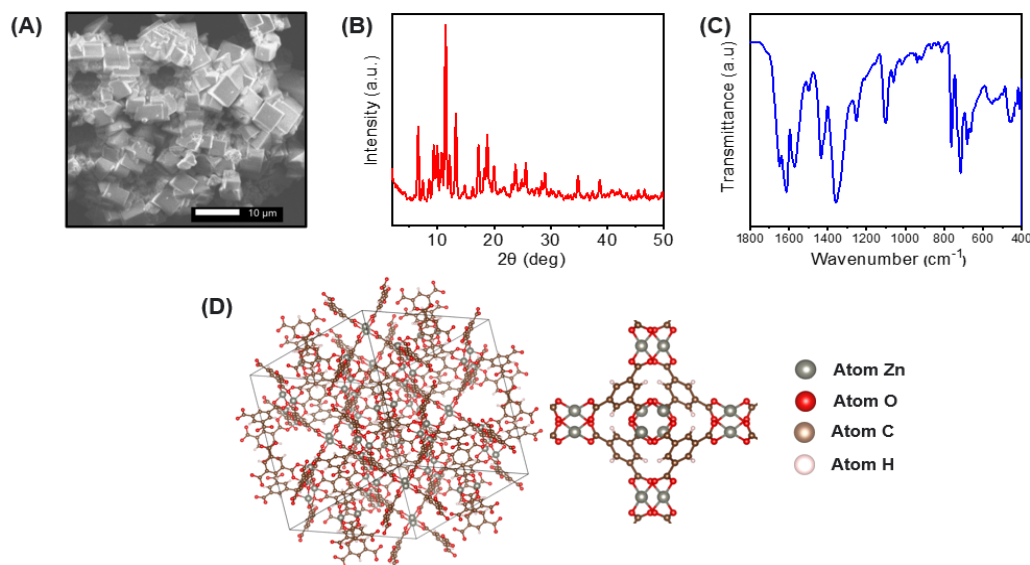


Figure 4: Confirmation of the obtained MOFs as Zn-HKUST-1. (A)(B)(C) Characterization of the obtained MOFs with surface topography (using SEM), structural information (using XRD), and chemical information (using FT-IR), respectively. (D) the chemical structure of the obtained MOFs as Zn-HKUST-1.

Table 2. XRD peak analysis of obtained Zn-HKUST-1

Peak 2 θ (°)	Spacing (Å)	hkl plane	Lattice constant (Å)
6.5	13.58	(200)	27.16
11.45	7.72	(222)	26.74
13.30	6.65	(400)	26.60
17.18	5.15	(511)	26.76
18.78	4.72	(440)	26.70

Finally, Fourier-transform Infrared spectroscopy (FTIR) is conducted to confirm Zn-HKUST-1 chemical compositions. The FTIR spectra of as-synthesized Zn-HKUST-1 is shown in **Figure 4C** with the peaks below 1200 cm^{-1} relate to the vibrations of the BTC linker. The range from 1250 cm^{-1} to 1650 cm^{-1} indicates the carboxylate linker's presence; the spectra peak observed at around 1610 cm^{-1} is attributed to the asymmetric stretching vibrations of the carboxylate group in BTC linker. Conversely, at 1420 cm^{-1} and 1370 cm^{-1} appeared two peaks that are related to the symmetric stretching vibrations of the carboxylate group. The influence of Zn^{2+} ion can be seen with the peak at 1630 cm^{-1} , usually attributed to the C=O stretching bond in BTC.[21] With the morphological, crystal, and chemical confirmation, we confirmed that our process successfully created Zn-HKUST-1 from the environmentally friendly salt Cl^- . The structure of the obtained Zn-HKUST-1 with a lattice constant of 26.7 Å is reconstructed in 3D, shown in **Figure 4D**.

IV. CONCLUSION AND PERSPECTIVES

In this report, we propose a new process to accelerate materials discovery for Green Chemistry. The process combines AI techniques for literature synthesis, using Large Language Models, and Robotic Liquid Handling Automation for materials synthesis. The process is successfully applied in the synthesis of a Metal Organic Framework named Zn-HKUST-1 from an environmental friendly precursor ZnCl_2 , instead of a common precursor $\text{Zn}(\text{NO}_3)_2$. First, a database about HKUST-1 MOFs is created from 22 papers, in which synthesis conditions are recorded and analyzed to identify the most promising synthesis condition for Zn-HKUST-1. Second, from this one suggested condition, experimental space is enlarged into 22 experiment conditions, which are simultaneously conducted with a robotic handler to fine tune the synthesis of the MOF. The robot procedure is modified to work with custom-made 3D printed well-plates to fit the needs of the experiments. Finally, an optimized condition is obtained and the resulting material are thoroughly characterized in its morphology, structure, and chemical composition, to be confirmed as Zn-HKUST-1.

Our result shows that AI techniques can be integrated seamlessly into Green Chemistry, in which environmentally friendly materials are used as alternatives for potentially harmful chemicals in existing synthesis schemes. In this work, we present the use of Large Language Models for materials synthesis and for automatic result verification. In the future work, we will integrate few-step suggestions into experiment design and image classification in material characterization. Coupling with high-throughput robotic solution handling, we believe this process would pave the way to the increasing use of environmentally friendly materials in the making and synthesis of materials in the lab and in industry.

CONFLICT OF INTERESTS

The authors declare no conflict of interests.

REFERENCES

- [1] D. Li, A. Yadav, H. Zhou, K. Roy, P. Thanasekaran, and C. Lee, "Advances and Applications of Metal-Organic Frameworks (MOFs) in Emerging Technologies: A Comprehensive Review," *Glob. Chall.*, vol. 8, no. 2, p. 2300244, 2024, doi: 10.1002/gch2.202300244.
- [2] M. S. Khan, Y. Li, D.-S. Li, J. Qiu, X. Xu, and H. Y. Yang, "A review of metal-organic framework (MOF) materials as an effective photocatalyst for degradation of organic pollutants," *Nanoscale Adv.*, vol. 5, no. 23, pp. 6318–6348, 2023, doi: 10.1039/D3NA00627A.
- [3] V. F. Yusuf, N. I. Malek, and S. K. Kailasa, "Review on Metal-Organic Framework Classification, Synthetic Approaches, and Influencing Factors: Applications in Energy, Drug Delivery, and Wastewater Treatment," *ACS Omega*, vol. 7, no. 49, pp. 44507–44531, Dec. 2022, doi: 10.1021/acsomega.2c05310.
- [4] Y. Lin, Y. Li, Y. Cao, and X. Wang, "Two-dimensional MOFs: Design & Synthesis and Applications", doi: 10.1002/asia.202100884.
- [5] Z. Zhang, Y. Zhao, Q. Gong, Z. Li, and J. Li, "MOFs for CO₂ capture and separation from flue gas mixtures: the effect of multifunctional sites on their adsorption capacity and selectivity," *Chem. Commun.*, vol. 49, no. 7, pp. 653–661, Dec. 2012, doi: 10.1039/C2CC35561B.
- [6] J. I. Feldblyum, M. Liu, D. W. Gidley, and A. J. Matzger, "Reconciling the Discrepancies between Crystallographic Porosity and Guest Access As Exemplified by Zn-HKUST-1," *J. Am. Chem. Soc.*, vol. 133, no. 45, pp. 18257–18263, Nov. 2011, doi: 10.1021/ja2055935.
- [7] A. A. Tehrani, V. Safarifard, A. Morsali, G. Bruno, and H. A. Rudbari, "Ultrasound-assisted synthesis of metal-organic framework nanorods of Zn-HKUST-1 and their templating effects for facile fabrication of zinc oxide nanorods via solid-state transformation," *Inorg. Chem. Commun.*, vol. 59, pp. 41–45, Sep. 2015, doi: 10.1016/j.inoche.2015.06.028.
- [8] M. J. Taghizadeh, S. S. S. Afghahi, and H. Ghasempoor, "Mechanical Synthesis of Zn-HKUST-1 Metal-Organic Framework and Investigation its Fluorescence Sensing Ability Toward to Detect Explosive-like Nitro aromatic Analytes Such as TNT," 2017.

- [9] J. M. Gregoire, L. Zhou, and J. A. Haber, "Combinatorial synthesis for AI-driven materials discovery," *Nat. Synth.*, vol. 2, no. 6, pp. 493–504, Jun. 2023, doi: 10.1038/s44160-023-00251-4.
- [10] M. M. Flores-Leonar *et al.*, "Materials Acceleration Platforms: On the way to autonomous experimentation," *Curr. Opin. Green Sustain. Chem.*, vol. 25, p. 100370, Oct. 2020, doi: 10.1016/j.cogsc.2020.100370.
- [11] B. P. MacLeod *et al.*, "Self-driving laboratory for accelerated discovery of thin-film materials," *Sci. Adv.*, vol. 6, no. 20, p. eaaz8867, May 2020, doi: 10.1126/sciadv.aaz8867.
- [12] M. L. Green *et al.*, "Fulfilling the promise of the materials genome initiative with high-throughput experimental methodologies," *Appl. Phys. Rev.*, vol. 4, no. 1, p. 011105, Mar. 2017, doi: 10.1063/1.4977487.
- [13] J. Yano *et al.*, "The case for data science in experimental chemistry: examples and recommendations," *Nat. Rev. Chem.*, vol. 6, no. 5, pp. 357–370, May 2022, doi: 10.1038/s41570-022-00382-w.
- [14] A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon, and E. D. Cubuk, "Scaling deep learning for materials discovery," *Nature*, vol. 624, no. 7990, pp. 80–85, Dec. 2023, doi: 10.1038/s41586-023-06735-9.
- [15] Z. Zheng, O. Zhang, C. Borgs, J. T. Chayes, and O. M. Yaghi, "ChatGPT Chemistry Assistant for Text Mining and the Prediction of MOF Synthesis," *J. Am. Chem. Soc.*, vol. 145, no. 32, pp. 18048–18062, Aug. 2023, doi: 10.1021/jacs.3c05819.
- [16] Z. Zheng, Z. Rong, N. Rampal, C. Borgs, J. T. Chayes, and O. M. Yaghi, "A GPT-4 Reticular Chemist for Guiding MOF Discovery**," *Angew. Chem. Int. Ed.*, vol. 62, no. 46, p. e202311983, 2023, doi: 10.1002/anie.202311983.
- [17] Y. Kang and J. Kim, "ChatMOF: An Autonomous AI System for Predicting and Generating Metal-Organic Frameworks".
- [18] "stanfordnlp/dspy." Stanford NLP, Apr. 28, 2024. Accessed: Apr. 28, 2024. [Online]. Available: <https://github.com/stanfordnlp/dspy>
- [19] "2,4-NONANEDIONE," *Org. Synth.*, vol. 47, p. 92, 1967, doi: 10.15227/orgsyn.047.0092.
- [20] R. . Ediaty, S. K. Dewi, M. R. Hasan, M. . Kahardina, I. K. Murwani, and M. . Nadjib, "MESOPOROUS HKUST-1 SYNTHESIZED USING SOLVOTHERMAL METHOD," *Rasayan J. Chem.*, vol. 12, no. 03, pp. 1653–1659, 2019, doi: 10.31788/RJC.2019.1231968.
- [21] Y.-K. Seo, G. Hundal, I. Jang, K. Y. Lee, C.-H. Jun, and J.-S. Chang, "Microwave synthesis of hybrid inorganic–organic materials including porous Cu₃(BTC)₂ from Cu(II)-trimesate mixture," *Microporous Mesoporous Mater.*, vol. 119, pp. 331–337, Mar. 2009, doi: 10.1016/j.micromeso.2008.10.035.