

Research Outcomes	Harnessing Large Language Models to Collect and Analyze Metal–Organic Framework Property Data Set
Performance Objectives	Published in a Top 10% JCR Journal
Type of Performance	<input checked="" type="checkbox"/> Research Article(Paper) <input type="checkbox"/> Patents <input type="checkbox"/> Researcher Exchange <input type="checkbox"/> Researcher Engagement <input type="checkbox"/> Information Exchange <input type="checkbox"/> Others
Description of Performance Type	Published in <b>Journal of the American Chemical Society</b> (JCR top 7.8%)
Research Institutes	Korea Advanced Institute of Science and Technology (KAIST) / Prof. Jihan Kim / Taeun Bae et al. (5 others)
Attachments (Image, Photograph, Ect.)	<p>a)</p> <p>&lt;Overview of the Data Mining Process Using Large Language Models&gt;</p>
Achievement Date	2025.01.21
Summary of Performance	<ul style="list-style-type: none"> <li>- Developed a tool that utilizes large language models (LLMs) to mine data from tables and text in metal-organic framework (MOF) literature.</li> <li>- Structured the extracted information into a database.</li> <li>- Published in <b>Journal of the American Chemical Society</b> (JCR top 7.8%).</li> </ul>
Description of Performance	<p>■ <b>Key Features</b></p> <ul style="list-style-type: none"> <li>- Data mining powered by large language models.</li> </ul> <p>■ <b>Performance</b></p> <ul style="list-style-type: none"> <li>- Achieves over 90% accuracy across all three stages of data extraction.</li> <li>- Capable of processing approximately 40,000 papers in a single batch.</li> </ul> <p>■ <b>Excellence of the Results</b></p> <ul style="list-style-type: none"> <li>- Outperforms existing data mining tools in accuracy.</li> <li>- Applicable to a wide range of papers with diverse formats.</li> </ul> <p>■ <b>Uniqueness of the Results</b></p> <ul style="list-style-type: none"> <li>- Unlike conventional methods that require retraining as the number of papers increases, this approach applies to new papers without additional training.</li> <li>- While initially developed for MOF literature, the method is easily adaptable to other scientific fields using large language models.</li> </ul>