Chemistry of Oxydic and Organic Interfaces Department
Institute of Functional Interfaces (IFG)

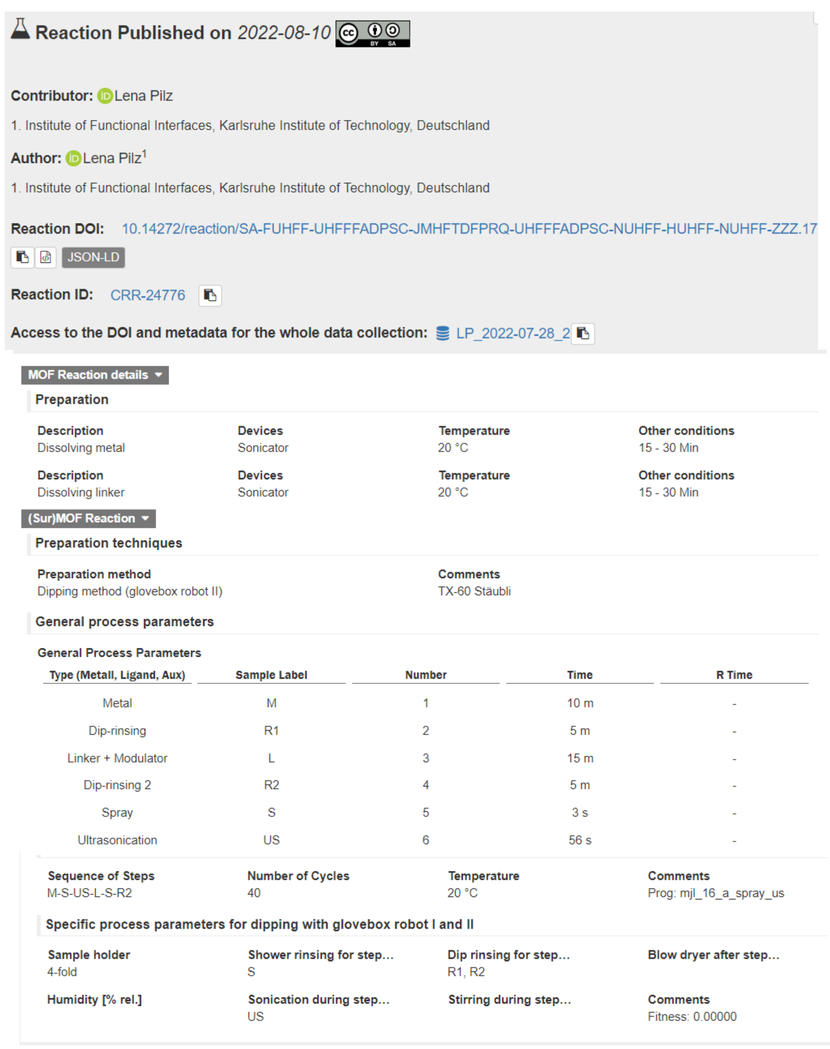
Karlsruhe Institute of Technology

Human–machine Interface to Interact with MOF Synthesis Data in Chemotion via a Large Language Model

Yi Luo[1], Pei-Chi Huang[2], Chia-Lin Lin[2], Lena Pilz[1], Manuel Tsotsalas[1], Nicloe Jung[2], Stefan Bräse[2], Christof Wöll [1*]

[1] Institute of Functional Interfaces (IFG), Karlsruhe Institute of Technology; [2] Institute of Organic Chemistry (IOC), Karlsruhe Institute of Technology;

Chemotion: an Electronic Laboratory Notebook (ELN) & Repository for Research Data



Reaction Published on *2022-08-10*

Contributor: Lena Pilz

1. Institute of Functional Interfaces, Karlsruhe Institute of Technology, Deutschland

Author: Lena Pilz[1]

1. Institute of Functional Interfaces, Karlsruhe Institute of Technology, Deutschland

Reaction DOI: 10.14272/reaction/SA-FUHFF-UHFFFADPSC-JMHFTDFPRQ-UHFFFADPSC-NUHFF-HUHFF-NUHFF-ZZZ.17

JSON-LD

Reaction ID: CRR-24776

Access to the DOI and metadata for the whole data collection: LP_2022-07-28_2

MOF Reaction details

Preparation

Description	Devices	Temperature	Other conditions
Dissolving metal	Sonicator	20 °C	15 - 30 Min
Dissolving linker	Sonicator	20 °C	15 - 30 Min

(Sur)MOF Reaction

Preparation techniques

Preparation method	Comments
Dipping method (glovebox robot II)	TX-60 Staubli

General process parameters

General Process Parameters

Type (Metall, Ligand, Aux)	Sample Label	Number	Time	R Time
Metal	M	1	10 m	-
Dip-rinsing	R1	2	5 m	-
Linker + Modulator	L	3	15 m	-
Dip-rinsing 2	R2	4	5 m	-
Spray	S	5	3 s	-
Ultrasonication	US	6	56 s	-

Sequence of Steps	Number of Cycles	Temperature	Comments
M-S-US-L-S-R2	40	20 °C	Prog: mjl_16_a_spray_us

Specific process parameters for dipping with glovebox robot I and II

Sample holder	Shower rinsing for step…	Dip rinsing for step…	Blow dryer after step…
4-fold	S	R1, R2	
Humidity [% rel.]	**Sonication during step…**	**Stirring during step…**	**Comments**
	US		Fitness: 0.00000

A SURMOF synthesis record in Chemotion Repository (https://dx.doi.org/10.14272/reaction/SA-FUHFF-UHFFFADPSC-JMHFTDFPRQ-UHFFFADPSC-NUHFF-HUHFF-NUHFF-ZZZ.17)

Chemotion is part of the ComPlat (Compound Platform) project, and it comprises two main components:

- Chemotion ELN (Electronic Laboratory Notebook): This tool is designed for academic researchers to integrate advanced Research Data Management techniques into their routine work..
- Chemotion Repository: This is a web-accessible data repository specifically tailored for Chemistry. It works in tandem with the ELN, allowing for direct, fast, secure transfer of research information for sharing and archiving purposes.

Large Language Model

Language Models (LMs):
Language models are algorithms that predict the probability distribution of language sequences. They are trained on large text corpora and can generate, understand, translate, or summarize text. Traditional LMs, like n-gram models, relied on fixed-size word sequences, but recent advancements have shifted towards neural network-based models.

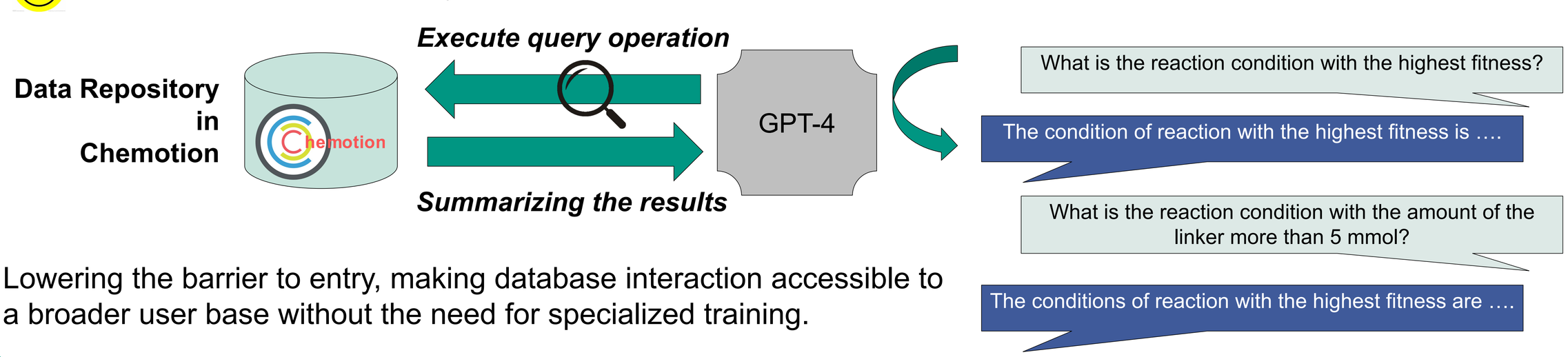
Generative Pre-trained Transformers (GPT):
GPT is a specific type of language model that utilizes the Transformer architecture. This architecture revolutionized NLP with its ability to process words in parallel and capture long-range dependencies in text.

Human–machine Interface to Interact with MOF Synthesis Data in Chemotion

Conversational Way to Interact a database

Studying specific technical knowledge of database schemas and query syntax → Employing structured query languages like SQL (Structured Query Language) to retrieve or manipulate data

Human–machine Interface Way to Interact a Database



Lowering the barrier to entry, making database interaction accessible to a broader user base without the need for specialized training.

In conclusion, our integration of GPT model with Chemotion represents a significant step forward in simplifying the analysis of MOF synthesis data. This approach not only streamlines data management but also makes complex information more accessible through intuitive, natural language queries. Currently, our focus is on MOF data, but we are ambitiously working towards expanding this technology to encompass all records of organic reactions within Chemotion. Our ultimate goal is to democratize data analysis in chemical research, making it more efficient and user-friendly for researchers of Chemotion.

1) Tremouilhac, Pierre, et al. *Journal of cheminformatics* 9.1 (2017)
2) Vaswani, Ashish, et al Advances in neural information processing systems 30 (2017).