

Supplementary Information

Fine-tuning Large Language Models for Chemical Text Mining

Wei Zhang,^{‡ab} Qinggong Wang,^{‡c} Xiangtai Kong,^{ab} Jiacheng Xiong,^{ab} Shengkun Ni,^{ab}
Duanhua Cao,^{ad} Buying Niu,^{ab} Mingan Chen,^{ae} Yameng Li,^g Runze Zhang,^{ab} Yitian
Wang,^{ab} Lehan Zhang,^{ab} Xutong Li,^{ab} Zhaoping Xiong,^g Qian Shi,^f Ziming Huang,^h
Zunyun Fu^{*a} and Mingyue Zheng^{*abc}

^aDrug Discovery and Design Center, State Key Laboratory of Drug Research, Shanghai Institute of
Materia Medica, Chinese Academy of Sciences, 555 Zuchongzhi Road, Shanghai 201203, China.

^bUniversity of Chinese Academy of Sciences, No. 19A Yuquan Road, Beijing 100049, China

^cNanjing University of Chinese Medicine, 138 Xianlin Road, Nanjing 210023, China

^dInnovation Institute for Artificial Intelligence in Medicine of Zhejiang University, College of
Pharmaceutical Sciences, Zhejiang University, Hangzhou, Zhejiang 310058, China

^eSchool of Physical Science and Technology, ShanghaiTech University, Shanghai 201210, China

^fLingang Laboratory, Shanghai 200031, China

^gProtonUnfold Technology Co., Ltd, Suzhou, China

^hMedizinische Klinik und Poliklinik I, Klinikum der Universität München, Ludwig-Maximilians-
Universität, Munich, Germany

[‡]Wei Zhang and Qinggong Wang contributed equally to this study.

*Correspondence should be addressed to:

Mingyue Zheng: myzheng@simm.ac.cn

Zunyun Fu: fuzunyun@simm.ac.cn

Contents

1. Dataset Preparation and Annotation	3
1.1. Overview of Datasets	3
1.2. Paragraph2MOFInfo Data Annotation.....	3
1.3. Paragraph2NMR Data Collection and Annotation.....	6
2. Examples of Prompt Engineering for Chemical Text Mining	8
2.1. Example for Paragraph2Compound Task	8
2.2. Example for Paragraph2RXNRole (Product Extraction) Task	9
2.3. Example for Paragraph2RXNRole (Reaction Role Labelling) Task	10
2.4. Example for Paragraph2MOFInfo Task.....	11
2.5. Example for Paragraph2NMR Task	12
2.6. Example for Paragraph2Action Task	13
2.7. Example for Zero-shot Prompts Designing on Paragraph2NMR Task.....	14
3. Processing Details	15
3.1. Processing for ChatGPT.....	15
3.2. Post-Processing for Open-source LLMs (Mistral, Llama3, Llama2).....	15
3.3. Pre-Processing for T5 and BART	15
3.4. Post-Processing for T5 and BART	15
4. Training Details	17
4.1. Hyperparameters Tuning.....	17
4.2. Hardware resources and memory cost.....	18
4.3. Multi-Task Learning of T5 and BART for Paragraph2MOFInfo	19
4.4. Multi-Task Learning of T5 and BART for Paragraph2NMR	20
5. Supplemental Results of the Performances	21
5.1. Performance of Paragraph2Compound	21
5.2. Performance of Paragraph2RXNRole	24
5.3. Performance of Paragraph2MOFInfo.....	25
5.4. Performance of Paragraph2NMR.....	28
5.5. Performance of Paragraph2Action	30
5.6. Performance Trends with Increasing Training Data	31
6. Analysis of Error Predictions by Fine-tuned ChatGPT	32
6.1. Error cases for Paragraph2Compound Task.....	32
6.2. Error cases for Paragraph2RXNRole Task	33
6.3. Error cases for Paragraph2MOFInfo Task	34
6.4. Error cases for Paragraph2NMR Task	35
6.5. Error cases for Paragraph2Action Task.....	36
7. Reference	37

1. Dataset Preparation and Annotation

1.1. Overview of Datasets

In this work, we fine-tuned ChatGPT on five complex tasks in chemical text mining: Paragraph2Compound, Paragraph2RXNRole, Paragraph2MOFInfo, Paragraph2NMR, and Paragraph2Action. (Table S1)

Table S1. Statistic of Dataset for the Five Chemical Text Mining Tasks.

Task	Data source and processing	Training set size	Evaluation set size
Paragraph2Compound	Based on USPTO’s data by Lowe ¹ , automatedly annotated by us.	10000, 1000, 100, 10	1000
Paragraph2RXNRole (Product Extraction)	Manually annotated by Guo et.al ² , post-processed by us.	6163	723
Paragraph2RXNRole (Reaction Role Labeling)	Manually annotated by Guo et.al ² post-processed by us.	599	111
Paragraph2MOFInfo	Based on raw data by Zheng et.al. ³ , re-annotated by us.	329	329
Paragraph2NMR	Manually collected and annotated by us.	300, 200, 100, 50, 25	300
Paragraph2Action	Manually annotated by Vaucher et.al. ⁴ .	14168, 1060	352

For Paragraph2Compound task, we compiled an automatically annotated dataset.

For Paragraph2RXNRole and Paragraph2Action, we utilized annotated data provided by others. For Paragraph2MOFInfo and Paragraph2NMR, we manually collected, annotated, and processed the data. We will illustrate the process of annotation for Paragraph2MOFInfo and Paragraph2NMR.

1.2. Paragraph2MOFInfo Data Annotation

For Paragraph2MOFInfo task, the raw data was collected from the work of Omar M. Yaghi et.al., and re-annotated by us (Fig. S1). The annotation process can be summarized as the following steps.

Step 1: we corrected symbols that are susceptible to mistakes or become garbled in the original data, especially symbols like "°C", "·", "µL", and "?" which are particularly prone to errors during the PDF parsing process.

Step 2: we annotated more complete information about substances' quantity. If both molecular mass and molar quantities are provided in the text, we provided exactly all these properties in the re-annotation.

Step 3: when a synthesis condition includes multiple components, we employed symbol "|" to separate them. The separated substances and quantities correspond separately.

Example 1:

Paragraph:

Synthesis of Zrbtba 1 In a 40 mL clear glass Wheaton sample vial with a rubber-lined cap, anhydrous ZrCl₄ (0.054 g, 0.23 mmol) and benzoic acid (2.400 g, 19.66 mmol) was dissolved in 12 mL of DMF using ultrasound bath for 30 min at 55 °C. The resulting solution was stored at room temperature for 15 minutes. To this clear solution btbaH4 (0.054 g, 0.070 mmol) was added and the resulting heterogeneous mixture was put in an ultrasound bath at room temperature for another 10 minutes. Finally, the vial was placed into an oven and heated with a ramp rate of 0.2 °C/minute to 120 °C for 96 hours followed by cooling to 25 °C at a rate of 0.5 °C / minute. Shiny cubic crystals of the compound 1 were found at the bottom and the walls of the vial at the end of the reaction. The product was isolated by filtration and cleaned with DMF (3 × 30 mL). Yield: 0.100 g (19 %, based on btbaH4).

Raw data:

compound name	metal source	metal amount	linker	linker amount	modulator	modulator amount or volume	solvent	solvent volume	reaction temperature	reaction time
1	ZrCl ₄	0.054 g	btbaH4	0.054 g	benzoic acid	2.400 g	DMF	12 mL	ramp rate of 0.2 °C/minute to 120 °C	96 hours

Re-annotated data:

compound name	metal source	metal amount	linker	linker amount	modulator	modulator amount or volume	solvent	solvent volume	reaction temperature	reaction time
Zrbtba 1	ZrCl ₄	0.054 g, 0.23 mmol	btbaH4	0.054 g, 0.070 mmol	benzoic acid	2.400 g, 19.66 mmol	DMF	12 mL	a ramp rate of 0.2 °C/minute to 120 °C	96 hours

Example 2:

Paragraph:

Preparation of NENU-500 and NENU-501. A mixture of Na₂MoO₄·2H₂O (618 mg, 2.55 mmol), Mo powder 99.99% (50 mg, 0.52 mmol), H₃PO₃ (20 mg, 0.25 mmol), zinc chloride (136 mg, 1.00 mmol), H₃BTB (130 mg, 0.30 mmol), tetrabutylammonium hydroxide 40 wt % solution in water (120 μL, 0.18 mmol), and H₂O (7 mL) was stirred for 20 min, and the pH was acidified to 4.8 with diluted HCl (2M). Then, the mixture was transferred and sealed in a 15 mL Teflonlined stainless steel container and heated at 180 °C for 72 h. After cooling to room temperature at 10 °C·h⁻¹, dark-red crystals (NENU500) suitable for XRD study were harvested (yield 68% based on H₃BTB). IR (Figure S3, KBr pellets, ν/cm⁻¹): 3450 (s), 2961 (m), 1599 (m), 1550 (w), 1466 (w), 1376 (m), 1110 (w), 935 (m), 814 (m), 780 (m), 705 (m), 587 (m). NENU-501 was isolated by an analogous method with NENU-500, only using H₃BPT in substitution for H₃BTB. IR (Figure S3, KBr pellets, ν/cm⁻¹): 3445 (w), 2961 (m), 2872 (m), 1559 (m), 1470 (m), 1350 (s), 1148 (w), 942 (s), 819 (s), 777 (s), 707 (m), 590 (m), 486 (w). S2.

Raw data:

compound name	metal source	metal amount	linker	linker amount	modulator	modulator amount or volume	solvent	solvent volume	reaction temperature	reaction time
NENU-500	Na ₂ MoO ₄ ·2H ₂ O, Mo powder 99.99%	618 mg, 50 mg	H ₃ BTB	130 mg	tetrabutylammonium hydroxide 40 wt % solution in water	120 μL	H ₂ O	7 mL	180 °C	72 h
NENU-501	Na ₂ MoO ₄ ·2H ₂ O, Mo powder 99.99%	618 mg, 50 mg	H ₃ BPT	N/A	tetrabutylammonium hydroxide 40 wt % solution in water	120 μL	H ₂ O	7 mL	180 °C	72 h

Re-annotated data:

compound name	metal source	metal amount	linker	linker amount	modulator	modulator amount or volume	solvent	solvent volume	reaction temperature	reaction time
NENU-500	Na ₂ MoO ₄ ·2H ₂ O Mo powder 99.99% zinc chloride	618 mg, 2.55 mmol 50 mg, 0.52 mmol 136 mg, 1.00 mmol	H ₃ BTB	130 mg, 0.30 mmol	tetrabutylammonium hydroxide 40 wt % solution in water	120 μL, 0.18 mmol	H ₂ O	7 mL	180 °C	72 h
NENU-501	Na ₂ MoO ₄ ·2H ₂ O Mo powder 99.99% zinc chloride	618 mg, 2.55 mmol 50 mg, 0.52 mmol 136 mg, 1.00 mmol	H ₃ BPT	N/A	tetrabutylammonium hydroxide 40 wt % solution in water	120 μL, 0.18 mmol	H ₂ O	7 mL	180 °C	72 h

Fig. S1. Example of data re-annotation for Paragraph2MOFInfo task.

1.3. Paragraph2NMR Data Collection and Annotation

The original data of Nuclear magnetic resonance (NMR) were collected from the PubMed database (https://ftp.ncbi.nlm.nih.gov/pub/pmc/oa_bulk/oa_comm/), and consisted of 600 data items selected from as many different articles as possible, including text passages in a variety of formats and expressions. Each data entry contains seven items, IUPAC name of the compound, a text field containing the hydrogen spectrum and carbon spectrum, the conditions for measuring the hydrogen spectrum and carbon spectrum, and the hydrogen spectrum and carbon spectrum data containing only numerical values. The paragraphs have been carefully selected, and should contain the following elements:

1. The paragraphs should contain IUPAC name, or the IUPAC name should be replaced with “the title compound” or “compound a”.
2. The paragraphs should contain the nuclear magnetic conditions, such as temperature, solvent, and equipment power.
3. The paragraphs should contain hydrogen spectrum and carbon spectrum.

After obtaining the text segments, we organized the data for fine-tuning the ChatGPT according to the following format:

1. If the IUPAC name is unavailable in the text, the pronoun provided in the text will be used instead, such as “The title compound” and “compound a”, will be utilized.
2. The **NMR-related text** often begins with as “¹H NMR” and “¹³C NMR”.

An example: ¹³C NMR (151 MHz, CDCl₃): δ 160.1, 134.6, 123.0, 115.3, 112.3, 55.5, 32.6, 30.4, 20.8

3. **NMR conditions** include equipment frequency, solvent, or temperature, are listed in sequence, and separated by “,”. Such as: 400 MHz, DMSO–d₆, 298 K

4. **NMR data** only retain numerical values of chemical shift, including range values.

NMR data is separated by comma and space(“ , ”). Such as: 4.26, 4.15, 3.81, 3.73,

2.87–2.80

2. Examples of Prompt Engineering for Chemical Text Mining

2.1. Example for Paragraph2Compound Task

Prompt Engineering of Paragraph2Compound

Zero-shot prompt:

Please just extract all compound names in the paragraph, the compound names should be split in " | ". If there is no compound name in whole paragraph, please return "N/A".

{Input Paragraph}

Few-shots prompt:

Please just extract all compound names in the paragraph, the compound names should be split in " | ". If there is no compound name in whole paragraph, please return "N/A".

Example 1:

input:

Compound 610 (102 mg, 0.366 mmol) was dissolved in DMF (3 mL), and 3-pyridylacetic acid hydrochloride (635 mg, 3.66 mmol), EDC hydrochloride (702 mg, 3.66 mmol), 1-hydroxybenzotriazole monohydrate (561 mg, 3.66 mmol) and triethylamine (0.510 mL, 3.66 mmol) were added thereto, followed by stirring at 80° C. for 10 hours. The reaction mixture was poured into a saturated aqueous solution of sodium hydrogencarbonate, and the precipitated solid was collected by filtration. The resulting solid was purified through silica gel column chromatography (hexane:ethyl acetate=2:1), followed by recrystallizing from a mixed solvent of ethanol and water to afford the entitled Compound 618 (74.6 mg, 52%) as a white solid.

Output:

Compound 610 | DMF | 3-pyridylacetic acid hydrochloride | EDC hydrochloride | 1-hydroxybenzotriazole monohydrate | triethylamine | sodium hydrogencarbonate | silica gel | hexane:ethyl acetate | ethanol | water | Compound 618

Example 2:

input:

6-amino-2-[[[(1S)-1-methylbutyl]oxy]-9-[5-(1-piperidinyl)pentyl]-7,9-dihydro-8H-purin-8-one (0.384 g, 0.98 mmol) was dissolved in isopropyl alcohol (4.6 mL, 12 vols) and heated to 40° C. Maleic acid (0.114 g, 0.98 mmol) was added. A clear solution was obtained. During cooling to room temperature, precipitation occurred. The slurry was filtered, washing with isopropyl alcohol (5 mL) and dried under reduced pressure at 40° C. to constant weight. 6-amino-2-[[[(1S)-1-methylbutyl]oxy]-9-[5-(1-piperidinyl)pentyl]-7,9-dihydro-8H-purin-8-one, maleate salt (0.305 g, 61%) was obtained as a white solid.

Output:

6-amino-2-[[[(1S)-1-methylbutyl]oxy]-9-[5-(1-piperidinyl)pentyl]-7,9-dihydro-8H-purin-8-one | isopropyl alcohol | Maleic acid | isopropyl alcohol | 6-amino-2-[[[(1S)-1-methylbutyl]oxy]-9-[5-(1-piperidinyl)pentyl]-7,9-dihydro-8H-purin-8-one, maleate salt

Example N:

...

{Input Paragraph}

Fig. S2. An Illustration of Prompt Engineering for Paragraph2Compound Task.

2.2. Example for Paragraph2RXNRole (Product Extraction) Task

Prompt Engineering of Product Extraction

Zero-shot prompt:
Identify and annotate the product names in the provided paragraph. Use the format "<Prod*compound*Prod>" to mark the product compounds, appending "<Prod*" immediately before and "*Prod>" right after each product name. These markers must be used as a continuous string without any separation or additional content. The output should be the paragraph with all product names appropriately annotated.
{Input Paragraph}

Few-shots prompt:
{Same prompt as zero-shot prompt} or {Just simple instructions}

Examples:

Input1:
As observed for the glycosylations of the 4-OH acceptor 7 , the use of equimolar proportions of the donor 6 and the acceptor 16 , in glycosylation reactions , led to the isolation of significant quantities of the glucal 9 and low yields of the disaccharides.

Output1:
As observed for the glycosylations of the 4-OH acceptor 7 , the use of equimolar proportions of the donor 6 and the acceptor 16 , in glycosylation reactions , led to the isolation of significant quantities of the glucal <Prod*9*Prod> and low yields of the <Prod*disaccharides.*Prod>

Input2:
The reactions of 1a with phenols containing electron - donating multisubstituted groups gave the corresponding products 2n and 2p in good yields with moderate to good ee values (Table 3 , entries 14 and 16) , while 2,4-dichlorophenol gave 42 % yield and 60 % ee (Table 3 , entry 15).

Output2:
The reactions of 1a with phenols containing electron - donating multisubstituted groups gave the corresponding products <Prod*2n*Prod> and <Prod*2p*Prod> in good yields with moderate to good ee values (Table 3 , entries 14 and 16) , while 2,4-dichlorophenol gave 42 % yield and 60 % ee (Table 3 , entry 15).

Input3:
Reduction of 15 gave the inverted alcohol 16 (64 % overall from 12 / 13) , which was converted into its pen- tafluorophenol thiono ester derivative 17 (100 %) and deoxygenated to give 18 (100 %).

Output3:
Reduction of 15 gave the inverted alcohol 16 (64 % overall from 12 / 13) , which was converted into its pen- tafluorophenol thiono ester derivative <Prod*17*Prod> (100 %) and deoxygenated to give <Prod*18*Prod> (100 %).

...

{Input Paragraph}

Fig. S3. An Illustration of Prompt Engineering for Paragraph2RXNRole (Product Extraction) Task.

2.3. Example for Paragraph2RXNRole (Reaction Role Labelling) Task

Prompt Engineering of Reaction Role Labelling

Zero-shot prompt:

Extract and label key chemical reaction details such as reactants, yield, solvents, catalysts, temperatures, reaction types, and durations from the provided paragraph. Use specific string concatenations to annotate each element: prepend "<Reactants*", "<Yield*", "<Solvent*", "<Catalyst_Reagents*", "<Temperature*", "<Reaction*", "<Time*" before each respective item, and append "**Reactants>", "**Yield>", "**Solvent>", "**Catalyst_Reagents>", "**Temperature>", "**Reaction>", "**Time>" after. Ensure that the opening and closing strings for each category match and are directly adjacent to the named item, without any intervening or additional content. The output should be the paragraph with all elements accurately annotated and labeled.

{Input Paragraph}

Few-shots prompt:

{Same prompt as zero-shot prompt} or {Just simple instructions}

Examples:

Input1:

Pivaloylation of the diol 10 gave a mixture of monopivaloates <Prod*12*Prod> and 13 (1 : 2) (100 %).

Output1:

<Reaction*Pivaloylation*Reaction> of the diol <Reactants*10*Reactants> gave a mixture of monopivaloates <Prod*12*Prod> and 13 (1 : 2) (<Yield*100 %*Yield>).

Input2:

For example , enantiomerically pure monotriflate (S)-2a was converted into methyl ester (S)-5lo and carboxylic acid <Prod*(S)-6*Prod> in high yields by palladium - catalyzed carbonylation2 (Scheme 3).

Output2:

For example , enantiomerically pure monotriflate <Reactants*(S)-2a*Reactants> was converted into methyl ester (S)-5lo and carboxylic acid <Prod*(S)-6*Prod> in <Yield*high*Yield> yields by <Catalyst_Reagents*palladium*Catalyst_Reagents> - catalyzed <Reaction*carbonylation2*Reaction> (Scheme 3).

Input3:

After 6 h in acetonitrile at 50 ° C , the corresponding dipeptide <Prod*13*Prod> was obtained in 61 % yield.

Output3:

After <Time*6 h*Time> in <Solvent*acetonitrile*Solvent> at <Temperature*50 ° C*Temperature> , the corresponding dipeptide <Prod*13*Prod> was obtained in <Yield*61 %*Yield> yield.

...

{Input Paragraph}

Fig. S4. An Illustration of Prompt Engineering for Paragraph2RXNRole (Reaction Role Labelling) Task.

2.4. Example for Paragraph2MOFInfo Task

Prompt Engineering of Paragraph2MOFInfo

Zero-shot prompt:

Answer the question as truthfully as possible using the provided context. Please summarize the following details in a dictionaries list: compound name or chemical formula (if the name is not provided), metal source, metal amount, organic linker(s), linker amount, modulator, modulator amount or volume, solvent(s), solvent volume(s), reaction temperature, and reaction time. If any information is not provided or you are unsure, use "N/A". The list can contain multiple dictionaries, corresponding to multiple pieces of synthetic information. The dictionaries should have 11 columns, all in lowercase, must end with [{"compound name": "text", "metal source": "text", "metal amount": "text", "linker": "text", "linker amount": "text", "modulator": "text", "modulator amount or volume": "text", "solvent": "text", "solvent volume": "text", "reaction temperature": "text", "reaction time": "text"}, ...] format reply.

{Input Paragraph}

Few-shots prompt:

{Same prompt as zero-shot prompt} or {Just simple instructions}

Examples:

Input1:

Optimized Reaction Conditions and Scale-Up Synthesis for CAU-21-BPDC (1)The HT synthesis optimization led to the following optimized synthesis procedure: to a mixture of 129.6 μL of 2 M $\text{Al}(\text{NO}_3)_3 \cdot 9\text{H}_2\text{O}$ in H_2O (259 μmol), 70 mg 4,4'-benzophenonedicarboxylic acid (H2BPDC, 259 μmol), and 10 mg (HBA, 82 μmol) were added 300 μL of N,N-dimethylformamide (3.86 mmol) and 70.4 μL of deionized water (5.36 mmol). The reactor was sealed and heated to 110 $^\circ\text{C}$ within 6 h, held at this temperature for 24 h, and then cooled to room temperature within 72 h. The product was filtered off, washed with N,N-dimethylformamide at 65 $^\circ\text{C}$ for 3 h, filtered again, and subsequently activated at 200 $^\circ\text{C}$ at 10–2 kPa for 48 h. The additional washing step in DMF is necessary because the raw product contains recrystallized linker molecules. Larger amounts of CAU-21-BPDC can be obtained using 648.2 μL of 2 M $\text{Al}(\text{NO}_3)_3 \cdot 9\text{H}_2\text{O}$ in H_2O (1.3 mmol), 350 mg of 4,4'-benzophenonedicarboxylic acid (H2BPDC, 1.3 mmol), 50 mg of benzoic acid (410 μmol), 1.5 mL of N,N-dimethylformamide (19.3 mmol), and 352 μL of deionized water (26.8 mmol) in a 30 mL Teflon lined steel autoclave. The temperature program and the additional washing and activation steps are the same as in the optimized synthesis conditions of CAU-21-BPDC (yield: 0.27 g, 66% based on H2BPDC).

Output1:

{"compound name": "CAU-21-BPDC",	{"compound name": "CAU-21-BPDC",
"metal source": "Al(NO3)3·9H2O",	"metal source": "Al(NO3)3·9H2O",
"metal amount": "129.6 μL , 259 μmol ",	"metal amount": "648.2 μL , 1.3 mmol",
"linker": "4,4'-benzophenonedicarboxylic acid (H2BPDC)",	"linker": "4,4'-benzophenonedicarboxylic acid (H2BPDC)",
"linker amount": "70 mg, 259 μmol ",	"linker amount": "350 mg, 1.3 mmol",
"modulator": "benzoic acid (HBA)",	"modulator": "benzoic acid",
"modulator amount or volume": "10 mg, 82 μmol ",	"modulator amount or volume": "50 mg, 410 μmol ",
"solvent": "N,N-dimethylformamide deionized water",	"solvent": "N,N-dimethylformamide deionized water",
"solvent volume": "300 μL 70.4 μL ",	"solvent volume": "1.5 mL 352 μL ",
"reaction temperature": "110 $^\circ\text{C}$ ",	"reaction temperature": "110 $^\circ\text{C}$ ",
"reaction time": "24 h",	"reaction time": "24 h"}]

...

{Input Paragraph}

Fig. S5. An Illustration of Prompt Engineering for Paragraph2MOFInfo Task.

2.5. Example for Paragraph2NMR Task

Prompt Engineering of Paragraph2NMR

Zero-shot prompt:

Extract text containing ¹H NMR and ¹³C NMR data, remove interference information such as reactants, raw materials, solvents and other non-final product names based on text semantics, and then extract the name, code or number of the final product. Please delete the IUPAC name Alias, numbers and ordinal numbers before and after fields, such as '2.1.3.', '(HL4)', '(9)', '(4d)'. NMR text should contain complete information, such as instrument power and solvent information, For example, "¹³C NMR text": "¹³C NMR (400 MHz, acetone-d6) 174.0 (C), 157.7 (C). Then split the NMR text. The content in NMR conditions is NMR instrument power and solvent information, such as "¹³C NMR conditions": "400MHz, acetone-d6". The contents in the NMR data are only numbers, such as "¹³C NMR data": "174.0, 157.7", "¹H NMR data": "174.0, 157.7". All responses must originate from information extracted from the given text, ensuring that the extracted content has not been modified or fragmented, and that capitalization and punctuation are exactly the same as the given text. Must end with {"IUPAC":"text","¹H NMR text":"text","¹H NMR conditions":"text","¹H NMR data":"text","¹³C NMR text":"text","¹³C NMR conditions":"text","¹³C NMR data":"text"} format reply.

{Input Paragraph}

Few-shots prompt:

{Same prompt as zero-shot prompt} or {Just simple instructions}

Examples:

Input1:

2,5-Dioxocyclopentyl 2-(2,2,2-trifluoroacetamido)acetate (TFA-Gly-OSu, 3b). Colorless amorphous mass. ¹H-NMR (270 MHz, CD₃OD) δ: 4.44 (s, 2H, CH₂NH), 2.84 (s, 4H, 2×CH₂) ppm. ¹³C NMR (67.5 MHz, ACETONE-d₆) δ: 170.1 (2×CO), 165.7, 158.2 (q, 2JCF = 37.4 Hz), 116.8 (q, 1JCF = 287.0 Hz), 39.5, 26.2 (2×CH₂) ppm. HRMS-ESI (m/z) [M + H]⁺ 291.0205, found 291.0208.

Output1:

```
{"IUPAC":"2,5-Dioxocyclopentyl 2-(2,2,2-trifluoroacetamido)acetate",
"1H NMR text":" 1H-NMR (270 MHz, CD3OD) δ: 4.44 (s, 2H, CH2NH), 2.84 (s, 4H, 2×CH2) ppm",
"1H NMR conds":"270 MHz, CD3OD",
"1H NMR data":"4.44, 2.84",
"13C NMR text":"13C NMR (67.5 MHz, ACETONE-d6) δ: 170.1 (2×CO), 165.7 158.2 (q, 2JCF = 37.4 Hz), 116.8
(q, 1JCF = 287.0 Hz), 39.5, 26.2 (2×CH2) ppm ",
"13C NMR conds":"67.5 MHz, ACETONE-d6",
"13C NMR data":"170.1, 165.7, 158.2, 116.8, 39.5, 26.2"}
...
```

{Input Paragraph}

Fig. S6. An Illustration of Prompt Engineering for Paragraph2NMR Task.

2.6. Example for Paragraph2Action Task

Prompt Engineering of Paragraph2Action

Few-shots prompt:

Convert experimental procedures to structured synthetic steps (action sequences). Please following the .The action types can be chosen from: ADD, COLLECTLAYER, CONCENTRATE, DEGAS, DRYSLID, DRYSLUTION, EXTRACT, FILTER, MAKESOLUTION, MICROWAVE, PARTITION, PH, PHASESEPARATION, PURIFY, QUENCH, RECRYSTALLIZE, REFLUX, SETTEMPERATURE, SONICATE, STIR, TRITURATE, WAIT, WASH, YIELD, FOLLOWOTHERPROCEDURE, INVALIDACTION, OTHERLANGUAGE, and NOACTION. The amounts needs to be given in the bracket after the compound. Here are some examples.

Examples

Input1:
After stirring at 70° C. for 15 hours, the reaction mixture was partitioned between ethyl acetate (15 mL) and water (15 mL).

Output1:
STIR for 15 hours at 70° C; PARTITION with ethyl acetate (15 mL) and water (15 mL).

Input2:
The resultant solid formed was collected by filtration and dried to provide 5-(2-(6-Methylpyridin-2-yl)pyridin-3-yl)-1H-indazole (118 mg, 47%) as a white solid.

Output2:
FILTER keep precipitate; DRYSLID; YIELD 5-(2-(6-Methylpyridin-2-yl)pyridin-3-yl)-1H-indazole (118 mg, 47%).

Input3:
After 8-14 hours, the reaction mixture was cooled to 0° C. and treated with sodium ethoxide (1012 µl, 0.682 mmol) and subsequently heated to reflux (135° C. bath temp).

Output3:
WAIT for 8-14 hours; SETTEMPERATURE 0° C; ADD sodium ethoxide (1012 µl, 0.682 mmol); REFLUX.

Input4:
The solvents were evaporated in vacuo; the residue was diluted with water (25 mL) and acidified with acetic acid (0.3 mL).

Output4:
CONCENTRATE; ADD water (25 mL); PH with acetic acid (0.3 mL) to pH acidic.

Input5:
After filtration, the white solid is sonicated in H2O (10 mL), then filtered to give a tan solid of the title compound.

Output5:
FILTER keep precipitate; ADD H2O (10 mL); SONICATE; FILTER keep precipitate; YIELD title compound.

...

{Input Paragraph}

Fig. S7. An Illustration of Prompt Engineering for Paragraph2Action Task.

2.7. Example for Zero-shot Prompts Designing on Paragraph2NMR Task

Paragraph:
Data for dimorpholine-dithiocarbamic acid 3-dimorpholinethiocarbamoylsulphanylmethyl-1,4-dioxo-1,4-dihydro-naphthalen-2-ylmethyl ester (3a): yellow solid (78.4%); mp 153–154 °C. ¹H NMR (400 MHz, CDCl₃) δ 8.11–8.13 (m, 2H, ArH), 7.75–7.77 (m, 2H, ArH), 4.89 (s, 4H, 2CH₂S), 3.99–4.28 (m, 8H, 4OCH₂), 3.76 (s, 8H, 4NCH₂). ¹³C NMR (100 MHz, CDCl₃) δ 196.1, 183.9, 143.7, 134.0, 132.0, 126.7, 66.3, 66.2, 34.0. HR-MS (ESI+) m/z: 509.0697 [M + H]⁺. Found: 509.0686 [M + H]⁺.

Prompt1:
"Strictly follow the IUPAC naming rules and the document format writing rules of nuclear magnetic spectrum data, extract the IUPAC name of the compound and the nuclear magnetic data text in the given text, and then split the nuclear magnetic data text into conditions and data, with {'IUPAC': 'text', '1H NMR condition': 'text', '1H NMR data': 'text', '13C NMR condition': 'text', '13C NMR data': 'text'} dictionary format reply, if the content does not exist, it will be N/A. The IUPAC name of the compound should remove information such as code name and number to ensure accuracy."

Output1:
'IUPAC': 'dimorpholine-dithiocarbamic acid 3-dimorpholinethiocarbamoylsulphanylmethyl-1,4-dioxo-1,4-dihydro-naphthalen-2-ylmethyl ester',
'1H NMR cond.': '(400 MHz, CDCl₃)',
'1H NMR data': 'δ 8.11–8.13 (m, 2H, ArH), 7.75–7.77 (m, 2H, ArH), 4.89 (s, 4H, 2CH₂S), 3.99–4.28 (m, 8H, 4OCH₂), 3.76 (s, 8H, 4NCH₂)',
'13C NMR cond.': '(100 MHz, CDCl₃)',
'13C NMR data': 'δ 196.1, 183.9, 143.7, 134.0, 132.0, 126.7, 66.3, 66.2, 34.0'

Prompt2:
"Strictly follow the IUPAC naming rules and the document format writing rules of nuclear magnetic spectrum data, extract the IUPAC name of the compound in the given text, the NMR data text field containing complete information, and split the NMR data text field into two parts: conditions and data, with {'IUPAC': 'text', '1H NMR condition': 'text', '1H NMR data': 'text', '13C NMR condition': 'text', '13C NMR data': 'text'} Dictionary format reply, if the content does not exist, it will be N/A. The IUPAC name of the compound should remove information such as code name and number to ensure accuracy."

Output2:
"IUPAC": "Dimorpholine-dithiocarbamic acid 3-dimorpholinethiocarbamoylsulphanylmethyl-1,4-dioxo-1,4-dihydro-naphthalen-2-ylmethyl ester",
"1H NMR cond.": "400 MHz, CDCl₃",
"1H NMR data": "δ 8.11–8.13 (m, 2H, ArH), 7.75–7.77 (m, 2H, ArH), 4.89 (s, 4H, 2CH₂S), 3.99–4.28 (m, 8H, 4OCH₂), 3.76 (s, 8H, 4NCH₂)",
"13C NMR cond.": "100 MHz, CDCl₃",
"13C NMR data": "δ 196.1, 183.9, 143.7, 134.0, 132.0, 126.7, 66.3, 66.2, 34.0"

Prompt3:
"Strictly follow the IUPAC naming rules and the document format writing rules of nuclear magnetic spectrum data, extract the IUPAC name of the compound in the given text, the NMR data text field containing complete information, and split the NMR data text field into two parts: conditions and data, the condition part contains information such as solvent and power, and the data part deletes the position information and only retains the numbers, as {'IUPAC': 'text', '1H NMR condition': 'text', '1H NMR data': 'text', '13C NMR condition': 'text', '13C NMR data': 'text'} Dictionary format reply, N if the content does not exist /A, IUPAC names of compounds should remove information such as code names and numbers to ensure accuracy."

Output3:
'IUPAC': 'dimorpholine-dithiocarbamic acid 3-dimorpholinethiocarbamoylsulphanylmethyl-1,4-dioxo-1,4-dihydro-naphthalen-2-ylmethyl ester',
'1H NMR cond.': '(400 MHz, CDCl₃)',
'1H NMR data': '8.11–8.13, 7.75–7.77, 4.89, 3.99–4.28, 3.76',
'13C NMR cond.': '(100 MHz, CDCl₃)',
'13C NMR data': '196.1, 183.9, 143.7, 134.0, 132.0, 126.7, 66.3, 66.2, 34.0'

Expected:
"IUPAC": "dimorpholine-dithiocarbamic acid 3-dimorpholinethiocarbamoylsulphanylmethyl-1,4-dioxo-1,4-dihydro-naphthalen-2-ylmethyl ester",
"1H NMR cond.": "400 MHz, CDCl₃",
"1H NMR data": "8.11–8.13, 7.75–7.77, 4.89, 3.99–4.28, 3.76",
"13C NMR cond.": "100 MHz, CDCl₃",
"13C NMR data": "196.1, 183.9, 143.7, 134.0, 132.0, 126.7, 66.3, 66.2, 34.0"}"

Fig. S8. The process of repeatedly revising prompts to extract expected formatting data.

3. Processing Details

3.1. Processing for ChatGPT

Thanks to its powerful comprehension and formatted capabilities, apart from transforming data into jsonl format before fine-tuning, no additional special preprocessing or postprocessing is required.

3.2. Post-Processing for Open-source LLMs (Mistral, Llama3, Llama2)

We found that the fine-tuned LLMs often finds it difficult to stop, as if it is difficult to generate the built-in "</s>" stop token. Therefore, we added "!!!" as a stop signal at the end of each output of the training set when fine-tuning LLMs. We then consider the content up to the first occurrence of "!!!" in the generated text as the response, by setting "!!!" as the stop token.

3.3. Pre-Processing for T5 and BART

We employ the multi-task learning strategy to fine-tune T5 and BART for Paragraph2MOFInfo and Paragraph2NMR, due to their limitation in generating multi-attribute long text. (Fig. S8, S9)

3.4. Post-Processing for T5 and BART

In Paragraph2RXNRole task, we found that LLMs such as ChatGPT, Mistral, Llama3, and Llama2 work well with little extra post-processing, but T5 and BART do necessitate additional, specialized post-processing. This requirement stems mainly from T5 and BART's limitation to accurately replicate text from the source, often resulting in missed or incorrectly copied characters. Also, T5 doesn't have tokens like

'<', '{', '}', and '/' in its vocabulary. Thus, the generated sentences by T5 always miss these tokens. To fix this for the Paragraph2RXNRole task, we modify the output by replacing 'Prod*' with '<Prod*', 'Reactants*' with '<Reactants*', 'Yield*' with '<Yield*', 'Solvent*' with '<Solvent*', 'Catalyst_Reagents*' with '<Catalyst_Reagents*', 'Temperature*' with '<Temperature*', 'Reaction*' with '<Reaction*', and 'Time*' with '<Time*'. This post-processing step ensures the inclusion of necessary tokens and improves the accuracy of the model's output for this task.

4. Training Details

4.1. Hyperparameters Tuning

Since ChatGPT fine-tuning does not allow for adjusting hyperparameters, we reported the performances of all models at the best epoch selected from the evaluation set for fair comparison. For the prompt-only ChatGPT, we provided the different number of examples and experimented with different prompts to enhance the model's performance. For T5-base and BART-base models, the hyperparameter we adjusted include epoch, learning rate and batch_size. In addition to epoch and learning rate, we also tried to adjust the lora_r and lora_alpha parameters for the llama2-13b-chat model, large lora_r is better (Table S2). For full fine-tuning Llama3-8b-instruct and Mistral-7b-instruct-v0.2, the hyperparameter we adjusted include epoch and learning rate. Default hyperparameters can be found in the GitHub. The best epochs are determined by the inference results of each epoch rather than evaluating loss.

Table S2. Hyperparameters for the models.

Model	Hyperparameters
GPT-3.5-turbo (prompt-only)	prompts; examples; number of examples
GPT-3.5-turbo (fine-tune)	epoch
Llama2-13b-chat (q-lora fine-tune)	epoch; lora_r {8, 64 }; lora_alpha {16, 128 }; learning rate {1e-3, 1e-4 , 1e-5}
Llama3-8b-instruct (full fine-tune)	epoch; learning rate {1e-5, 5e-6 , 1e-6}; batch size {2, 4}
Mistral-7b-instruct-v0.2 (full fine-tune)	epoch; learning rate {1e-5, 5e-6 , 1e-6}; batch size {2, 4}
T5-base (full fine-tune)	epoch; learning rate {1e-3, 1e-4 , 1e-5}; batch size {4, 8}
BART-base (full fine-tune)	epoch; learning rate {1e-3, 1e-4, 1e-5 }; batch size {4, 8}

The hyperparameters in black bold are recommended.

4.2. Hardware resources and memory cost

Table S3. Fine-tuning and Inference Cost

Task	Model	Strategy	Number of Fine-tuning Data	Fine-tuning Cost	Number of Inference Data	Inference Cost	Hyperparameters
Paragraph2Action	GPT-3.5-turbo-0613	full fine-tuning	1060	4 epochs, total 1.5 h, \$ 4	352	~ 7 min (using API)	-
	Mistral-7b-instruct-v0.2	full fine-tuning		~ 4 epochs × 2 min/epoch on 4×40GB A100 (~24 GB/GPU)		~ 10 s on 1×40GB A100 (using vllm)	lr = 5e-6, bs = 4, max_l = 4096
	Llama3-8b-instruct	full fine-tuning		~ 3 epochs × 10 min/epoch on 4×40GB A100 (~26 GB/GPU)		~ 10 s on 1×40GB A100 (using vllm)	lr = 5e-6, bs = 4, max_l = 4096
	Llama2-13b-chat-hf	q-lora fine-tuning		~ 5 epochs × 7 min/epoch on 1×40GB A100 (~12 GB/GPU)		~ 20 s on 1×40GB A100 (using vllm)	lr = 1e-4, bs = 2, max_l = 4096
	Bart-base	full fine-tuning		~ 5 epoch × 10 s/epoch on 1×40GB A100 (~7 GB/GPU)		~ 35 s on 1×40GB A100	lr = 1e-4, bs = 8, max_l = 512
	T5-base	full fine-tuning		~ 31 epoch × 20 s/epoch on 1×40GB A100 (~11 GB/GPU)		~ 80 s on 1×40GB A100	lr = 1e-4, bs = 8, max_l = 512
Paragraph2Compound	GPT-3.5-turbo-0613	full fine-tuning	1000	5 epochs, total 1.5 h, \$ 92	1000	~ 7 min (using API)	-
	Mistral-7b-instruct-v0.2	full fine-tuning		~ 1 epoch × 20 min/epoch on 4×40GB A100 (~28 GB/GPU)		~ 10 s on 1×40GB A100 (using vllm)	lr = 5e-6, bs = 4, max_l = 4096
	Llama3-8b-instruct	full fine-tuning		~ 2 epoch × 100 min/epoch on 4×40GB A100 (~32 GB/GPU)		~ 10 s on 1×40GB A100 (using vllm)	lr = 5e-6, bs = 4, max_l = 4096
	Llama2-13b-chat-hf	q-lora fine-tuning		~ 3 epochs × 100 min/epoch on 1×40GB A100 (~20 GB/GPU)		~ 20 s on 1×40GB A100 (using vllm)	lr = 1e-5, bs = 2, max_l = 4096
	Bart-base	full fine-tuning		~ 10 epochs × 3 min/epoch on 1×40GB A100 (~16 GB/GPU)		~ 40 s on 1×40GB A100	lr = 1e-5, bs = 8, max_l = 512
	T5-base	full fine-tuning		~ 19 epochs × 5 min/epoch on 1×40GB A100 (~30 GB/GPU)		~ 60 s on 1×40GB A100	lr = 1e-4, bs = 8, max_l = 512
Paragraph2NMR	GPT-3.5-turbo-0613	full fine-tuning	1000	3 epochs, total 1.5 h, \$ 9	1000	~ 23 min (using API)	-
	Mistral-7b-instruct-v0.2	full fine-tuning		~ 5 epochs × 3 min/epoch on 4×40GB A100 (~26 GB/GPU)		~ 1 min on 1×40GB A100 (using vllm)	lr = 5e-6, bs = 2, max_l = 4096
	Llama3-8b-instruct	full fine-tuning		~ 2 epochs × 8 min/epoch on 4×40GB A100 (~29 GB/GPU)		~ 1 min on 1×40GB A100 (using vllm)	lr = 5e-6, bs = 2, max_l = 4096
	Llama2-13b-chat-hf	q-lora fine-tuning		~ 3 epochs × 11 min/epoch on 1×40GB A100 (~20 GB/GPU)		~ 3 min on 1×40GB A100 (using vllm)	lr = 1e-5, bs = 2, max_l = 4096
	Bart-base	full fine-tuning		~ 36 epochs × 20 s/epoch on 1×40GB A100 (~13 GB/GPU)		~ 12 min on 1×40GB A100	lr = 1e-5, bs = 8, max_l = 512
	T5-base	full fine-tuning		~ 30 epochs × 40 s/epoch on 1×40GB A100 (~21 GB/GPU)		~ 9 min on 1×40GB A100	lr = 1e-5, bs = 8, max_l = 512
Paragraph2Prod	GPT-3.5-turbo-0613	full fine-tuning	10000	3 epochs, total 3 h, \$ 90	1000	~ 23 min (using API)	-
	Mistral-7b-instruct-v0.2	full fine-tuning		~ 2 epochs × 28 min/epoch on 4×40GB A100 (~28 GB/GPU)		~ 1 min on 1×40GB A100 (using vllm)	lr = 5e-6, bs = 2, max_l = 4096
	Llama3-8b-instruct	full fine-tuning		~ 2 epochs × 80 min/epoch on 4×40GB A100 (~32 GB/GPU)		~ 1 min on 1×40GB A100 (using vllm)	lr = 5e-6, bs = 2, max_l = 4096
	Llama2-13b-chat-hf	q-lora fine-tuning		~ 5 epochs × 110 min/epoch on 1×40GB A100 (~20 GB/GPU)		~ 3 min on 1×40GB A100 (using vllm)	lr = 1e-5, bs = 2, max_l = 4096
	Bart-base	full fine-tuning		~ 42 epochs × 3 min/epoch on 1×40GB A100 (~16 GB/GPU)		~ 12 min on 1×40GB A100	lr = 1e-5, bs = 8, max_l = 512
	T5-base	full fine-tuning		~ 38 epochs × 6 min/epoch on 1×40GB A100 (~22 GB/GPU)		~ 9 min on 1×40GB A100	lr = 1e-4, bs = 8, max_l = 512
Paragraph2Role	GPT-3.5-turbo-0613	full fine-tuning	200	3 epochs, total 30 min, \$ 15	300	~ 15 min (using API)	-
	Mistral-7b-instruct-v0.2	full fine-tuning		~ 2 epochs × 2 min/epoch on 4×40GB A100 (~38 GB/GPU)		~ 2 min on 1×40GB A100 (using vllm)	lr = 5e-6, bs = 2, max_l = 4096
	Llama3-8b-instruct	full fine-tuning		~ 4 epochs × 4 min/epoch on 4×40GB A100 (~33 GB/GPU)		~ 2 min on 1×40GB A100 (using vllm)	lr = 5e-6, bs = 2, max_l = 4096
	Llama2-13b-chat-hf	q-lora fine-tuning		~ 3 epochs × 5 min/epoch on 1×40GB A100 (~20 GB/GPU)		~ 3 min on 1×40GB A100 (using vllm)	lr = 1e-5, bs = 2, max_l = 4096
	Bart-base	full fine-tuning		~ 17 epochs × 30 s/epoch on 1×40GB A100 (~13 GB/GPU)		~ 4 min on 1×40GB A100	lr = 1e-5, bs = 8, max_l = 512
	T5-base	full fine-tuning		~ 19 epochs × 60 s/epoch on 1×40GB A100 (~23 GB/GPU)		~ 7 min on 1×40GB A100	lr = 1e-4, bs = 8, max_l = 512
Paragraph2MOFInfo	GPT-3.5-turbo-0613	full fine-tuning	613	1 epoch, total 1h, \$ 7	723	~ 12 min (using API)	-
	Mistral-7b-instruct-v0.2	full fine-tuning		~ 1 epoch × 14 min/epoch on 4×40GB A100 (~22 GB/GPU)		~ 15 s on 1×40GB A100 (using vllm)	lr = 5e-6, bs = 2, max_l = 4096
	Llama3-8b-instruct	full fine-tuning		~ 1 epoch × 40 min/epoch on 4×40GB A100 (~24 GB/GPU)		~ 15 s on 1×40GB A100 (using vllm)	lr = 5e-6, bs = 2, max_l = 4096
	Llama2-13b-chat-hf	q-lora fine-tuning		~ 2 epoch × 40 min/epoch on 1×40GB A100 (~12 GB/GPU)		~ 20 s on 1×40GB A100 (using vllm)	lr = 1e-5, bs = 2, max_l = 4096
	Bart-base	full fine-tuning		~ 28 epochs × 1 min/epoch on 1×40GB A100 (~8 GB/GPU)		~ 50 s on 1×40GB A100	lr = 1e-5, bs = 8, max_l = 512
	T5-base	full fine-tuning		~ 9 epochs × 2 min/epoch on 1×40GB A100 (~16 GB/GPU)		~ 60 s on 1×40GB A100	lr = 1e-4, bs = 8, max_l = 512
Paragraph2MOFInfo	GPT-3.5-turbo-0613	full fine-tuning	599	5 epochs, total 1h, \$ 5	111	~ 4 min (using API)	-
	Mistral-7b-instruct-v0.2	full fine-tuning		~ 6 epochs × 2 min/epoch on 4×40GB A100 (~21 GB/GPU)		~ 7 s on 1×40GB A100 (using vllm)	lr = 5e-6, bs = 2, max_l = 4096
	Llama3-8b-instruct	full fine-tuning		~ 10 epochs × 5 min/epoch on 4×40GB A100 (~23 GB/GPU)		~ 7 s on 1×40GB A100 (using vllm)	lr = 5e-6, bs = 2, max_l = 4096
	Llama2-13b-chat-hf	q-lora fine-tuning		~ 2 epoch × 5 min/epoch on 1×40GB A100 (~11 GB/GPU)		~ 10 s on 1×40GB A100 (using vllm)	lr = 1e-4, bs = 2, max_l = 4096
	Bart-base	full fine-tuning		~ 39 epochs × 10 s/epoch on 1×40GB A100 (~7 GB/GPU)		~ 15 s on 1×40GB A100	lr = 1e-5, bs = 8, max_l = 512
	T5-base	full fine-tuning		~ 22 epochs × 20 s/epoch on 1×40GB A100 (~12 GB/GPU)		~ 20 s on 1×40GB A100	lr = 1e-5, bs = 8, max_l = 512
Paragraph2MOFInfo	GPT-3.5-turbo-0613	full fine-tuning	329	3 epochs, total 30 min, \$ 15	329	~ 15 min (using API)	-
	Mistral-7b-instruct-v0.2	full fine-tuning		~ 6 epochs × 2 min/epoch on 4×40GB A100 (~29 GB/GPU)		~ 2 min on 1×40GB A100 (using vllm)	lr = 5e-6, bs = 2, max_l = 4096
	Llama3-8b-instruct	full fine-tuning		~ 2 epochs × 4 min/epoch on 4×40GB A100 (~31 GB/GPU)		~ 2 min on 1×40GB A100 (using vllm)	lr = 5e-6, bs = 2, max_l = 4096
	Llama2-13b-chat-hf	q-lora fine-tuning		~ 3 epochs × 5 min/epoch on 1×40GB A100 (~16 GB/GPU)		~ 3 min on 1×40GB A100 (using vllm)	lr = 1e-5, bs = 2, max_l = 4096
	Bart-base	full fine-tuning		~ 160 epochs × 1 min/epoch on 1×40GB A100 (~11 GB/GPU)		~ 2 min on 1×40GB A100	lr = 1e-5, bs = 8, max_l = 1024
	T5-base	full fine-tuning		~ 140 epochs × 2 min/epoch on 1×40GB A100 (~22 GB/GPU)		~ 2 min on 1×40GB A100	lr = 1e-4, bs = 8, max_l = 1024

4.3. Multi-Task Learning of T5 and BART for Paragraph2MOFInfo

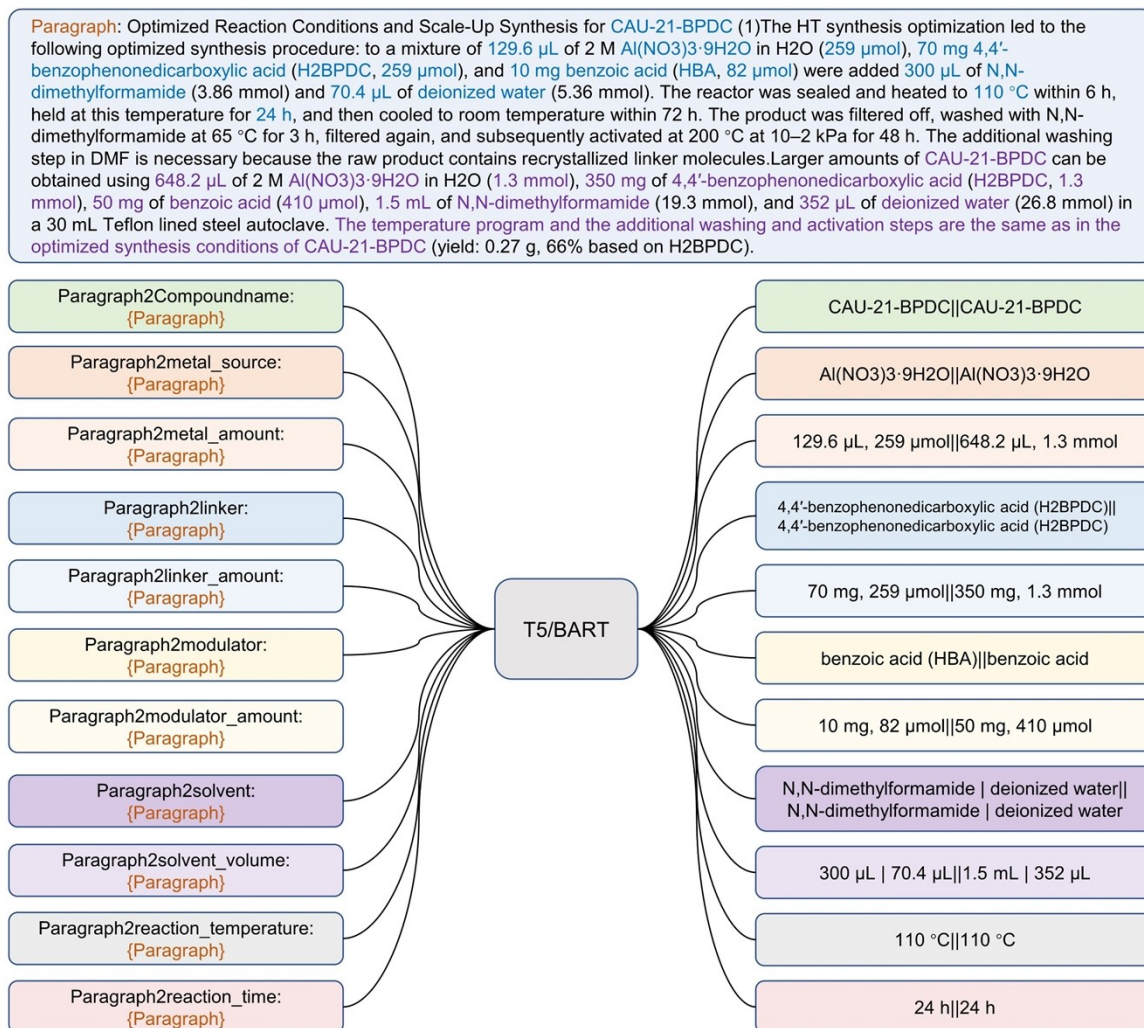


Fig. S9. Multi-task Learning Scheme of T5 and BART for Paragraph2MOF Task. Multiple components are separated with “||”. The different values of the same component are separated with “|”.

4.4. Multi-Task Learning of T5 and BART for Paragraph2NMR

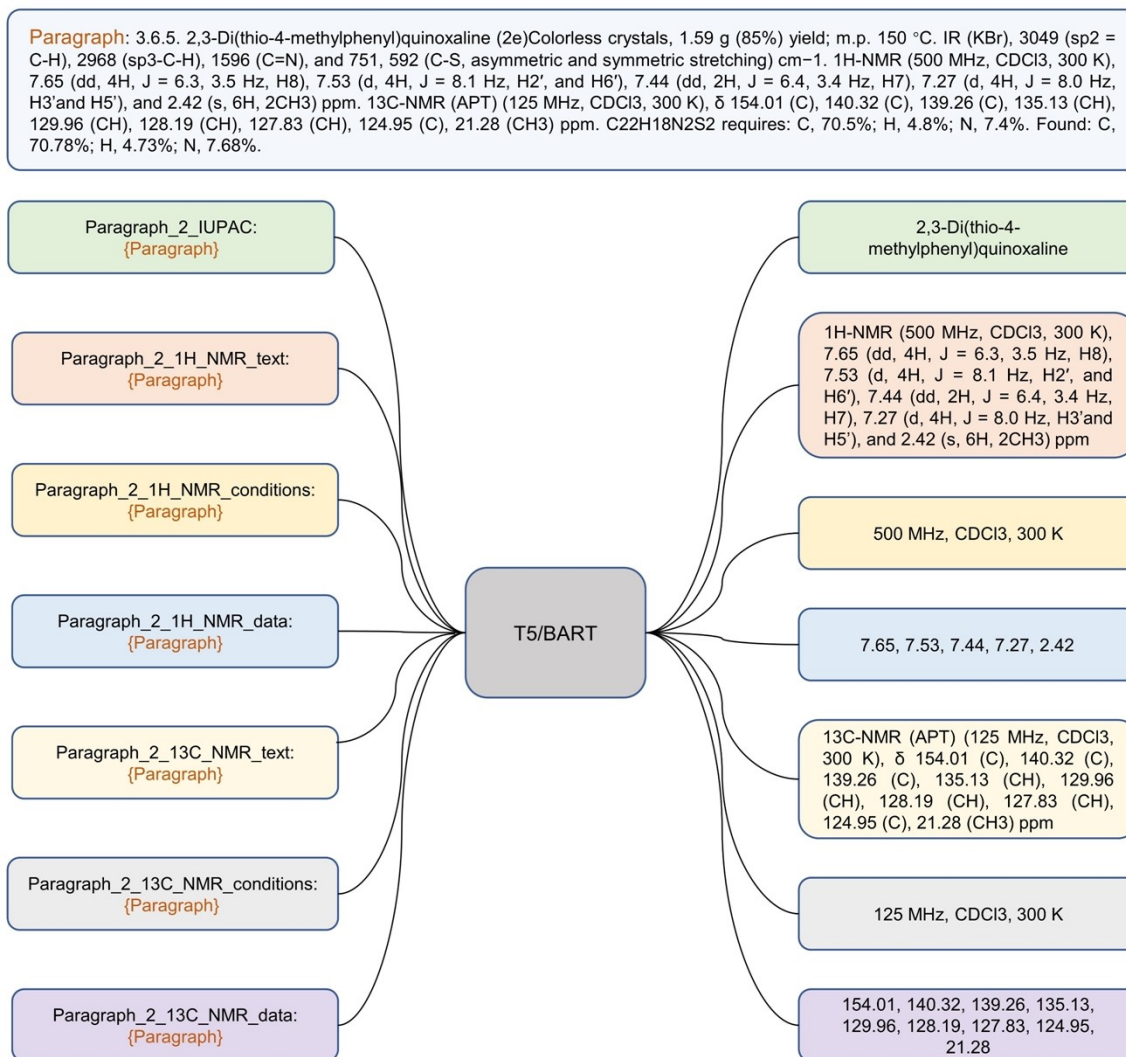


Fig. S10. Multi-Task Learning Scheme of T5 and BART for Paragraph2NMR Task.

5. Supplemental Results of the Performances

5.1. Performance of Paragraph2Compound

Table S4 Performance on Trial 1 for Paragraph2Compound

Model	Train set	Test set	Precision	Recall	F1 score	Jaccard index
GPT-3.5-turbo-0613 (zero-shot)	0	1000	0.648	0.590	0.600	0.479
GPT-3.5-turbo-0613 (three-shots)	0	1000	0.732	0.721	0.711	0.579
GPT-4-0613 (zero-shot)	0	1000	0.789	0.662	0.705	0.575
GPT-4-0613 (three-shots)	0	1000	0.793	0.731	0.748	0.625
GPT-3.5-turbo-0613 (fine-tuned)	10	1000	0.739	0.761	0.727	0.596
GPT-3.5-turbo-0613 (fine-tuned)	100	1000	0.825	0.859	0.832	0.736
GPT-3.5-turbo-0613 (fine-tuned)	1000	1000	0.849	0.883	0.860	0.776
GPT-3.5-turbo-0613 (fine-tuned)	1000	10000	0.851	0.879	0.859	0.775
GPT-3.5-turbo-0613 (fine-tuned)	10000	1000	0.878	0.897	0.883	0.812
Llama3-8b-instruct (fine-tuned)	10	1000	0.535	0.624	0.545	0.434
Llama3-8b-instruct (fine-tuned)	100	1000	0.790	0.816	0.791	0.683
Llama3-8b-instruct (fine-tuned)	1000	1000	0.854	0.848	0.842	0.752
Llama3-8b-instruct (fine-tuned)	1000	10000	0.854	0.841	0.839	0.747
Llama3-8b-instruct (fine-tuned)	10000	1000	0.859	0.873	0.860	0.778
Mistral-7b-instruct-v0.2 (fine-tuned)	10	1000	0.642	0.765	0.666	0.539
Mistral-7b-instruct-v0.2 (fine-tuned)	100	1000	0.809	0.836	0.811	0.711
Mistral-7b-instruct-v0.2 (fine-tuned)	1000	1000	0.848	0.872	0.853	0.767
Mistral-7b-instruct-v0.2 (fine-tuned)	1000	10000	0.853	0.866	0.853	0.768
Mistral-7b-instruct-v0.2 (fine-tuned)	10000	1000	0.875	0.889	0.878	0.804
Llama2-13b-chat (q-lora fine-tuned)	10	1000	0.625	0.621	0.579	0.443
Llama2-13b-chat (q-lora fine-tuned)	100	1000	0.758	0.794	0.756	0.643
Llama2-13b-chat (q-lora fine-tuned)	1000	1000	0.830	0.853	0.830	0.737
Llama2-13b-chat (q-lora fine-tuned)	1000	10000	0.829	0.844	0.826	0.731
Llama2-13b-chat (q-lora fine-tuned)	10000	1000	0.856	0.882	0.863	0.783
T5-base (fine-tuned)	10	1000	0.691	0.734	0.693	0.564
T5-base (fine-tuned)	100	1000	0.806	0.829	0.810	0.708
T5-base (fine-tuned)	1000	1000	0.823	0.863	0.837	0.746
T5-base (fine-tuned)	1000	10000	0.829	0.861	0.839	0.750
T5-base (fine-tuned)	10000	1000	0.865	0.880	0.870	0.796
BART-base (fine-tuned)	10	1000	0.607	0.547	0.552	0.415
BART-base (fine-tuned)	100	1000	0.732	0.640	0.666	0.539
BART-base (fine-tuned)	1000	1000	0.777	0.690	0.718	0.599
BART-base (fine-tuned)	1000	10000	0.786	0.685	0.717	0.596
BART-base (fine-tuned)	10000	1000	0.793	0.723	0.747	0.631

Table S5 Performance on Trial 2 for Paragraph2Compound

Model	Train set	Test set	Precision	Recall	F1 score	Jaccard index
GPT-3.5-turbo-0613 (zero-shot)	0	1000	0.663	0.612	0.611	0.480
GPT-3.5-turbo-0613 (three-shots)	0	1000	0.683	0.710	0.680	0.542
GPT-4-0613 (zero-shot)	0	1000	0.776	0.670	0.704	0.571
GPT-4-0613 (three-shots)	0	1000	0.776	0.708	0.729	0.600
GPT-3.5-turbo-0613 (fine-tuned)	10	1000	0.741	0.746	0.720	0.593
GPT-3.5-turbo-0613 (fine-tuned)	100	1000	0.817	0.856	0.826	0.729
GPT-3.5-turbo-0613 (fine-tuned)	1000	1000	0.849	0.880	0.857	0.772
GPT-3.5-turbo-0613 (fine-tuned)	10000	1000	0.879	0.891	0.881	0.810
Llama3-8b-instruct (fine-tuned)	10	1000	0.578	0.635	0.572	0.482
Llama3-8b-instruct (fine-tuned)	100	1000	0.759	0.779	0.752	0.633
Llama3-8b-instruct (fine-tuned)	1000	1000	0.821	0.853	0.844	0.755
Llama3-8b-instruct (fine-tuned)	10000	1000	0.863	0.878	0.864	0.783
Mistral-7b-instruct-v0.2 (fine-tuned)	10	1000	0.656	0.744	0.674	0.545
Mistral-7b-instruct-v0.2 (fine-tuned)	100	1000	0.778	0.824	0.788	0.681
Mistral-7b-instruct-v0.2 (fine-tuned)	1000	1000	0.851	0.870	0.853	0.768
Mistral-7b-instruct-v0.2 (fine-tuned)	10000	1000	0.880	0.887	0.879	0.807
Llama2-13b-chat (q-lora fine-tuned)	10	1000	0.620	0.617	0.582	0.445
Llama2-13b-chat (q-lora fine-tuned)	100	1000	0.758	0.772	0.745	0.629
Llama2-13b-chat (q-lora fine-tuned)	1000	1000	0.837	0.832	0.825	0.728
Llama2-13b-chat (q-lora fine-tuned)	10000	1000	0.867	0.873	0.866	0.787
T5-base (fine-tuned)	10	1000	0.706	0.710	0.688	0.558
T5-base (fine-tuned)	100	1000	0.791	0.822	0.797	0.692
T5-base (fine-tuned)	1000	1000	0.829	0.858	0.838	0.748
T5-base (fine-tuned)	10000	1000	0.869	0.877	0.870	0.796
BART-base (fine-tuned)	10	1000	0.562	0.463	0.477	0.342
BART-base (fine-tuned)	100	1000	0.715	0.655	0.668	0.536
BART-base (fine-tuned)	1000	1000	0.819	0.633	0.693	0.572
BART-base (fine-tuned)	10000	1000	0.795	0.714	0.740	0.625

Table S6 Performance on Trial 3 for Paragraph2Compound

Model	Train set	Test set	Precision	Recall	F1 score	Jaccard index
GPT-3.5-turbo-0613 (zero-shot)	0	1000	0.643	0.676	0.638	0.503
GPT-3.5-turbo -0613 (three-shots)	0	1000	0.675	0.75	0.694	0.56
GPT-4-0613 (zero-shot)	0	1000	0.805	0.672	0.716	0.584
GPT-4-0613 (three-shots)	0	1000	0.781	0.779	0.770	0.654
GPT-3.5-turbo-0613 (fine-tuned)	10	1000	0.704	0.765	0.711	0.584
GPT-3.5-turbo-0613 (fine-tuned)	100	1000	0.807	0.861	0.823	0.724
GPT-3.5-turbo-0613 (fine-tuned)	1000	1000	0.851	0.880	0.859	0.775
GPT-3.5-turbo-0613 (fine-tuned)	10000	1000	0.879	0.893	0.881	0.809
Llama3-8b-instruct (fine-tuned)	10	1000	0.661	0.603	0.605	0.479
Llama3-8b-instruct (fine-tuned)	100	1000	0.763	0.768	0.751	0.635
Llama3-8b-instruct (fine-tuned)	1000	1000	0.853	0.841	0.839	0.748
Llama3-8b-instruct (fine-tuned)	10000	1000	0.867	0.872	0.863	0.782
Mistral-7b-instruct-v0.2 (fine-tuned)	10	1000	0.674	0.744	0.671	0.550
Mistral-7b-instruct-v0.2 (fine-tuned)	100	1000	0.810	0.825	0.805	0.703
Mistral-7b-instruct-v0.2 (fine-tuned)	1000	1000	0.856	0.861	0.851	0.766
Mistral-7b-instruct-v0.2 (fine-tuned)	10000	1000	0.878	0.887	0.879	0.807
Llama2-13b-chat (q-lora fine-tuned)	10	1000	0.624	0.632	0.595	0.461
Llama2-13b-chat (q-lora fine-tuned)	100	1000	0.750	0.788	0.748	0.636
Llama2-13b-chat (q-lora fine-tuned)	1000	1000	0.819	0.844	0.820	0.726
Llama2-13b-chat (q-lora fine-tuned)	10000	1000	0.860	0.882	0.866	0.787
T5-base (fine-tuned)	10	1000	0.630	0.721	0.648	0.513
T5-base (fine-tuned)	100	1000	0.793	0.821	0.798	0.692
T5-base (fine-tuned)	1000	1000	0.833	0.856	0.839	0.750
T5-base (fine-tuned)	10000	1000	0.868	0.878	0.870	0.795
BART-base (fine-tuned)	10	1000	0.590	0.496	0.515	0.381
BART-base (fine-tuned)	100	1000	0.708	0.653	0.665	0.536
BART-base (fine-tuned)	1000	1000	0.786	0.671	0.708	0.585
BART-base (fine-tuned)	10000	1000	0.785	0.717	0.739	0.624

5.2. Performance of Paragraph2RXNRole

Table S7 Performance of Paragraph2Prod

Model	Precision (%)	Recall (%)	F1 (%)	Ratio of Post-Process-free
OPSIN	18.8	5.4	8.4	-
BERT	78.8	56.8	66.0	-
BioBERT	76.4	61.3	68.0	-
ChemBERT	84.6	69.4	76.2	-
GPT-3.5-turbo-0613 (20-shots)	14.7	54.1	23.1	67.5
GPT-4-0613 (20-shots)	9.3	74.8	16.6	96.1
BART-base (fine-tuned)	66.2	42.3	51.6	40.1
T5-base (fine-tuned)	71.3	55.9	62.6	39.3
Llama2-13b-chat (q-lora fine-tuned)	75.0	72.9	73.9	99.5
Llama3-8b-instruct (fine-tuned)	78.3	58.6	67.0	99.7
Mistral-7b-instruct-v0.2 (fine-tuned)	71.7	73.0	72.3	99.2
GPT-3.5-turbo-0613 (fine-tuned)	78.5	75.7	77.1	100.0

Model	Precision (%)	Recall (%)	F1 (%)	Ratio of Post-Process-free (%)
BERT	69.2	69.2	69.2	-
BioBERT	73.3	75.5	74.3	-
ChemBERT	77.0	76.4	76.7	-
ChemRxnBERT	79.3	78.1	78.7	-
GPT-3.5-turbo-0613 (20-shots)	81.1	45.2	58.1	64.9
GPT-4-0613 (20-shots)	59.4	71.7	65.0	94.6
BART-base (fine-tuned)	82.4	55.7	66.5	27.9
T5-base (fine-tuned)	77.6	72.6	74.9	27.9
Llama2-13b-chat (q-lora fine-tuned)	80.1	72.8	76.2	98.2
Llama3-8b-instruct (fine-tuned)	77.0	76.5	76.8	95.5
Mistral-7b-instruct-v0.2 (fine-tuned)	78.0	80.0	79.0	98.2
GPT-3.5-turbo-0613 (fine-tuned)	84.7	81.3	83.0	100.0

Table S8 Performance of Paragraph2Role

5.3. Performance of Paragraph2MOFInfo

Table S9 Mean performance of Levenshtein similarity and Exact match accuracy for extracting paragraphs containing single reactions and multiple reactions respectively by different models (Paragraph2MOFInfo)

Model	Mean Levenshtein Similarity for single reactions	Mean Exact Accuracy for single reactions	Mean Levenshtein Similarity for multi reactions	Mean Exact Accuracy for multi reactions
BART-base (full fine-tuned)	0.8683	0.7488	0.5338	0.3326
T5-base (full fine-tuned)	0.8797	0.7369	0.5736	0.4037
GPT-3.5-turbo-0613 (zero-shot)	0.7887	0.6326	0.656	0.5083
GPT-4-0613 (zero-shot)	0.8418	0.7217	0.6683	0.5337
Llama2-13b-chat (q-lora fine-tuned)	0.8477	0.7544	0.3942	0.3285
Llama3-8b-instruct (full fine-tuned)	0.8976	0.8166	0.7343	0.608
Mistral-7b-instruct-v0.2 (full fine-tuned)	0.8946	0.816	0.7693	0.609
GPT-3.5-turbo-0613 (full fine-tuned)	0.9052	0.8272	0.7844	0.6882

Table S10 Levenshtein similarity for 11 parameters in Paragraph2MOFInfo

Model	BART-base (full fine-tuned)	T5-base (full fine-tuned)	GPT-3.5-turbo (zero-shot)	GPT-4 (zero-shot)	Llama2-13b-chat (q-lora fine-tuned)	Llama3-8b-instruct (full fine-tuned)	Mistral-7b-instruct-v0.2 (full fine-tuned)	GPT-3.5-turbo (full fine-tuned)
Compound Name	0.7688	0.8183	0.7664	0.8411	0.7187	0.8641	0.9014	0.9129
Metal Source	0.8312	0.8286	0.8825	0.8825	0.8056	0.9342	0.9298	0.9124
Metal Amount	0.7856	0.8128	0.7468	0.7753	0.7641	0.8655	0.8705	0.8575
Linker	0.793	0.8269	0.7695	0.7927	0.71	0.8513	0.8598	0.8766
Linker Amount	0.7632	0.7905	0.6608	0.737	0.714	0.8404	0.8231	0.8374
Modulator	0.724	0.7305	0.663	0.771	0.65	0.7762	0.7808	0.8294
Modulator Amount	0.7048	0.7228	0.6652	0.7397	0.673	0.781	0.7948	0.8419
Solvent	0.7733	0.7996	0.7353	0.7678	0.691	0.8371	0.8371	0.8506
Solvent Volume	0.7481	0.7686	0.6838	0.7263	0.7022	0.8178	0.8249	0.8254
Reaction Temperature	0.8741	0.8885	0.8841	0.8882	0.8372	0.945	0.9496	0.9397
Reaction Time	0.8962	0.8768	0.8661	0.8767	0.8544	0.9271	0.9361	0.9521

Table S11 Exact match accuracy for 11 parameters in Paragraph2MOFInfo

Parameter	BART-base (full fine-tuned)	T5-base (full fine-tuned)	GPT-3.5-turbo (zero-shot)	GPT-4 (zero-shot)	Llama2-13b-chat (q-lora fine-tuned)	Llama3-8b-instruct (full fine-tuned)	Mistral-7b-instruct-v0.2 (full fine-tuned)	GPT-3.5-turbo (full fine-tuned)
Compound name	0.4961	0.5299	0.5299	0.613	0.5065	0.5974	0.6312	0.7507
Metal Source	0.7273	0.3117	0.8156	0.8078	0.7325	0.8883	0.8727	0.8468
Metal Amount	0.6546	0.7065	0.5403	0.5948	0.6909	0.7766	0.7792	0.761
Linker	0.6286	0.7117	0.6286	0.6416	0.626	0.7714	0.761	0.8
Linker Amount	0.5403	0.6494	0.4234	0.5247	0.6052	0.7481	0.7143	0.7351
Modulator	0.6494	0.6779	0.5844	0.7221	0.5974	0.7143	0.7143	0.7844
Modulator Amount	0.652	0.6156	0.5714	0.6623	0.5974	0.7221	0.7091	0.761
Solvent	0.6078	0.6649	0.4675	0.613	0.587	0.7039	0.7065	0.7247
Solvent Volume	0.5481	0.6364	0.4234	0.574	0.574	0.6935	0.7065	0.7195
Reaction Temperature	0.7584	0.8597	0.8468	0.8571	0.8156	0.9195	0.9427	0.9169
Reaction Time	0.8416	0.8571	0.7974	0.8286	0.8338	0.8935	0.9065	0.9299

5.4. Performance of Paragraph2NMR

Table S12 Levenshtein similarity of Paragraph2NMR

Model	IUPAC	¹ H NMR text	¹ H NMR conditions	¹ H NMR data	¹³ C NMR text	¹³ C NMR conditions	¹³ C NMR data
GPT-3.5-turbo-0613 (finetuned w/ prompt)	99.1	97.2	98.4	99.5	96.8	98.1	99.7
GPT-3.5-turbo-0613 (finetuned w/o prompt)	98.6	97.9	98.6	99.7	97.6	98.1	99.8
Mistral-7b-instruct-v0.2 (finetuned w/ prompt)	98.2	99.0	99.2	99.4	99.0	98.6	99.8
Mistral-7b-instruct-v0.2 (finetuned w/o prompt)	98.4	99.3	98.8	99.6	98.9	98.5	99.5
Llama3-8b-instruct (finetuned w/ prompt)	97.2	96.6	97.1	97.5	96.4	96.1	96.2
Llama3-8b-instruct (finetuned w/o prompt)	97.8	97.1	97.9	98.8	96.6	97.6	98.8
Llama2-13b-instruct (q-lora finetuned w/ prompt)	96.8	96.3	91.6	96.9	96.8	90.1	96.9
Llama2-13b-instruct (q-lora finetuned w/o prompt)	97.0	96.8	92.1	97.9	97.2	90.3	98.3
T5-base (finetuned)	97.2	92.3	97.6	93.6	86.2	94.0	87.8
Bart-base (finetuned)	95.3	77.2	97.2	73.7	80.4	96.7	81.0
GPT-4-0613 (one-shot)	98.6	96.6	89.6	95.6	96.2	88.0	96.7
GPT-4-0613 (zero-shot)	94.5	97.4	91.8	97.1	97.0	90.3	97.6
GPT-3.5-turbo-0613 (one-shot)	97.6	95.7	94.9	97.4	95.9	96.0	97.9
GPT-3.5-turbo-0613 (zero-shot)	91.6	95.2	90.5	95.4	94.8	90.1	97.4

Table S13 Exact match accuracy of Paragraph2NMR

Model	IUPAC	¹ H NMR text	¹ H NMR conditions	¹ H NMR data	¹³ C NMR text	¹³ C NMR conditions	¹³ C NMR data
GPT-3.5-turbo-0613 (finetuned w/ prompt)	96.0	86.7	93.0	94.0	87.0	92.7	95.7
GPT-3.5-turbo-0613 (finetuned w/o prompt)	96.3	85.3	92.3	94.7	85.7	91.0	95.0
Mistral-7b-instruct-v0.2 (finetuned w/ prompt)	95.7	82.0	92.3	92.3	84.7	90.3	95.3
Mistral-7b-instruct-v0.2 (finetuned w/o prompt)	94.7	84.3	92.7	93.0	86.3	92.7	94.3
Llama3-8b-instruct (finetuned w/ prompt)	89.3	83.7	91.0	90.7	83.0	90.3	89.3
Llama3-8b-instruct (finetuned w/o prompt)	92.7	84.3	91.3	92.7	82.7	91.0	92.3
Llama2-13b-instruct (q-lora finetuned w/ prompt)	91.0	82.3	85.3	90.0	83.0	82.7	89.7
Llama2-13b-instruct (q-lora finetuned w/o prompt)	89.3	82.3	86.3	91.3	86.0	84.3	93.0
T5-base (finetuned)	88.7	5.0	86.0	76.7	4.7	80.3	73.0
Bart-base (finetuned)	64.3	2.0	90.3	9.0	8.0	87.0	13.3
GPT-4-0613 (one-shot)	94.7	84.7	79.3	72.3	82.0	78.3	90.0
GPT-4-0613 (zero-shot)	48.0	78.0	84.3	83.7	75.0	83.3	89.0
GPT-3.5-turbo-0613 (one-shot)	92.0	67.0	85.7	72.3	70.0	88.0	89.7
GPT-3.5-turbo-0613 (zero-shot)	37.4	61.3	76.3	71.7	63.7	75.7	88.7

5.5. Performance of Paragraph2Action

Table S14 Performance of GPT-3.5-Turbo-0126 using In-context Learning and Fine-tuning for Paragraph2Action

Model	Strategy	100% acc	90% acc	75% acc	Modified BLEU score	Levenshtein similarity
GPT-3.5-turbo-0126 (30-shots)		0.1790	0.2840	0.4744	0.5016	0.6742
GPT-3.5-turbo-0126 (60-shots)	In-context	0.1960	0.2983	0.4772	0.5098	0.6663
GPT-3.5-turbo-0126 (60-shots)	Learning	0.2131	0.3153	0.5170	0.4895	0.6943
GPT-3.5-turbo-0126 (120-shots)		0.2273	0.3153	0.5057	0.5440	0.6826
GPT-3.5-turbo-0126 (30-data)		0.3807	0.4659	0.6477	0.6889	0.7548
GPT-3.5-turbo-0126 (60-data)	Fine-tuning	0.2642	0.3580	0.5398	0.5885	0.6765
GPT-3.5-turbo-0126 (90-data)		0.3892	0.4716	0.6648	0.6987	0.7616
GPT-3.5-turbo-0126 (120-data)		0.4204	0.5057	0.6818	0.7224	0.7812

5.6. Performance Trends with Increasing Training Data

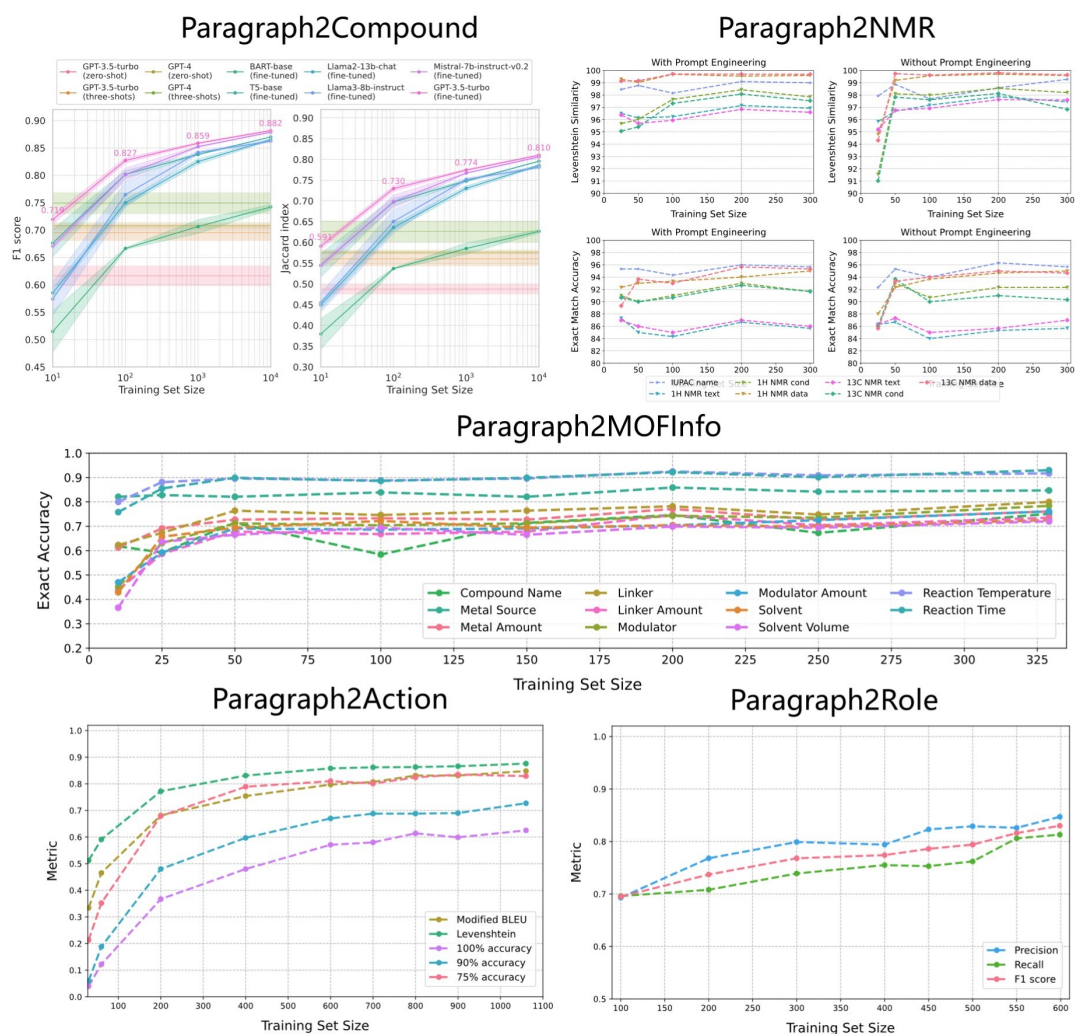


Fig. S11. Performance Improvement with Increasing Training Data

6. Analysis of Error Predictions by Fine-tuned ChatGPT

For each task, we have taken some examples where the prediction does not completely match to the ground truth label.

6.1. Error cases for Paragraph2Compound Task

Error Cases of Paragraph2Compound

Example 1:
4-Amino-6-chloro-5-nitropyrimidine (Boon et al., J. Chem. Soc., 96-102 (1951)) (1.5 g, 8.6 mmol) was added to a solution of sodium (0.23 g, 9.9 mmol) in benzyl alcohol (14 mL). The solution was heated in a 130° C. oil bath for 3.5 h, and was poured into benzene (50 mL). A yellow solid was collected by filtration and washed with benzene. Crystallization from benzene/ether afforded an analytically pure sample of 3a: yield, 0.71 g (34%); mp 149-150° C.; UV (pH 1) λ_{max} 284 nm ($\epsilon=0.368 \times 10^4$), 333 (0.488×10^4); (pH 6.9) 284 (0.329×10^4), 336 (0.470×10^4); (pH 13) 290 (0.344×10^4), 333 (0.494×10^4); ¹H NMR δ 5.50 (s, 2H, ArCH₂), 7.33-7.49 (m, 5H, ArH), 8.12-8.24 (br d, 2H, NHa and NHb, exchange with D₂O), 8.24 (s, 1H, H-2); MS (EI) calcd. m/z for C₁₁H₁₀N₄O₃ 246.0752, found 246.0751; Anal. (C₁₁H₁₀N₄O₃) C, H, N.

Ground Truth: ['4-Amino-6-chloro-5-nitropyrimidine', 'sodium', 'benzyl alcohol', 'benzene', 'benzene', 'benzene/ether', '3a']
Prediction: ['4-Amino-6-chloro-5-nitropyrimidine', 'sodium', 'benzyl alcohol', 'benzene', 'benzene', 'benzene/ether']

Example 2:
A solution of 2-methyl-2-thiopseudourea sulfate (6.9 g, 0.025 mol) in 30 ml of sodium hydroxide (4%) was added a solution of 4-methylbenzoyl chloride (3.4 g, 0.022 mol) in ether (10 ml) at room temperature. The reaction mixture was stirred overnight and the precipitated solid was filtered, washed with water, later hexanes and dried under high vacuum. N-(4-methylbenzoyl)-S-methylisothiourea: yield 4.60 g (quantitative); purity 98% (HPLC); ¹H-NMR (CD₃OD) δ 2.4 (s, 3H, CH₃), 2.6 (s, 3H, SMe), 7.2 (d, 2H, ArH), 8.1 (d, 2H, ArH).

Ground Truth: ['2-methyl-2-thiopseudourea sulfate', 'sodium hydroxide', '4-methylbenzoyl chloride', 'ether', 'water', 'hexanes']
Prediction: ['2-methyl-2-thiopseudourea sulfate', 'sodium hydroxide', '4-methylbenzoyl chloride', 'ether', 'water', 'hexanes', 'N-(4-methylbenzoyl)-S-methylisothiourea']

Example 3:
Fermentation of dodecane with strain H534 under the standard fermentation conditions (Example 20) produced approximately 139 g/l within 233 hrs with a substrate conversion efficiency of 32.1%. The final production rate was 0.58 g/l/hr. The product was 82.7% dodecanedioic acid. The remaining product was predominantly adipic acid. With methyl laurate as the substrate, H534 produced 115.3 g/l dicarboxylic acid within 223 hrs with a substrate conversion efficiency of 34.6%. The production rate was 0.49 g/l/hr. The product was 89.1% dodecanedioic acid. The remaining product was predominantly adipic acid.

Ground Truth: ['dodecanedioic acid', 'adipic acid', 'methyl laurate', 'dicarboxylic acid']
Prediction: ['dodecane', 'dodecanedioic acid']

Fig. S12. Example of incorrect predictions for Paragraph2Compound task.

6.2. Error cases for Paragraph2RXNRole Task

Error Cases of Paragraph2RXNRole

Example 1:

Reaction of diphenylacetylene with complex 19A led to only cycloheptadienone 23A in 30 % yield ; with (phenylcyclopropyl)- carbene complex 19B , cycloheptadienone 25 was produced in 53 % yield.

Ground Truth: Reaction of diphenylacetylene with complex 19A led to only <Prod*cycloheptadienone*Prod> <Prod*23A*Prod> in 30 % yield ; with (phenylcyclopropyl)- carbene complex 19B , cycloheptadienone <Prod*25*Prod> was produced in 53 % yield.

Prediction: Reaction of diphenylacetylene with complex 19A led to only cycloheptadienone <Prod*23A*Prod> in 30 % yield ; with (phenylcyclopropyl)- carbene complex 19B , cycloheptadienone <Prod*25*Prod> was produced in 53 % yield.

Example 2:

Treatment of C with a suitable base leads to intramolecular displacement of the mesylate to give cyclic derivatives , which on deprotection can give 1 and 2.

Ground Truth: Treatment of C with a suitable base leads to intramolecular displacement of the mesylate to give cyclic derivatives , which on deprotection can give 1 and 2

Prediction: Treatment of C with a suitable base leads to intramolecular displacement of the mesylate to give cyclic derivatives , which on deprotection can give <Prod*1*Prod> and <Prod*2.*Prod>

Example 3:

The preparation (Scheme 2) of the title compound <Prod*4*Prod> through Br / Li interchange of the known18 α -bromoalkene 6 in Et₂O or t-BuOMe with n- butyllithium (n-BuLi) in hexane or cyclopentane was already described.¹⁹ However , this method was unprofitable in THF as the solvent20 since 4 coupled quickly with its coproduct 1- bromobutane (n-BuBr) to give 8 and LiBr.

Ground Truth: The preparation (Scheme 2) of the title compound <Prod*4*Prod> through <Reaction*Br / Li interchange*Reaction> of the known18 <Reactants* α -bromoalkene*Reactants> <Reactants*6*Reactants> in <Solvent*Et₂O*Solvent> or <Solvent*t-BuOMe*Solvent> with <Catalyst_Reagents*n-butyllithium*Catalyst_Reagents> (n-BuLi) in <Solvent*hexane*Solvent> or <Solvent*cyclopentane*Solvent> was already described.¹⁹ However , this method was unprofitable in THF as the solvent20 since 4 coupled quickly with its coproduct 1- bromobutane (n-BuBr) to give 8 and LiBr.

Prediction: The preparation (Scheme 2) of the title compound <Prod*4*Prod> through <Reaction*Br / Li interchange*Reaction> of the known18 α -bromoalkene <Reactants*6*Reactants> in <Solvent*Et₂O*Solvent> or <Solvent*t-BuOMe*Solvent> with <Reactants*n- butyllithium*Reactants> (<Reactants*n-BuLi*Reactants>) in <Solvent*hexane*Solvent> or <Solvent*cyclopentane*Solvent> was already described.¹⁹ However , this method was unprofitable in THF as the solvent20 since 4 coupled quickly with its coproduct 1- bromobutane (n-BuBr) to give 8 and LiBr.

Fig. S13. Example of incorrect predictions for Paragraph2RXNRole task.

6.3. Error cases for Paragraph2MOFInfo Task

Error Cases of Paragraph2MOFInfo										
<p>Example 1: MIL-140D or ZrO[O2C-C12N2H6Cl2-CO2] was synthesized in a similar manner as MIL-140A with a 23mL Teflon lined steel autoclave starting from 1mmol (340 mg) of 3,3'-dichloro-4,4'-azobenzenedicarboxylic acid (Cl2AzoBDC), 0.5 mmol (117 mg) of ZrCl4 (Alfa Aesar, 99.5+%), 285 μL (299 mg, 5 mmol) of acetic acid (CH3CO2H) and 5mL (4.75 g, 65 mmol) of DMF. The autoclave was placed in an oven at 180 °C for both syntheses and kept at this temperature for 16 hours.</p>										
compound name	metal source	metal amount	linker	linker amount	modulator	modulator amount or volume	solvent	solvent volume	reaction temperature	reaction time
MIL-140D or ZrO[O2C-C12N2H6Cl2-CO2]	ZrCl4	117mg, 0.5 mmol	3,3'-dichloro-4,4'-azobenzenedicarboxylic acid (Cl2AzoBDC)	340 mg, 1 mmol	acetic acid (CH3CO2H)	285 μ L, 299 mg, 5 mmol	DMF	5 mL, 4.75 g, 65 mmol	180 °C	16 hours
MIL-140D or ZrO[O2C-C12N2H6Cl2-CO2]	ZrCl4	0.5 mmol, 117 mg	3,3'-dichloro-4,4'-azobenzenedicarboxylic acid (Cl2AzoBDC)	1 mmol, 340 mg	acetic acid (CH3CO2H)	285 μ L, 299 mg, 5 mmol	DMF	5 mL, 4.75 g, 65 mmol	180 °C	16 hours
<p>Example 2: Synthesis of {[Yb4(L)3(μ3-OH)4(H2O)4](NO3)2·10H2O·12DMF}n (2). Yb(NO3)3·5H2O (44.8 mg, 0.1 mmol), H2L (9.6 mg, 0.03 mmol) and DMF/EtOH/H2O (8 mL, v/v/v: 5/2/1) were loaded into a 15 mL bottle. The bottle was sealed and heated at 90 °C for three days. Colourless block crystals of 2 were obtained in 51% yield based on H2L. Anal. Calcd for C93H149N17O48Yb4: C, 37.67; H, 5.06; N, 8.03%. Found: C, 38.28; H, 4.73; N, 7.73%. IR (KBr, Fig. S9b?): ν (cm⁻¹) 3372 (s), 1650 (s), 1587 (s), 1535 (s), 1400 (s), 1175 (w), 1093 (w), 1004 (w), 853 (m), 775 (s), 698 (m), 664 (w). Synthesis of {[Lu4(L)3(μ3-OH)4(H2O)4](NO3)2·14H2O·12DMF}n (3). Complex 3 was obtained by the same procedure used for the preparation of 2 except that Yb(NO3)2·5H2O was replaced by Lu(NO3)3·H2O (37.8 mg, 0.1 mmol). Colourless block crystals of 3 were obtained in 47% yield based on H2L. Anal. Calcd for C93H157N17O52Lu4: C, 36.68; H, 5.20; N, 7.82%. Found: C, 36.81; H, 5.35; N, 7.59%. IR (KBr, Fig. S9c?): ν (cm⁻¹) 3375 (s), 1650 (s), 1587 (s), 1538 (s), 1400 (s), 1176 (w), 1092 (w), 852 (m), 802 (w), 773 (s), 698 (m), 667 (w).</p>										
compound name	metal source	metal amount	linker	linker amount	modulator	modulator amount or volume	solvent	solvent volume	reaction temperature	reaction time
{[Yb4(L)3(μ 3-OH)4(H2O)4](NO3)2·10H2O·12DMF}n	Yb(NO3)3·5H2O	44.8 mg, 0.1 mmol	H2L	9.6 mg, 0.03 mmol	N/A	N/A	DMF/EtOH/H2O (v/v/v: 5/2/1)	8 mL	90 °C	three days
{[Lu4(L)3(μ 3-OH)4(H2O)4](NO3)2·14H2O·12DMF}n	Lu(NO3)3·H2O	37.8 mg, 0.1 mmol	H2L	9.6 mg, 0.03 mmol	N/A	N/A	DMF/EtOH/H2O (v/v/v: 5/2/1)	8 mL	90 °C	three days
{[Yb4(L)3(μ 3-OH)4(H2O)4](NO3)2·10H2O·12DMF}n	Yb(NO3)3·5H2O	44.8 mg, 0.1 mmol	H2L	9.6 mg, 0.03 mmol	N/A	N/A	DMF/EtOH/H2O	8 mL, v/v/v: 5/2/1	90 °C	three days
{[Lu4(L)3(μ 3-OH)4(H2O)4](NO3)2·14H2O·12DMF}n	Lu(NO3)3·H2O	37.8 mg, 0.1 mmol	H2L	9.6 mg, 0.03 mmol	N/A	N/A	DMF/EtOH/H2O	8 mL, v/v/v: 5/2/1	90 °C	three days

Fig. S14. Example of incorrect predictions for Paragraph2MOFInfo task. The cells in green represented the annotated label (ground truth). The cells in blue represented the predictions.

6.4. Error cases for Paragraph2NMR Task

Error Cases of Paragraph2NMR

Example 1:

2.3.4. 7,9-Dibromo-2-methyl-1,10b-dihydro-5H-pyrazolo[1,5-c][1,3]benzoxazin-5-thione (6d) White powder, yield, 26%; m.p. 268–270 °C; 1H NMR (300 MHz, CDCl₃) δ 7.76 (d, J = 1.7 Hz, 1H, Ar[sbond]H), 7.21 (d, J = 0.8 Hz, 1H, Ar[sbond]H), 5.25–5.15 (m, 1H, 10b-H), 3.47 (dd, J = 17.1, 10.0 Hz, 1H, 1-Hb), 3.21 (dd, J = 16.2, 11.9 Hz, 1H, 1-Hb), 2.29 (s, 3H, CH₃). 13C NMR (75 MHz, CDCl₃) δ 175.1 (C[dbond]S), 164.5 (C[dbond]N), 146.6 (C[sbond]O), 136.3, 126.7, 125.1, 118.2, 111.6, 57.5 (C-10b), 42.9 (C-1), 16.9 (CH₃). TOF-HRMS: m/z [M + H]⁺ calcd for C₁₁H₉Br₂N₂O₂S: 376.8776; found: 376.8777.

Ground Truth:

{¹IUPAC:"7,9-Dibromo-2-methyl-1,10b-dihydro-5H-pyrazolo[1,5-c][1,3]benzoxazin-5-thione",¹H NMR text:"1H NMR (300 MHz, CDCl₃) δ 7.76 (d, J = 1.7 Hz, 1H, Ar[sbond]H), 7.21 (d, J = 0.8 Hz, 1H, Ar[sbond]H), 5.25–5.15 (m, 1H, 10b-H), 3.47 (dd, J = 17.1, 10.0 Hz, 1H, 1-Hb), 3.21 (dd, J = 16.2, 11.9 Hz, 1H, 1-Hb), 2.29 (s, 3H, CH₃)",¹³C NMR text:"13C NMR (75 MHz, CDCl₃) δ 175.1 (C[dbond]S), 164.5 (C[dbond]N), 146.6 (C[sbond]O), 136.3, 126.7, 125.1, 118.2, 111.6, 57.5 (C-10b), 42.9 (C-1), 16.9 (CH₃)",¹H NMR conditions:"300 MHz, CDCl₃",¹H NMR data:"7.76, 7.21, 5.25–5.15, 3.47, 3.21, 2.29",¹³C NMR conditions:"75 MHz, CDCl₃",¹³C NMR data:"175.1, 164.5, 146.6, 136.3, 126.7, 125.1, 118.2, 111.6, 57.5, 42.9, 16.9"}

Prediction:

{¹IUPAC:"7,9-Dibromo-2-methyl-1,10b-dihydro-5H-pyrazolo[1,5-c][1,3]benzoxazin-5-thione",¹H NMR text:"1H NMR (300 MHz, CDCl₃) δ 7.76 (d, J = 1.7 Hz, 1H, ArH), 7.21 (d, J = 0.8 Hz, 1H, ArH), 5.25–5.15 (m, 1H, 10b-H), 3.47 (dd, J = 17.1, 10.0 Hz, 1H, 1-Hb), 3.21 (dd, J = 16.2, 11.9 Hz, 1H, 1-Hb), 2.29 (s, 3H, CH₃)",¹³C NMR text:"13C NMR (75 MHz, CDCl₃) δ 175.1 (CS), 164.5 (CN), 146.6 (CO), 136.3, 126.7, 125.1, 118.2, 111.6, 57.5 (C-10b), 42.9 (C-1), 16.9 (CH₃)",¹H NMR conditions:"300 MHz, CDCl₃",¹H NMR data:"7.76, 7.21, 5.25–5.15, 3.47, 3.21, 2.29",¹³C NMR conditions:"75 MHz, CDCl₃",¹³C NMR data:"175.1, 164.5, 146.6, 136.3, 126.7, 125.1, 118.2, 111.6, 57.5, 42.9, 16.9"}

Example 2:

7-Isopropoxy-6-methoxy-2H-chromen-2-one (11) White needle crystals; 58.73% yield; m.p. 108.4–108.5 °C; IR ν_{max} (KBr) cm⁻¹: 3098, 2979, 2939, 1706, 1613, 1560, 1513, 1464, 1424, 1383, 1269, 1245, 1148, 1025, 846. 1H-NMR (DMSO-d₆, δ ppm): 1.31 (d, 6H, J = 8 Hz, CH₃), 3.80 (s, 3H, OCH₃), 4.71–4.80 (m, 1H, CH), 6.29 (d, 1H, J = 8 Hz, C3-H), 7.08 (s, 1H, C8-H), 7.25 (s, 1H, C5-H), 7.95 (d, 1H, J = 8 Hz, C4-H) (Figure S5). 13C-NMR (DMSO-d₆, δ ppm): 22.03, 56.24, 71.15, 102.02, 109.63, 111.48, 112.96, 144.78, 146.91, 149.84, 151.03, 161.09 (Figure S6). MS (m/z): [M]⁺ 234. Anal. Calcd. for C₁₃H₁₄O₄: C, 66.67%; H, 6.02%. Found: C, 66.82%; H, 6.21%.

Ground Truth:

{¹IUPAC:"7-Isopropoxy-6-methoxy-2H-chromen-2-one",¹H NMR text:"1H-NMR (DMSO-d₆, δ ppm): 1.31 (d, 6H, J = 8 Hz, CH₃), 3.80 (s, 3H, OCH₃), 4.71–4.80 (m, 1H, CH), 6.29 (d, 1H, J = 8 Hz, C3-H), 7.08 (s, 1H, C8-H), 7.25 (s, 1H, C5-H), 7.95 (d, 1H, J = 8 Hz, C4-H) (Figure S5)",¹³C NMR text:"13C-NMR (DMSO-d₆, δ ppm): 22.03, 56.24, 71.15, 102.02, 109.63, 111.48, 112.96, 144.78, 146.91, 149.84, 151.03, 161.09 (Figure S6)",¹H NMR conditions:"DMSO-d₆, δ ppm",¹H NMR data:"1.31, 3.80, 4.71–4.80, 6.29, 7.08, 7.25, 7.95",¹³C NMR conditions:"DMSO-d₆, δ ppm",¹³C NMR data:"22.03, 56.24, 71.15, 102.02, 109.63, 111.48, 112.96, 144.78, 146.91, 149.84, 151.03, 161.09"}

Prediction:

{¹IUPAC:"7-Isopropoxy-6-methoxy-2H-chromen-2-one",¹H NMR text:"1H-NMR (DMSO-d₆, δ ppm): 1.31 (d, 6H, J = 8 Hz, CH₃), 3.80 (s, 3H, OCH₃), 4.71–4.80 (m, 1H, CH), 6.29 (d, 1H, J = 8 Hz, C3-H), 7.08 (s, 1H, C8-H), 7.25 (s, 1H, C5-H), 7.95 (d, 1H, J = 8 Hz, C4-H)",¹³C NMR text:"13C-NMR (DMSO-d₆, δ ppm): 22.03, 56.24, 71.15, 102.02, 109.63, 111.48, 112.96, 144.78, 146.91, 149.84, 151.03, 161.09",¹H NMR conditions:"DMSO-d₆, δ ppm",¹H NMR data:"1.31, 3.80, 4.71–4.80, 6.29, 7.08, 7.25, 7.95",¹³C NMR conditions:"DMSO-d₆, δ ppm",¹³C NMR data:"22.03, 56.24, 71.15, 102.02, 109.63, 111.48, 112.96, 144.78, 146.91, 149.84, 151.03, 161.09"}

Fig. S15. Example of incorrect predictions for Paragraph2NMR task.

6.5. Error cases for Paragraph2Action Task

Error Cases of Paragraph2Action

Example 1:
A solution was prepared by dissolving 4-(1-{2-fluoro-1-[(3S)-pyrrolidin-3-yl]ethyl}-1H-pyrazol-4-yl)-7-[[2-(trimethylsilyl)ethoxy]methyl]-7H-pyrrolo[2,3-d]pyrimidine (from Example 70, Step 7; 25 mg, 0.058 mmol) and DIPEA (2.0E1 μ L, 0.12 mmol) in NMP (0.2 mL).

Ground Truth:

- ADD 4-(1-{2-fluoro-1-[(3S)-pyrrolidin-3-yl]ethyl}-1H-pyrazol-4-yl)-7-[[2-(trimethylsilyl)ethoxy]methyl]-7H-pyrrolo[2,3-d]pyrimidine (25 mg, 0.058 mmol);
- ADD DIPEA (2.0E1 μ L, 0.12 mmol);
- ADD NMP (0.2 mL).

Prediction:

- MAKESOLUTION with 4-(1-{2-fluoro-1-[(3S)-pyrrolidin-3-yl]ethyl}-1H-pyrazol-4-yl)-7-[[2-(trimethylsilyl)ethoxy]methyl]-7H-pyrrolo[2,3-d]pyrimidine (25 mg, 0.058 mmol) and DIPEA (2.0E1 μ L, 0.12 mmol) and NMP (0.2 mL)

Example 2:
To a stirred and cooled (ice-bath) mixture of 13.0 g of 3,4-difluorophenol and 200 ml of N,N-dimethylformamide, there were added 4.8 g of a dispersion of sodium hydride in mineral oil (50%) and, after stirring for 1 hour, a solution of 22.8 g of (S)-(oxiranylmethyl) 4-methylbenzenesulfonate (ester) in some N,N-dimethylformamide.

Ground Truth:

- MAKESOLUTION with 3,4-difluorophenol (13.0 g) and N,N-dimethylformamide (200 ml);
- ADD SLN;
- ADD sodium hydride in mineral oil (50%) (4.8 g) at 0 °C;
- STIR for 1 hour;
- MAKESOLUTION with (S)-(oxiranylmethyl) 4-methylbenzenesulfonate (ester) (22.8 g) and N,N-dimethylformamide;
- ADD SLN.

Predicted Result:

- MAKESOLUTION with 3,4-difluorophenol (13.0 g) and N,N-dimethylformamide (200 ml);
- ADD SLN;
- ADD dispersion of sodium hydride in mineral oil (50%) (4.8 g) at 0 °C;
- STIR for 1 hour;
- MAKESOLUTION with (S)-(oxiranylmethyl) 4-methylbenzenesulfonate (22.8 g) and N,N-dimethylformamide;
- ADD SLN.

Example 3:
A mixture of 0.8 g of 2-[[[1-(1-carbamoyl-1,2-dimethylpropyl)amino]methyl]-5-ethylnicotinic acid (2.7 mmol) and 0.56 g sodium acetate (6.8 mmol) in 10 mL acetic acid is warmed until homogeneous and cooled to room temperature.

Ground Truth:

- INVALIDACTION.

Predicted Result:

- MAKESOLUTION with 2-[[[1-(1-carbamoyl-1,2-dimethylpropyl)amino]methyl]-5-ethylnicotinic acid (0.8 g, 2.7 mmol) and sodium acetate (0.56 g, 6.8 mmol) and acetic acid (10 mL);
- ADD SLN;
- SETTEMPERATURE room temperature.

Fig. S16. Example of incorrect predictions for Paragraph2Action task.

7. Reference

1. D. Lowe, Chemical reactions from US patents (1976-Sep2016), https://figshare.com/articles/dataset/Chemical_reactions_from_US_patents_1976-Sep2016_/5104873, (accessed August 29, 2023, DOI: 10.6084/m9.figshare.5104873.v1).
2. J. Guo, A. S. Ibanez-Lopez, H. Gao, V. Quach, C. W. Coley, K. F. Jensen and R. Barzilay, Automated chemical reaction extraction from scientific literature, *J. Chem. Inf. Model.*, 2021, **62**, 2035-2045.
3. 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.*, 2023, **145**, 18048-18062.
4. A. C. Vaucher, F. Zipoli, J. Geluykens, V. H. Nair, P. Schwaller and T. Laino, Automated extraction of chemical synthesis actions from experimental procedures, *Nat. Comm.*, 2020, **11**, 3601.