Zero-Shot Information Extraction via Chatting with ChatGPT

Xiang Wei[1], **Xingyu Cui**[1], **Ning Cheng**[1], **Xiaobin Wang**[2], **Xin Zhang**, **Shen Huang**[2], **Pengjun Xie**[2], **Jinan Xu**[1], **Yufeng Chen**[1], **Meishan Zhang**, **Yong Jiang**[2], **and Wenjuan Han**[1]

[1] Beijing Jiaotong University, Beijing, China

[2] DAMO Academy, Alibaba Group, China

Abstract

Zero-shot information extraction (IE) aims to build IE systems from the unannotated text. It is challenging due to involving little human intervention. Challenging but worthwhile, zero-shot IE reduces the time and effort that data labeling takes. Recent efforts on large language models (LLMs, *e.g.*, GPT-3, ChatGPT) show promising performance on zero-shot settings, thus inspiring us to explore prompt-based methods. In this work, we ask whether strong IE models can be constructed by directly prompting LLMs. Specifically, we transform the zero-shot IE task into a multi-turn question-answering problem with a two-stage framework (ChatIE). With the power of ChatGPT, we extensively evaluate our framework on three IE tasks: entity-relation triple extract, named entity recognition, and event extraction. Empirical results on six datasets across two languages show that ChatIE achieves impressive performance and even surpasses some full-shot models on several datasets (*e.g.*, NYT11-HRL). We believe that our work could shed light on building IE models with limited resources. [1]

1 Introduction

Information extraction aims to extract structured information from unstructured text into structured data formats, including tasks such as entity-relation triple extract (RE), named entity recognition (NER), event extraction (EE) (Tjong Kim Sang, 2002; Ratinov and Roth, 2009; Wei et al., 2020; Zheng et al., 2021; Li et al., 2020a), *etc*. It is a fundamental and crucial task in natural language processing (Sarawagi et al., 2008). Working with an enormous amount of labeling data is always hectic, labor-intensive, and time-consuming. Hence, many organizations and companies rely on IE techniques to automate manual work with zero/few-shot methods, *e.g.*, clinical IE (Agrawal et al., 2022).

Recent works (Agrawal et al., 2022; Jeblick et al., 2022; Zhang et al., 2022) on large-scale pre-trained language models (LLM), such as GPT-3 (Brown et al., 2020), InstructGPT (Ouyang et al., 2022) and ChatGPT [2], suggest that LLMs perform well in various downstream tasks even without tuning the parameters but only with a few examples as instructions. Hence, it is a timing question: Is it feasible to prompt LLMs to do zero-shot IE tasks under a unified framework? It is challenging because the structured data containing multiple dependent elements are difficult to extract through one-time prediction, especially for some complex tasks like RE. Previous works decompose these complex tasks into different parts and train several modules to solve each part. For example, in the RE task, the pipeline method PURE (Zhong and Chen, 2021) first identifies two entities and then predicts the relations between them. However, supervision from labeled data is required in this model. Additionally, Li et al. (2019b) regard RE as a question-answering process by first extracting subjects and then objects according to the relation templates.

Based on these clues, in this paper, we turn to ChatGPT and hypothesize that ChatGPT is born with the right abilities to deposit a unified zero-shot IE model in an interactive mode. More specifically, we propose ChatIE by transforming the zero-shot IE task into a multi-turn question-answering problem with a two-stage framework. In the first stage, we aim to find out the corresponding element types that may exist in a sentence. Then in the second stage, we perform a chained information extraction to each element type from Stage I. Each stage is implemented with a multi-turn QA process. In each turn, we construct prompts based on designed templates and previously extracted information as input to ask ChatGPT. Finally, we compose the results of each turn into structured data. We conduct extensive experiments on IE, NER, and EE

[1] https://github.com/cocacola-lab/ChatIE

[2] https://openai.com/blog/chatgpt

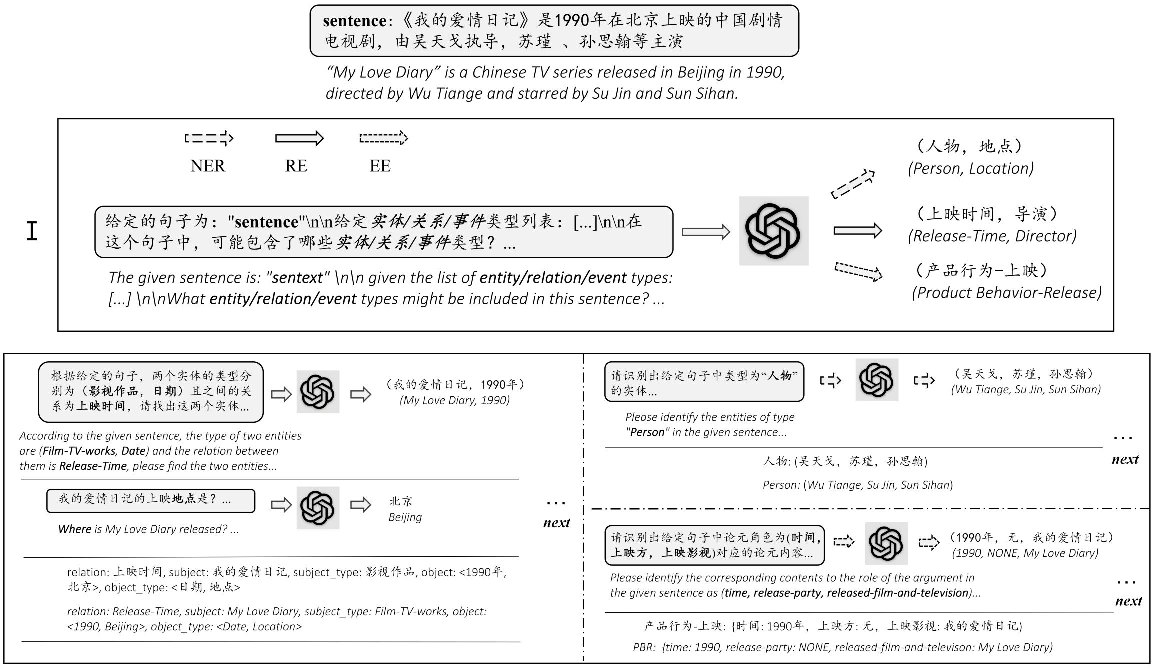

Figure 1: Illustration for the framework. For convenience, we use the samples of DuIE2.0 as examples of three tasks to show.

tasks, including six datasets across two languages: English and Chinese. Empirical results show that while vanilla ChatGPT without using ChatIE fails in solving IE with original task instruction, our proposed two-stage framework instantiated on ChatGPT succeeds when the IE task is decomposed into multiple simpler and easier sub-tasks. Surprisingly, ChatIE achieves impressive performance and even surpasses some full-shot models on several datasets (*e.g.*, NYT11-HRL (Takanobu et al., 2019)).

2 ChatIE

2.1 Multi-Turn QA framework for zero-shot IE

We decompose the IE task into two stages, each containing several turns of QA, which refer to the dialogue with ChatGPT. In the first stage, we aim to find out the existing types of entities, relations, or events in the sentence respectively in three tasks. In this way, we filter out the element types that do not exist to reduce the search space and computational complexity, conducing to extracting information. Then in the second stage, we further extract relevant information based on the element types extracted in the first stage as well as the corresponding task-specific scheme. The overview of our framework is shown in Figure 1, which we will describe in detail later.

Stage I: For one sample, this stage generally includes only one turn of QA. In order to find the element types presented in the sentence, we first utilize the task-specific TypeQuesTemplates and the list of element types to construct the question. Then we combine the question and sentence as input to ChatGPT. To facilitate answer extraction, we ask the system to reply in the list form. If the sentence does not contain any element types, the system will generate a response with NONE Token.

Stage II: This stage generally includes multiple QA turns. In advance, we design a series of specific ChainExtractionTemplates for element types according to the scheme of the task. The ChainExtractionTemplates define a chain of question templates and the length of the chain is usually one. But for complicated schemes such as complex-object value extraction in entity-relation triple extraction, the length of the chain is greater than one. At this point, the extraction of an element may depend on another previous element, so we call it chained templates. We perform multi turns QA in the order of previously extracted element types as well as the order of ChainExtractionTemplates. To generate a question, we need to retrieve the template with the element type and fill the corresponding slots if necessary. Then we access ChatGPT and get a response. Finally, we compose structured information based on the elements extracted in each turn. Similarly, for the convenience of answer extraction, we ask the system to reply in table form. If nothing is extracted, the system will generate a response with NONE token.

2.2 Applying the Framework to IE tasks

After curating the unified framework, ChatIE, we'll then start applying the framework to information extraction tasks, to process and build models for each task.

2.2.1 Entity-Relation Triple Extraction

Given a sentence x and question prompt q, the model is desired to predict triples $T(x) = \{(s_1, r_1, o_1), \cdots, (s_n, r_n, o_n)\}$, where $type((s_i, r_i, o_i)) \in TT$, a list of potential triple types. Formally for an output triple (s, r, o), we can express the process as:

$$p((s,r,o)|x,q) = \underbrace{p(r|x,q_1)}_{\textit{Stage I}} \underbrace{p((s,o)|q_r) \overbrace{\cdots\cdots}^{\textit{complex object}}}_{\textit{Stage II}} \quad (1)$$

Where q_1 is the question generated using relation types list R and the corresponding template in Stage I. And q_r is the question generated using the template related to the previously extracted relation type in Stage II. It is worth noting that we omit x in Stage II, because ChatGPT can record the relevant information of each turn QA. In addition, we need further several turns QA for samples with complex-object values. The complex-object value refers to an object with multiple attributes.

2.2.2 Named Entity Recognition

For the NER task, the first stage is to filter out the existing entity types in the sentence given the desired type list. Once we get the entity types, we can construct the input for the second stage accordingly. In the second stage, each turn aims to extract the entities of one type. So the number of turns in Stage II is up to the number of entities obtained in Stage I, and Stage II is omitted if the first stage gets no types at all. In the experiment, the BIO annotation is not considered since it's kind of hard for ChatGPT. Besides, because the type of entities is few (3 types in conllpp, 4 types in msra) in the datasets, the first stage is skipped in the actual experiment, asking for every type of entity in the second stage.

2.2.3 Event Extraction

ChatIE divides the zero-shot EE task into two subtasks: event classification and argument extraction, solved with two stages in a pipelined fashion. The first stage is designed for event classification, formalized as a text classification problem getting event types from a given text. The second stage is then devoted to argument extraction, formalized as an extractive machine read comprehension (MRC) problem that identifies arguments of specific roles associated with predicted event types from the Stage I.

3 Experiment

3.1 Setup

Datasets

RE. NYT11-HRL (Takanobu et al., 2019) is a preprocessed version of NYT11 (Riedel et al., 2010; Hoffmann et al., 2011) and contains 12 predefined relation types. DuIE2.0 (Li et al., 2019a) is the industry's largest schema-based Chinese RE dataset and contains 48 predefined relation types. Some of the objects in the triples have multiple attributes, called complex-object values.

NER. The conllpp (Wang et al., 2019) dataset is a modified version of the conll2003 (Tjong Kim Sang and De Meulder, 2003) and contains 4 entity types. MSRA (Levow, 2006) is a Chinese named entity recognition dataset for the news field and contains 3 entity types.

EE. DuEE1.0 (Li et al., 2020b) is a Chinese event extraction dataset released by Baidu, which contains 65 event types. The ACE05 [3] corpus provides event annotations in document and sentence levels from a variety of domains such as newswires and online forums.

Evaluation Metrics

RE. We report the standard micro F1 measure and adopt two evaluate metrics: 1) *border* evaluation (BE): an extracted relation triple *(subject, relation, object)* is considered as correct if the whole entity span of both subject and object and relation are all correct. 2) *strict* evaluation (SE): in addition to what is required in the *border* evaluation, the type of both subject and object also must be correct. We use BE on NYT11-HRL because there is no annotation of entity types and use SE on DuIE2.0.

[3] https://catalog.ldc.upenn.edu/LDC2006T06

	RE						NER						EE					
	DuIE2.0			NYT11-HRL			MSRA			collnpp			DuEE1.0			ACE05		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1
fs-1	0.0	0.0	0.0	0.0	0.0	0.0	14.7	7.9	9.7	2.71	17.2	4.66	0.4	0.2	0.3	0.0	0.0	0.0
fs-5	0.0	0.0	0.0	0.0	0.0	0.0	34.5	10.3	15.5	2.53	16.65	4.38	0.2	0.6	0.3	0.0	0.0	0.0
fs-10	16.5	0.1	0.2	0.0	0.0	0.0	60.0	30.9	40.6	2.49	18.54	4.38	2.1	0.7	1.0	0.0	0.0	0.0
fs-20	41.4	0.4	0.8	3.4	2.7	0.5	63.4	44.8	52.5	2.48	19.36	4.41	1.7	0.8	1.1	4.6	0.1	0.2
fs-50	45.7	2.5	4.7	11.7	1.9	3.3	71.6	62.4	66.6	41.94	11.55	8.93	3.2	8.5	4.6	6.7	1.6	2.6
fs-100	50.8	7.2	12.0	34.8	6.2	10.6	81.3	76.1	78.6	50.26	24.97	32.89	8.7	12.0	10.1	8.0	4.9	6.0
full-shot	68.9	72.2	70.5	47.9	55.1	51.3	96.33	95.63	95.98	94.18	94.61	94.39	50.9	42.8	46.5	45.3	54.3	49.4
FCM	-	-	-	43.2	29.4	35.0	-	-	-	-	-	-	-	-	-	-	-	-
MultiR	-	-	-	32.8	30.6	31.7	-	-	-	-	-	-	-	-	-	-	-	-
single	17.8	7.7	10.7	10.8	5.7	7.4	56.3	**57.3**	56.8	61.4	43.0	50.6	61.7	77.5	68.7	18.2	23.9	20.7
ChatIE	**74.6**	**67.5**	**70.9**	**30.6**	**48.4**	**37.5**	**58.4**	57.0	**57.7**	**62.3**	**55.0**	**58.4**	**66.5**	**78.5**	**72.0**	**25.3**	**35.5**	**29.5**

Table 1: F1 score on six datasets over two languages.

NER. We only consider the complete matching and use the micro F1 to evaluate NER task. Only when both the border and the type of the predicted entity and the true entity are the same will we regard it as a correct prediction.

EE. We adopt the different evaluation metrics on the DuEE1.0 dataset and ACE05 dataset. For the DuEE1.0 dataset, F-measure (F1[4]) is scored according to the word-level matching. For the ACE05 dataset, the predicted argument results are matched with the manually marked argument results at the entity level and evaluated by the micro F1.

3.2 Main Results

We summarize the main results in Table 1. We observe that while vanilla ChatGPT (Row **single**, ChatGPT using a single-turn QA instead of ChatIE) performs poorly in solving IE, our proposed two-stage framework based on ChatGPT (Row **ChatIE**) succeeds. ChatIE generally improves performance over six widely used IE datasets by 18.98% points significantly on average.

Notably, the gains become more significant compared with few-shot approaches. For each few-shot experiment, we randomly select 3 sets of the training data, and train 3 times on each set to get an average result. The baselines are PaddleNLP LIC2021 IE [5] and CaseRel (Wei et al., 2020) for RE, AdaSeq Bert-CRF [6] for NER, PaddleNLP LIC2021 EE [7] and Text2event (Lu et al., 2021) for EE. ChatIE is comparable to fs-20 on MSRA, or outperforms fs-100 on NYT11-HRL, collnpp and ACE05, or even surpasses the full-shot on DuIE2.0 and DuEE1.0 in terms of performance.

More surprisingly, compared with two supervised models FCM (Gormley et al., 2015) and MultiR (Hoffmann et al., 2011) on NYT11-HRL, ChatIE surpassed them by 2.5% and 5.8% respectively. Supervised learning models are computationally-intensive and require high-quality labeled data. Additionally, for each task, an individual model is trained from scratch. In contrast, ChatIE works without any finetuning and training to update parameters. It vastly reduces the computation and time investment. With all these benefits, ChatIE still outperforms these supervised learning.

4 Case Study

The first sentence "Just as the JAMA article was being published, three dozen children began dying of acute renal failure at two hospitals in ***Delhi***, ***India***." is an RE case where the same pair of entities belong to two different types of relations. The triples are *(India, location-contains, Delhi)* and *(Delhi, administration_division-country, India)*. In the first stage, ChatIE detects the two relation types. Then in the second stage, ChatIE further extract the words Delhi and India and confirms which one is the source entity and which is the target. This shows ChatIE's ability to give different labels to the same entity in different relations. It is worth noting that we convert *location-contains* to *location-located_in* to predict in the actual experiment, which means we regard *(Delhi, location-located_in, India)* and *(India, location-contains, Delhi)* as equivalent.

[4] https://github.com/PaddlePaddle/PaddleNLP/tree/develop/examples/information_extraction/DuEE

[5] github.com/PaddlePaddle/PaddleNLP/tree/develop/examples/information_extraction/DuIE

[6] github.com/modelscope/AdaSeq/tree/master/examples/bert_crf

[7] github.com/PaddlePaddle/PaddleNLP/tree/develop/examples/information_extraction/DuEE

The second sentence "Four other ***Google*** executives the chief financial officer, ***George Reyes***; the senior vice president for business operations, ***Shona Brown***; the chief legal officer, ***David Drummond***; and the senior vice president for product management, ***Jonathan Rosenberg*** earned salaries of $ 250,000 each." is an RE example where one relation involves multiple triples. It's hard for many methods to extract but it is accomplished by ChatIE. The extracted triples are *(George Reyes, person-company, Google)*, *(Shona Brown, person-company, Google)*, *(David Drummond, person-company, Google)* and *(Jonathan Rosenberg, person-company, Google)*. ChatIE first filters out the *person-company* and outputs the 4 triples related to the relation at the same time in the second stage.

The third sentence "Score on the first day of the four-day Sheffield Shield match between ***Tasmania*** and ***Victoria*** at Bellerive Oval on Friday." is a NER example with confusing entities. The word "Tasmania" and "Victoria" can be categorized as "LOCATION" types, but are actually team names in this sentence, which are "ORGANIZATION" types. ChatIE can recognize the confusing point, showing its advantage in understanding the sentence and choosing the right word meanings.

The last sentence "***Clinton*** suffered greatly over the ***19 Rangers*** that ***died***, 18 on the ***3rd of October*** and MattReersen (ph) ***three days later***." is an EE example. In the first stage, ChatIE gets the event type when scanning the word "died". Then it goes from this word to catch the victim "19 rangers", further detects the agent "Clinton" before the predicate, and targets on "3rd of October" and "three days later".

5 Vanilla Prompt *vs*. Our Chat-based Prompt

Table 2, 3 and 4 demonstrate the comparison of vanilla prompts and our Chat-based prompts in terms of IE. [8]

6 Related Work

Working with an enormous amount of labeling data is always hectic, labor-intensive, and time-consuming. Hence, researchers focus on zero/few-shot technologies even though IE is challenging in low-resource scenarios, such as few-shot relation classification or extraction (Sainz et al., 2021; Han et al., 2018), few-shot event argument extraction (Sainz et al., 2022a) and few-shot information extraction(Sainz et al., 2022b).

ChatGPT has gained widespread attention recently. Many fields received its impacts and evolving fast, such as Medicine (Jeblick et al., 2022; King, 2022) and Online Exam (Susnjak, 2022). In the NLP community, there are new investigations with ChatGPT in several tasks as well. For example, (Zhang et al., 2022) use ChatGPT achieved state-of-the-art performance on Stance Detection, (Guo et al., 2023) evaluated its helpfulness on question answering, (Jiao et al., 2023) state that it is a good translator for spoken language. We try to dig into its information extraction ability, suggesting a simple zero-shot IE framework.

7 Conclusion

We presented ChatIE, a multi-turn QA framework for zero-shot information extraction based on ChatGPT. Through this interactive mode, ChatIE can decompose complex IE tasks into several parts and compose the results of each turn into a final structured result. We apply this framework to RE, NER, and EE tasks and conduct extensive experiments on six datasets across two languages to validate its effectiveness. Surprisingly, ChatIE achieves impressive performance and even surpasses some full-shot models on several datasets. This work paves the way for a new paradigm for zero-shot IE, where the experts decompose IE task into multiple simpler and easier sub-tasks, define chat-like prompts, and directly runs those specifications without training and finetuning.

[8]The experiments are conducted using the version of ChatGPT prior to January 30, 2023.

References

Monica Agrawal, Stefan Hegselmann, Hunter Lang, Yoon Kim, and David Sontag. 2022. Large language models are zero-shot clinical information extractors. *arXiv preprint arXiv:2205.12689*.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

Matthew R Gormley, Mo Yu, and Mark Dredze. 2015. Improved relation extraction with feature-rich compositional embedding models. In *Proceedings of the*

1	Vanilla Prompt	Chat-based Prompt
STAGE I	***Question:*** Suppose you are an entity-relationship triple extraction model. I'll give you list of head entity types: subject_types, list of tail entity types: object_types, list of relations: relations. Give you a sentence, please extract the subject and object in the sentence based on these three lists, and form a triplet in the form of (subject, relation, object). The given sentence is "Bono said that President Jacques Chirac of France had spoken eloquently of the need to support Africa , though he added that France had not yet come through with the resources ." relations:['location-located_in','administrative_division-country', 'person-place_lived', 'person-company', 'person-nationality', 'company-founders', 'country-administrative_divisions', 'person-children', 'country-capital','deceased_person-place_of_death','neighborhood-neighborhood_of', 'person-place_of_birth'] subject_types: ['organization', 'person', 'location', 'country'] object_types: ['person', 'location', 'country', 'organization', 'city'] In the given sentence, what triples might be contained? Please answer in the form (subject, relation, object): - - - - - - - - - - - - - - - - ***Expected Output:*** [(Jacques Chirac, person-nationality, France)] ***Output:*** []	***Question:*** The given sentence is " Bono said that President Jacques Chirac of France had spoken eloquently of the need to support Africa , though he added that France had not yet come through with the resources ." List of given relations: ['location-located_in','administrative_division-country', 'person-place_lived', 'person-company', 'person-nationality', 'company-founders', 'country-administrative_divisions', 'person-children', 'country-capital','deceased_person-place_of_death','neighborhood-neighborhood_of', 'person-place_of_birth'] What relations in the given list might be included in this given sentence? If not present, answer: none. Respond as a tuple, e.g. (relation 1, relation 2, ...): - - - - - - - - - - - - - - - - ***Expected Output:*** (person-nationality) ***Output:*** (person-nationality)
STAGE II	None	***Question:*** According to the given sentence, the two entities are of type ('person', 'country') and the relation between them is 'person-nationality', find the two entities and list them all by group if there are multiple groups. If not present, answer: none. Respond in the form of a table with two columns and a header of ('person', 'country'): - - - - - - - - - - - - - - - - ***Expected Output:*** (Jacques Chirac, France) ***Output:*** (Jacques Chirac, France)

Table 2: Illustration of vanilla prompts vs our Chat-based prompts in terms of RE. The text highlighted with red represents the prompt template. The text following ***Question:*** represents the prompt that is used in ChatIE.

1	Vanilla Prompt	Chat-based Prompt
STAGE I	***Question:*** I'm going to give you a sentence and ask you to identify the entities and label the entity category. There will only be 4 types of entities: ['LOC', 'MISC', 'ORG', 'PER']. Please present your results in list form. "Japan then laid siege to the Syrian penalty area and had a goal disallowed for offside in the 16th minute." Make the list like: ['entity name1', 'entity type1'],['entity name2', 'entity type2']...... ***Expected Output:*** ["Japan", "LOC"], ["Syrian", "MISC"] ***Output:*** []	***Question:*** Given sentence: "Japan then laid siege to the Syrian penalty area and had a goal disallowed for offside in the 16th minute." The known entity types are: ['LOC', 'MISC', 'ORG', 'PER']. Please answer: What types of entities are included in this sentence? ***Expected Output:*** LOC, MISC ***Output:*** LOC, MISC
STAGE II	None	***Question:*** According to the sentence above, please output the entities of 'LOC' in the form of list like: ['entity name1', 'entity type1'], ['entity name2', 'entity type2']...... According to the sentence above, please output the entities of 'MISC' in the form of list like: ['entity name1', 'entity type1'], ['entity name2', 'entity type2']...... ***Expected Output:*** ["Japan", "LOC"], ["Syrian", "MISC"]***Output:*** ["Japan", "LOC"], ["Syrian", "LOC"]

Table 3: Illustration of vanilla prompts vs our Chat-based prompts in terms of NER. The text highlighted with red represents the prompt template. The text following ***Question:*** represents the prompt that is used in ChatIE.

1	Vanilla Prompt	Chat-based Prompt
Stage I	***Question:*** The list of argument roles corresponding to the event type ‘Contact:Phone-Write’ is [‘Entity’, ‘Time’], The list of argument roles corresponding to the event type ‘Business:Declare-Bankruptcy’ is [‘Org’, ‘Time’, ‘Place’], The list of argument roles corresponding to the event type ‘Justice:Arrest-Jail’ is [‘Person’, ‘Agent’, ‘Crime’, ‘Time’, ‘Place’], The list of argument roles corresponding to the event type ‘Life:Die’ is [‘Agent’, ‘Victim’, ‘Instrument’, ‘Time’, ‘Place’], The list of argument roles corresponding to the event type ‘Personnel:Nominate’ is [‘Person’, ‘Agent’, ‘Position’, ‘Time’, ‘Place’], The list of argument roles corresponding to the event type ‘Conflict:Attack’ is [‘Attacker’, ‘Target’, ‘Instrument’, ‘Time’, ‘Place’], The list of argument roles corresponding to the event type ‘Justice:Sue’ is [‘Plaintiff’, ‘Defendant’, ‘Adjudicator’, ‘Crime’, ‘Time’, ‘Place’], The list of argument roles corresponding to the event type ‘Life:Marry’ is [‘Person’, ‘Time’, ‘Place’]. Give a sentence:"What I do know is Saddam Hussein has butchered over a million of his own citizens.", please extract the event arguments according to the argument roles, and return them in the form of a table.The header of the table is ‘event type’, ‘argument role’, ‘argument content’. If no argument role has a corresponding argument content, the argument content returns "None". - - - - - ***Expected Output:*** "event_type": "Life:Die", "arguments": ["role": "Victim", "argument": "over a million of his own citizens" , { "role": "Agent", "argument": "Saddam Hussein" } ***Output:*** None	***Question:*** The list of event types: [‘Life:Die’, ‘Justice:Arrest-Jail’, ‘Contact:Phone-Write’, ‘Life:Marry’, ‘Conflict:Attack’, ‘Personnel:Nominate’, ‘Business:Declare-Bankruptcy’, ‘Justice:Sue’] Give a sentence: "What I do know is Saddam Hussein has butchered over a million of his own citizens." What types of events are included in this sentence? Please return the most likely answer according to the list of event types above. Require the answer in the form: Event type - - - - - ***Expected Output:*** Life:Die ***Output:*** Life:Die
Stage II	None	***Question:*** The list of argument roles corresponding to the event type ‘Life: Die’ is [‘Agent’, ‘Victim’, ‘Instrument’, ‘Time’, ‘Place’]. please extract the event arguments in the given sentence according to the argument roles, and return them in the form of a table. The header of the table is ‘event type’, ‘argument role’, ‘argument content’. If no argument role has a corresponding argument content, the argument content returns "None". - - - - - ***Expected Output:*** "arguments": ["role": "Victim", "argument": "over a million of his own citizens" , { "role": "Agent", "argument": "Saddam Hussein" } ***Output:*** "arguments": ["role": "Victim", "argument": "over a million of his own citizens" , { "role": "Agent", "argument": "Saddam Hussein" }

Table 4: Illustration of vanilla prompts vs our Chat-based prompts in terms of EE. The text highlighted with red represents the prompt template. The text following ***Question:*** represents the prompt that is used in ChatIE.

2015 Conference on Empirical Methods in Natural Language Processing*, pages 1774–1784.

Biyang Guo, Xin Zhang, Ziyuan Wang, Minqi Jiang, Jinran Nie, Yuxuan Ding, Jianwei Yue, and Yupeng Wu. 2023. How close is chatgpt to human experts? comparison corpus, evaluation, and detection. *arXiv preprint arxiv:2301.07597*.

Xu Han, Hao Zhu, Pengfei Yu, Ziyun Wang, Yuan Yao, Zhiyuan Liu, and Maosong Sun. 2018. FewRel: A large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4803–4809, Brussels, Belgium. Association for Computational Linguistics.

Raphael Hoffmann, Congle Zhang, Xiao Ling, Luke Zettlemoyer, and Daniel S Weld. 2011. Knowledge-based weak supervision for information extraction of overlapping relations. In *Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies*, pages 541–550.

Katharina Jeblick, Balthasar Schachtner, Jakob Dexl, Andreas Mittermeier, Anna Theresa Stüber, Johanna Topalis, Tobias Weber, Philipp Wesp, Bastian Sabel, Jens Ricke, et al. 2022. Chatgpt makes medicine easy to swallow: An exploratory case study on simplified radiology reports. *arXiv preprint arXiv:2212.14882*.

Wenxiang Jiao, Wenxuan Wang, Jen-tse Huang, Xing Wang, and Zhaopeng Tu. 2023. Is chatgpt a good translator? a preliminary study. *arXiv preprint arXiv:2301.08745*.

Michael R King. 2022. The future of ai in medicine: a perspective from a chatbot. *Annals of Biomedical Engineering*, pages 1–5.

Gina-Anne Levow. 2006. The third international Chinese language processing bakeoff: Word segmentation and named entity recognition. In *Proceedings of the Fifth SIGHAN Workshop on Chinese Language Processing*, pages 108–117, Sydney, Australia. Association for Computational Linguistics.

Fayuan Li, Weihua Peng, Yuguang Chen, Quan Wang, Lu Pan, Yajuan Lyu, and Yong Zhu. 2020a. Event extraction as multi-turn question answering. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 829–838.

Shuangjie Li, Wei He, Yabing Shi, Wenbin Jiang, Haijin Liang, Ye Jiang, Yang Zhang, Yajuan Lyu, and Yong Zhu. 2019a. Duie: A large-scale chinese dataset for information extraction. In *CCF International Conference on Natural Language Processing and Chinese Computing*, pages 791–800. Springer.

Xiaoya Li, Fan Yin, Zijun Sun, Xiayu Li, Arianna Yuan, Duo Chai, Mingxin Zhou, and Jiwei Li. 2019b. Entity-relation extraction as multi-turn question answering. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1340–1350.

Xinyu Li, Fayuan Li, Lu Pan, Yuguang Chen, Weihua Peng, Quan Wang, Yajuan Lyu, and Yong Zhu. 2020b. Duee: a large-scale dataset for chinese event extraction in real-world scenarios. In *CCF International Conference on Natural Language Processing and Chinese Computing*, pages 534–545. Springer.

Yaojie Lu, Hongyu Lin, Jin Xu, Xianpei Han, Jialong Tang, Annan Li, Le Sun, Meng Liao, and Shaoyi Chen. 2021. Text2event: Controllable sequence-to-structure generation for end-to-end event extraction. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 2795–2806.

Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *arXiv preprint arXiv:2203.02155*.

Lev Ratinov and Dan Roth. 2009. Design challenges and misconceptions in named entity recognition. In *Proceedings of the thirteenth conference on computational natural language learning (CoNLL-2009)*, pages 147–155.

Sebastian Riedel, Limin Yao, and Andrew McCallum. 2010. Modeling relations and their mentions without labeled text. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 148–163. Springer.

Oscar Sainz, Itziar Gonzalez-Dios, Oier Lopez de Lacalle, Bonan Min, and Eneko Agirre. 2022a. Textual entailment for event argument extraction: Zero- and few-shot with multi-source learning. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 2439–2455, Seattle, United States. Association for Computational Linguistics.

Oscar Sainz, Oier Lopez de Lacalle, Gorka Labaka, Ander Barrena, and Eneko Agirre. 2021. Label verbalization and entailment for effective zero and few-shot relation extraction. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 1199–1212, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Oscar Sainz, Haoling Qiu, Oier Lopez de Lacalle, Eneko Agirre, and Bonan Min. 2022b. ZS4IE: A toolkit for zero-shot information extraction with simple verbalizations. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: System Demonstrations*, pages

27–38, Hybrid: Seattle, Washington + Online. Association for Computational Linguistics.

Sunita Sarawagi et al. 2008. Information extraction. *Foundations and Trends® in Databases*, 1(3):261–377.

Teo Susnjak. 2022. Chatgpt: The end of online exam integrity? *arXiv preprint arXiv:2212.09292*.

Ryuichi Takanobu, Tianyang Zhang, Jiexi Liu, and Minlie Huang. 2019. A hierarchical framework for relation extraction with reinforcement learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 7072–7079.

Erik F. Tjong Kim Sang. 2002. Introduction to the CoNLL-2002 shared task: Language-independent named entity recognition. In *COLING-02: The 6th Conference on Natural Language Learning 2002 (CoNLL-2002)*.

Erik F. Tjong Kim Sang and Fien De Meulder. 2003. Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. In *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*, pages 142–147.

Zihan Wang, Jingbo Shang, Liyuan Liu, Lihao Lu, Jiacheng Liu, and Jiawei Han. 2019. Crossweigh: Training named entity tagger from imperfect annotations. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5154–5163.

Zhepei Wei, Jianlin Su, Yue Wang, Yuan Tian, and Yi Chang. 2020. A novel cascade binary tagging framework for relational triple extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1476–1488.

Bowen Zhang, Daijun Ding, and Liwen Jing. 2022. How would stance detection techniques evolve after the launch of chatgpt? *arXiv preprint arXiv:2212.14548*.

Hengyi Zheng, Rui Wen, Xi Chen, Yifan Yang, Yunyan Zhang, Ziheng Zhang, Ningyu Zhang, Bin Qin, Xu Ming, and Yefeng Zheng. 2021. Prgc: Potential relation and global correspondence based joint relational triple extraction. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 6225–6235.

Zexuan Zhong and Danqi Chen. 2021. A frustratingly easy approach for entity and relation extraction. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 50–61.