



蘇州大學

SOOCHOW UNIVERSITY

工作汇报

刘承伟

2023-05-03

工作

► 文献阅读：

1. 《Is ChatGPT a Good Sentiment Analyzer? A Preliminary Study》 Nanjing University of Science and Technology.
 - Provide a preliminary evaluation of ChatGPT on the understanding of opinions, sentiments, and emotions contained in the text.
2. 《 Investigating Chain-of-thought with ChatGPT for Stance Detection on Social Media》 Shenzhen Technology University.
 - Investigate the efficacy of the chain-of-thought (CoT) prompting strategy when applied to ChatGPT (GPT-3.5) for stance detection tasks



Is ChatGPT a Good Sentiment Analyzer? A Preliminary Study

► Motivation

1. Investigate whether ChatGPT can serve as a universal sentiment analyzer.
2. We are particularly curious how ChatGPT performs on the sentiment analysis tasks, i.e., Can it really understand the opinions, sentiments, and emotions contained in the text?

► Solution

1. We evaluate it in four settings, including **standard evaluation**, **polarity shift evaluation**, **open-domain evaluation**, and **sentiment inference evaluation**.
2. The above evaluation involves 18 benchmark datasets and 5 representative sentiment analysis tasks, and we compare ChatGPT with finetuned BERT and corresponding state-of-the-art (SOTA) models on end-task.



Is ChatGPT a Good Sentiment Analyzer? A Preliminary Study

Set:

Ask ChatGPT to generate the **task instruction** for each task to elicit its ability to the corresponding task. Taking the E2E-ABSA task as an example, our query is:

1. Please give me three concise prompts for eliciting your ability to perform Aspect-Based Sentiment Analysis (i.e., extract the aspect terms and sentiment polarity)
2. Examine the generated three prompts on a small-scale (e.g., **50 examples**) example set driving from the corresponding training set.
3. The final prompts adopted for each task are shown below



Is ChatGPT a Good Sentiment Analyzer? A Preliminary Study

Task	Prompt
SC	Given this text, what is the sentiment conveyed? Is it positive or negative? Text: {sentence}
ABSC	Sentence: {sentence} What is the sentiment polarity of the aspect {aspect} in this sentence?
E2E-ABSA	Given a review, extract the aspect term(s) and determine their corresponding sentiment polarity. Review: {sentence}
ECE	Document: {doc} \n Each line in the above document represents a clause and the number at the beginning of each line indicates the clause ID. Clauses expressing emotions are referred to as “emotion clause” and clauses causing emotions are referred to as “cause clauses”. It has been identified that the clause with ID {emo_id}, {emotion clause} is an emotion clause, and the corresponding emotion keyword is {emotion}. Based on the above information, complete the following tasks: 1. Describe in one sentence the cause of the emotion clause with ID {emo_id}. 2. Based on the result of Task 1, output the ID of the cause clause that best fits the requirements. 3. According to the result of Task 2, match clauses with causality into pairs in the form “(emotion clause ID, cause clause ID)” and output all pairs as a set, such as (1,2),(3,4). Note: the emotion clause and the cause clause may be the same clause, and only the most obvious pairs need to be outputted.
ECPE	Document: {doc} \n Each line in the above document represents a clause and the number at the beginning of each line indicates the clause ID. Clauses expressing emotions are referred to as “emotion clause” and clauses causing emotions are referred to as “cause clauses”. Based on the above information, complete the following tasks: 1. Describe the emotions and their corresponding causes contained in the document in one sentence. 2. Output the ID of the emotion clause in task 1, you only need to find the one with the strongest intensity. 3. For each emotion clause in task 2, find the corresponding cause clause and output the cause clause ID, you only need to find the most suitable one. 4. Match clauses with causality into pairs in the form “(emotion clause ID, cause clause ID)” and output all pairs as a set, such as (1,2),(3,4). Note: the emotion clause and the cause clause may be the same clause, and only the most obvious pairs need to be outputted.



Is ChatGPT a Good Sentiment Analyzer? A Preliminary Study

Task	Datasets	Metric	Fine-tuned		Zero-shot
			BERT	SOTA	ChatGPT
SC	SST-2	Acc	95.47	97.50 ^{α}	93.00
ABSC	14-Restaurant	Acc / F1	83.94 / 75.28	89.54 / 84.86 ^{β}	84.00 / 70.33
	14-Laptop	Acc / F1	77.85 / 73.20	83.70 / 80.13 ^{γ}	76.00 / 67.00
E2E-ABSA	14-Restaurant	F1	77.75	78.68 ^{δ}	69.14
	14-Laptop	F1	66.05	70.32 ^{δ}	49.11

Table 3: Performance comparison among ChatGPT, fine-tuned BERT, and SOTA models on 5 datasets. α , β , γ , and δ denote the model performance reported in Raffel et al. (2020), Zhang et al. (2022c), Li et al. (2022) and Fei et al. (2022) on the corresponding task, respectively. The best results are in **bold**.



Is ChatGPT a Good Sentiment Analyzer? A Preliminary Study

► **ChatGPT vs fine-tuned small language models (Bert)**

1. ChatGPT is on par with fine-tuned small language models in sentiment classification tasks..
2. ChatGPT can match or even surpass fine-tuned BERT in accuracy for SC and ABSC.
3. E2E-ABSA: The poorer performance Laptop is due to the presence of more proprietary terms and specific expressions in this domain, which the language model has encountered less frequently during pretraining.
4. ChatGPT still lags behind to some extent since they have more parameters or complicated model architecture designs to some extent than fine-tuned BERT models.



Is ChatGPT a Good Sentiment Analyzer? A Preliminary Study

Human Evaluation

Model	14-Rest.	14-Laptop
<i>Fully-supervised results</i>		
BERT	77.75	66.05
SOTA ^{δ}	78.68	70.32
<i>Zero-shot results</i>		
ChatGPT (Auto Eval.)	69.14	49.11
ChatGPT (Human Eval.)	83.86	72.77

It can demonstrate that the prediction results of ChatGPT indeed align with human preferences owing to RLHF and the potential of ChatGPT as a universal sentiment analyzer. Case Study.



Is ChatGPT a Good Sentiment Analyzer? A Preliminary Study

Few-shot Prompting Results













ABSC (Zero-shot)  Sentence: I did swap out the hard drive for a Samsung 830 SSD which I highly recommend . What is the sentiment polarity of the aspect hard drive in this sentence?  Sentiment polarity: negative Ground Truth: neutral 	E2E-ABSA (Zero-shot)  Given a review, extract the aspect term(s) and determine their corresponding sentiment polarity. Review: I can say that I am fully satisfied with the performance that the computer has supplied .  Aspect terms: computer performance Sentiment polarity: positive Ground Truth: [[performance, positive]] 
ABSC (Few-shot)  Sentence: The keyboard is too slick . What is the sentiment polarity of the aspect keyboard in this sentence? Label: negative Sentence: Decent wine at reasonable prices . What is the sentiment polarity of the aspect wine in this sentence? Label: positive Sentence: Amazing Quality ! What is the sentiment polarity of the aspect Quality in this sentence? Label: positive Sentence: I did swap out the hard drive for a Samsung 830 SSD which I highly recommend . What is the sentiment polarity of the aspect hard drive in this sentence?  Label: neutral Ground Truth: neutral 	E2E-ABSA (Few-shot)  Given a review, extract the aspect term(s) and determine their corresponding sentiment polarity. Here are some examples: Review: It runs perfectly . Label: [[runs, positive]] Review: The service is awful . Label: [[service, negative]] Review: lots of extra space but the keyboard is ridiculously small . Label: [[space, positive], [keyboard, negative]] Review: I can say that I am fully satisfied with the performance that the computer has supplied .  Label: [[performance, positive]] Ground Truth: [[performance, positive]] 

Figure 3: Case study for ChatGPT on ABSC and E2E-ABSA in zero-shot and few-shot settings. The text in blue, black, green and red denote the given prompts, the examples to be evaluated, the responses of ChatGPT and the ground truths, respectively.

Is ChatGPT a Good Sentiment Analyzer? A Preliminary Study













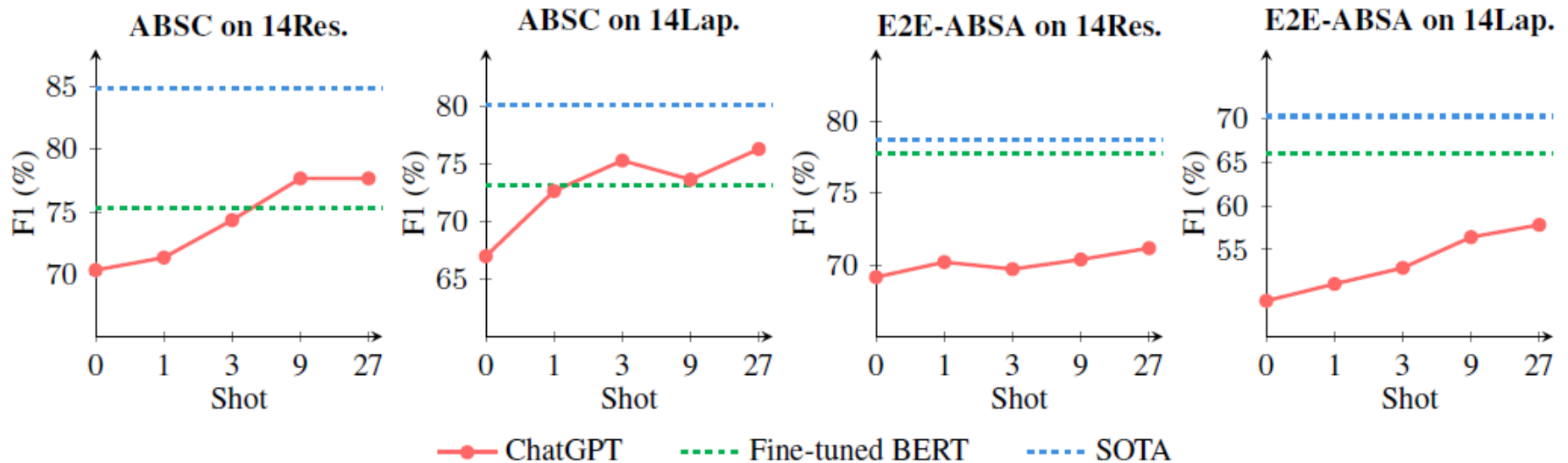
SC (Negation)  Given this text, what is the sentiment conveyed? Is it positive or negative? Text: unfortunately , it 's not silly fun unless you enjoy really bad movies .  The sentiment conveyed in this text is negative. Ground Truth: negative 	SC (Speculation)  Given this text, what is the sentiment conveyed? Is it positive or negative? Text: it can't decide if it wants to be a mystery/thriller , a romance or a comedy .  The sentiment conveyed in the text is negative. Ground Truth: negative 
ABSC (Negation)  Sentence: not so much faster and sleeker looking PUNCT . What is the sentiment polarity of the aspect looking in this sentence?  Label: positive Ground Truth: negative 	ABSC (Speculation)  Sentence: if only it were super fast and had outstanding graphics PUNCT . What is the sentiment polarity of the aspect graphics in this sentence?  Label: positive Ground Truth: negative 

Figure 4: Case study for ChatGPT on SC and ABSC in case of the linguistic phenomena such as negation and speculation. The text in blue, black, green and red denote the given prompts, the examples to be evaluated, the responses of ChatGPT and the ground truths, respectively.

Is ChatGPT a Good Sentiment Analyzer? A Preliminary Study



In light of this, more advanced and effective in-context learning technologies can be employed to improve the few-shot prompting performance, e.g., better selection and formatting of demonstration examples.

Is ChatGPT a Good Sentiment Analyzer? A Preliminary Study

Polarity Shift Evaluation (Sample Figure4)

Task	Shifting Type	Dataset	Fine-tuned	Zero-shot
			BERT	ChatGPT
SC	Negation	SST-2-Neg.	90.68	91.00
	Speculation	SST-2-Spec.	92.05	92.00
ABSC	Negation	14-Res-Neg.	70.93	80.00
			61.90	69.00
		14-Lap-Neg.	60.25	73.00
			53.97	67.00
	Speculation	14-Res-Spec.	64.29	77.00
			60.53	68.00
		14-Lap-Spec.	40.86	47.00
			39.40	47.00

Is ChatGPT a Good Sentiment Analyzer? A Preliminary Study

Shifting Type	Seed Words
Negation	n't, no, not, never, neither, nor, unless, but, however, rather than, not yet, not only, nonetheless, despite, although, even though, in spite of, unlikely
Speculation	if, would, could, should, seems, might, maybe, whether, unless, even if, if only, can't believe, grant that, guessing, suspect, hope, wish, let's probably



Is ChatGPT a Good Sentiment Analyzer? A Preliminary Study

Open Domain Evaluation

Trained on → Evaluated on ↓	Fine-tuned BERT								ChatGPT	
	14-Res.		14-Lap.		The Rest		Domain-Specific		Zero-shot	
	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1
14-Res.	81.11	74.99	84.44	78.76	80.00	69.63	81.11	74.99	83.00	61.00
14-Lap.	77.78	70.60	77.78	72.84	76.67	59.83	77.78	72.84	73.00	53.00
Books	57.78	41.91	57.78	42.84	62.22	46.11	71.11	57.17	60.00	51.00
Clothing	74.44	55.00	76.67	56.21	76.67	61.66	80.00	58.15	70.00	60.00
Hotel	86.67	51.73	86.67	51.29	85.56	50.23	87.78	51.98	87.00	83.50
Device	86.67	85.35	86.67	88.92	94.44	98.11	100.00	100.00	97.00	99.00
Service	71.11	67.91	71.11	67.59	81.11	79.29	74.44	62.69	77.00	65.00
14-Twitter	62.22	54.11	62.22	56.16	70.00	67.83	62.22	60.99	67.00	64.00
Finance	75.56	62.75	74.44	55.59	31.11	31.58	82.22	79.07	87.00	72.00
METS-Cov	53.33	47.06	50.00	37.56	38.89	35.65	61.11	58.53	77.00	56.00
Average	72.67	61.14	72.78	60.78	69.67	59.99	77.78	67.64	77.80	66.45



Is ChatGPT a Good Sentiment Analyzer? A Preliminary Study

Sentiment Inference Evaluation (Sample Figure 6)

emotion cause extraction (ECE)

emotion cause pair extraction (ECPE)

Task	Models	P	R	F1
ECE	<i>Fine-tuned Results</i>			
	PAE-DGL [†]	75.25	64.50	69.46
	<i>Zero-shot Results</i>			
	ChatGPT	75.68	72.41	74.01
ECPE	<i>Fine-tuned Results</i>			
	ECPE-2D [†]	69.89	56.03	62.20
	<i>Zero-shot Results</i>			
	ChatGPT	54.13	50.86	52.44



Is ChatGPT a Good Sentiment Analyzer? A Preliminary Study

ECE

文档：“

- 1, 对于鲁女士遇见老人跌倒立刻出手相助
- 2, 还陪着老人一起等民警的行为
- 3, 周围路过的多位居民都对她表示了赞扬
- 4, 出警的民警也告诉记者
- 5, 跌倒的老人真的需要有人帮助
- 6, 否则可能会发生二次事故
- 7, 因扶老人而被讹的事情是极少数的”

以上文档的每一行代表一个子句，每行开头的数字表示该子句的编号。我们将包含情绪表达的子句称为“情绪子句”，将导致情绪发生的子句称为“原因子句”。我们注意到编号为3的子句“周围路过的多位居民都对她表示了赞扬”是情绪子句，情绪关键词是“赞扬”。基于以上内容完成以下任务。1. 用一句话描述编号为3的情绪子句对应的原因。2. 根据任务1的结果，输出最符合要求的原因子句的编号。3. 根据任务2的结果，将具有因果关系的子句匹配成(情绪子句编号, 原因子句编号)的形式，最后将所有的编号配对按照集合的形式输出，例如{(1,2),(3,4)}。注意：情绪子句和原因子句有可能是同一个子句，且最终只需要输出情绪和原因最明显的配对。



1. “周围路过的多位居民都对她表示了赞扬”这个情绪子句的原因是鲁女士帮助跌倒老人。
2. 根据任务1的结果，原因子句的编号是1。
3. 可以将配对{(1,3)}输出，因为鲁女士帮助跌倒老人导致周围居民对她表示赞扬。

Ground Truth: {(3,1)}



ECE

Document: “

- 1, For Ms. Lu, she immediately helped when she met the fallen old man
- 2, Accompanying the old man and waiting for the behavior of the police
- 3, Many residents passing by praised her
- 4, The police who called out also told reporters
- 5, Old people who fall really need help
- 6, Otherwise, a second accident may occur
- 7, It is very rare to be corrupted for helping the elderly ”

Each line in the above document represents a clause and the number at the beginning of each line indicates the clause ID. Clauses expressing emotions are referred to as “emotion clause” and clauses causing emotions are referred to as “cause clauses”. It has been identified that the clause with ID 3, “Many residents passing by praised her” is an emotion clause, and the corresponding emotion keyword is “praise”. Based on the above information, complete the following tasks: 1. Describe in one sentence the cause of the emotion clause with ID 3; 2. Based on the result of Task 1, output the ID of the cause clause that best fits the requirements. 3. According to the result of Task 2, match clauses with causality into pairs in the form “(emotion clause ID, cause clause ID)” and output all pairs as a set, such as (1,2),(3,4). Note: the emotion clause and the cause clause may be the same clause, and only the most obvious pairs need to be outputted.



1. "Many residents passing by praised her" The reason for this emotion clause is that Ms. Lu helps the fallen old man.
2. According to the result of task 1, the cause clause is numbered 1.
3. The pair {(1,3)} can be output, because Ms. Lu helped the fallen old man and the surrounding residents praised her.

Ground Truth: {(3,1)}



Is ChatGPT a Good Sentiment Analyzer? A Preliminary Study

ECPE

文档：“

- 1, 对于鲁女士遇见老人跌倒立刻出手相助
- 2, 还陪着老人一起等民警的行为
- 3, 周围路过的多位居民都对她表示了赞扬
- 4, 出警的民警也告诉记者
- 5, 跌倒的老人真的需要有人帮助
- 6, 否则可能会发生二次事故
- 7, 因扶老人而被讹的事情是极少数的”

以上文档的每一行代表一个子句，每行开头的数字表示该子句的编号。我们将包含情绪表达的子句称为“情绪子句”，将导致情绪发生的子句称为“原因子句”。基于以上内容，分别完成以下任务。1. 用一句话描述以上文档中包含的情绪以及它对应的原因。2. 输出任务1中情绪子句的编号，你只需要找到情绪强度最强的那一个子句。3. 对于任务2中的每个情绪子句，找到导致其情绪的原因子句，并输出原因子句的编号，你只需要找到最符合要求的那一个子句。4. 将以上具有因果关系的子句匹配成(情绪子句编号, 原因子句编号)的形式，最后将所有的编号配对按照集合的形式输出，例如 {(1,2),(3,4)}。注意：情绪子句和原因子句有可能是同一个子句，且最终只需要输出情绪和原因最明显的配对。



1. 情绪为赞扬，原因为鲁女士及其行为。
2. 情绪子句编号为3。
3. 情绪子句3的导致情绪的原因子句为2，即鲁女士陪着老人等民警的行为。
4. 匹配结果为 {(3,2)}，即3号情绪子句由2号原因子句导致。

Ground Truth: {(3,1)}



ECPE

Document: “

- 1, For Ms. Lu, she immediately helped when she met the fallen old man
- 2, Accompanying the old man and waiting for the behavior of the police
- 3, Many residents passing by praised her
- 4, The police who called out also told reporters
- 5, Old people who fall really need help
- 6, Otherwise, a second accident may occur
- 7, It is very rare to be corrupted for helping the elderly ”

Each line in the above document represents a clause and the number at the beginning of each line indicates the clause ID. Clauses expressing emotions are referred to as “emotion clause” and clauses causing emotions are referred to as “cause clauses”. Based on the above information, complete the following tasks: 1. Describe the emotions and their corresponding causes contained in the document in one sentence. 2. Output the ID of the emotion clause in task 1, you only need to find the one with the strongest intensity. 3. For each emotion clause in task 2, find the corresponding cause clause and output the cause clause ID, you only need to find the most suitable one. 4. Match clauses with causality into pairs in the form “(emotion clause ID, cause clause ID)” and output all pairs as a set, such as (1,2),(3,4). Note: the emotion clause and the cause clause may be the same clause, and only the most obvious pairs need to be outputted.



1. Emotions are praise because of Ms. Lu and her actions.
2. The emotion clause is numbered 3.
3. The cause clause of emotion clause 3 is 2, that is, Ms. Lu accompanied the old man and other police officers.
4. The matching result is { (3,2) }, that is, emotion clause No. 3 is caused by cause clause No. 2.

Ground Truth: {(3,1)}



Investigating Chain-of-thought with ChatGPT for Stance Detection on Social Media

► **Direct question-answering (DQA)** (Thought-inducing)

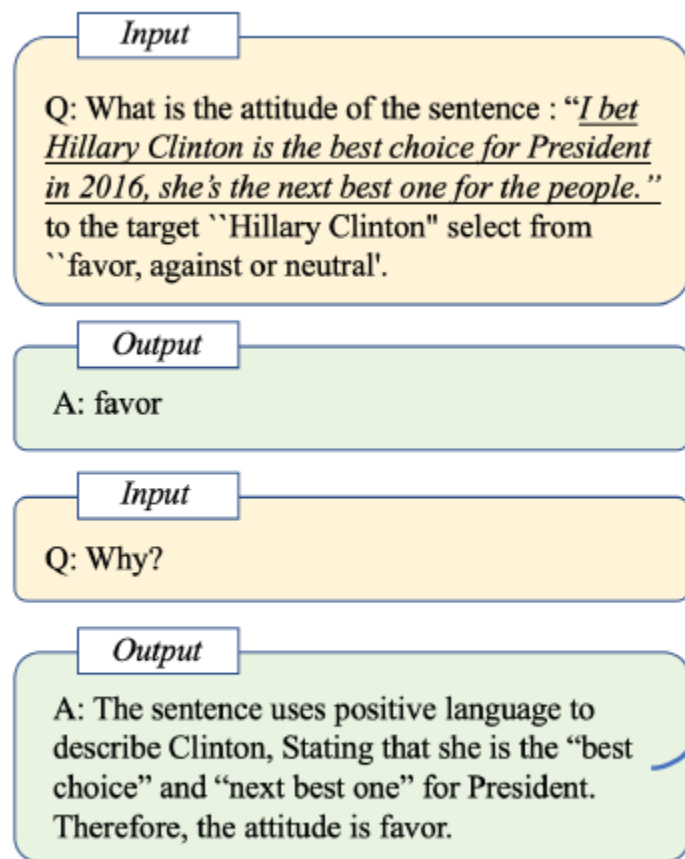
1. We directly ask the ChatGPT model the stance polarity of a certain tweet towards a specific target.

► **step-by-step question-answering (StSQA)** (one-shot)

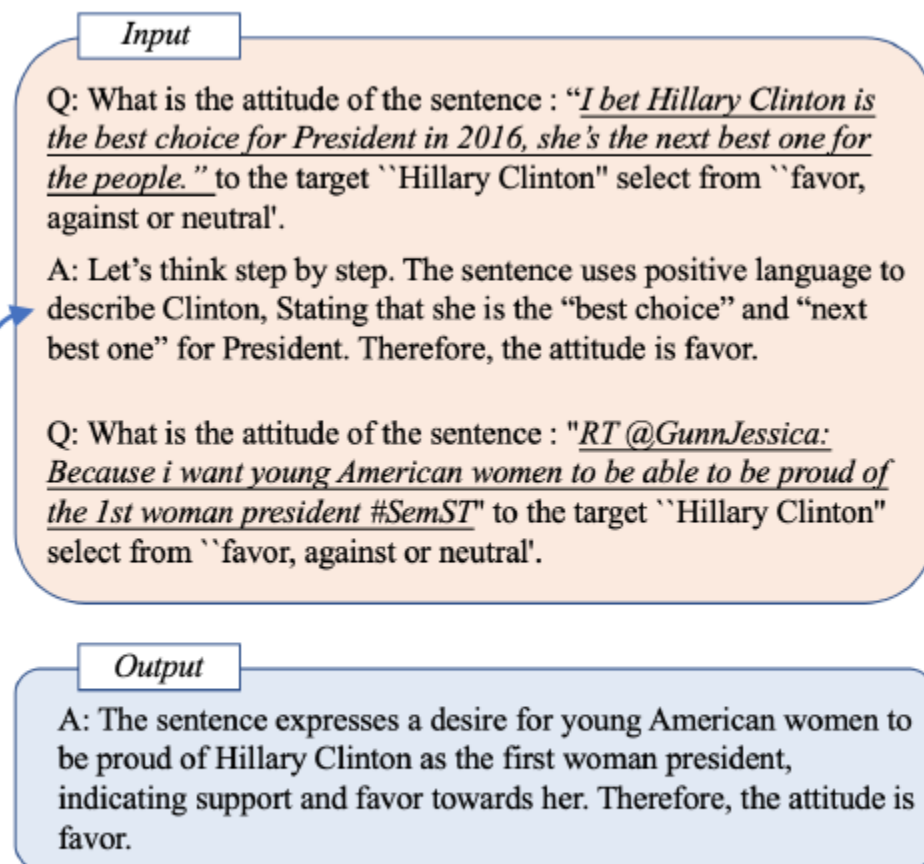
1. Thought-inducing. The first stage is to construct the question-answer pair (QAP).
2. The second stage encompasses inferencing the tweet's stance using the provided QAP.



Investigating Chain-of-thought with ChatGPT for Stance Detection on Social Media



(a) Thought-inducing



(b) 1-shot inference

Investigating Chain-of-thought with ChatGPT for Stance Detection on Social Media

Model	SEM16 (F_{avg})				VAST (F_m)
	HC	FM	LA	DT	
Bicond [†]	32.7	40.6	34.4	30.5	41.0
CrossNet [†]	38.3	41.7	38.5	35.6	45.5
SEKT	50.1	44.2	44.6	46.8	41.1
TPDG [†]	50.9	53.6	46.5	47.3	51.9
Bert_Spc [†]	49.6	41.9	44.8	40.1	65.3
Bert-GCN [†]	50.0	44.3	44.2	42.3	68.6
PT-HCL [†]	54.5	54.6	50.9	50.1	71.6
DQA	78.0	69.0	59.3	71.3	62.3
StSQA	78.9	68.7	61.5	71.6	68.9



Thank you

