

A Details of Few-shot Prompts

This section contains the details of few-shot prompts used by ChatGPT or Vicuna during the construction of the ECR-Chain set and in our experiments. We present templates for each type of prompt along with exemplars.

- **<reasoning> prompt.** Table 6 presents the detailed format of the few-shot prompt used for reasoning along the ECR-Chain, as discussed in Section 3.3. The specific few-shot exemplars used are shown in Figure 5. This table is the complete version of Figure 2. We employ this prompt in the first stage of constructing the ECR-Chain set and in the few-shot learning experiments with ChatGPT. Besides, we use a similar version with Vicuna during supervised learning. The difference is that when using it with Vicuna, the requirements for listing “Appraisals” and “Stimuli” no longer necessitate specifying corresponding “Reaction” and “Appraisal” item indexes (as shown in steps 3 & 4 of the 5-step reasoning process described in Table 6), which is to maintain consistency with the format of the constructed ECR-Chain data after consolidation.

- **<rationalization> prompt.** Table 7 shows the few-shot prompt used for the rationalization stage of ECR-Chain set construction with ChatGPT. The specific 4-shot exemplars we employed in this prompt are shown in Figure 6. In each exemplar of the prompt, we first provide the conversation context, target utterance, and causal utterance requiring a supplemental reasoning chain as the sample question. Subsequently, we offer a four-step sample answer following the “explanation-stimulus-appraisal-reaction” format. When the model processes actual questions, it similarly starts by explaining the entire process of how the specified causal utterance stimulates the target emotion in free-text form and then provides the supplemental chain.

- **<consolidation> prompt.** Table 8 shows the few-shot prompt used for the consolidation stage of ECR-Chain set construction with ChatGPT. The specific 3-shot exemplars we employed in this prompt are shown in Figure 7. In each question, we provide the conversation context, target utterance, all causal utterances, and also the simple combined reasoning chains of all chains retained after filtering and those supplemented by rationalization. We ask ChatGPT to check the “Reactions”, “Appraisals”, and “Stimuli” parts of these chains, consolidating items in each part that are semantically similar into a single one to streamline the constructed rationale. The simplified “Reaction-Appraisal-Stimulus” chains are then appended to the “Theme” initially generated in the reasoning step, forming the final constructed ECR-Chain data.

- **<answer> prompt.** Table 9 shows the prompt that asks language models to directly predict causal utterances based on the input of the Causal Emotion Entailment (CEE) task. We select questions the same as used in the exemplars of the <reasoning> prompt, as shown in Figure 5. The difference lies in that the corresponding answer in each exemplar only contains the final causal utterances and does not include the reasoning chain part. In the multi-task training of Vicuna discussed in Section 3.4.2, we employ both this

prompt and the <reasoning> prompt to construct two types of instructional inputs for each question sample. Besides, in the few-shot learning experiments with ChatGPT described in Section 4.3, we use this prompt to obtain baseline results of predicting answers directly without reasoning.

B Details of the GPT-4 Evaluation

In Section 4.5, we ask language models to perform emotion-cause reasoning in an explainable manner. We utilize GPT-4 to evaluate the quality of the reasoning chains generated by each model. In this section, we detail the settings for the evaluation conducted with GPT-4 and show GPT-4’s specific evaluations in several cases.

B.1 Evaluation Settings

Following the evaluation template provided in [Wang *et al.*, 2023], we design a prompt template tailored to guide GPT-4 in evaluating generated emotion-cause reasoning chains, as shown in Table 10. We conduct pairwise comparisons among the three models, namely Vanilla-Vicuna, MT-Vicuna, and ChatGPT. Each time, we select a question sample from the test set, including the conversation context and the specified target utterance, and insert it into the [Question] part of the template. Subsequently, we collect two rationales generated by the two models to be compared and fill in the [Answer of Assistant 1] and [Answer of Assistant 2], respectively. The GPT-4 is instructed to compare these two rationales based on the three aspects we state in the prompt and to provide an evaluation for each of the rationales. Following the “evidence calibration evaluation” approach proposed in [Wang *et al.*, 2023], we require the model to generate the explanation first and then give the score. This ensures that the score is calibrated with the evaluation evidence. Note that we do not provide golden labels to GPT-4 to make it focus more on the quality of causal descriptions than the causal utterance predictions, since predictions that differ from the labels may also have plausible explanations.

For a specific model, we compare its generated rationale for a given test question with those of the other two models separately. The model’s score for this test sample is calculated as the average of the two scores it receives in the two comparisons. And the final score of a model is the average of its scores across all test set samples. Besides, the order of the two rationales in the evaluation prompt is randomized (not model-specific) to avoid the positional bias mentioned in [Wang *et al.*, 2023].

B.2 Evaluation Cases

We present several cases from the explainable ECR experiment to show GPT-4’s evaluative analysis. Figure 8 shows the same case as Figure 4. The second row in Figure 8 details GPT-4’s analysis of the multi-task model in comparison to the other two models (we omit the comparative analysis between Vanilla-Vicuna and ChatGPT). Note that GPT-4 was blind to the actual names of the models. We replace the model identifiers in GPT-4’s output with the actual model names for clarity. Figures 9 and Figure 10 follow the same format to show two additional cases.

| |
|--|
| <p><reasoning></p> <p>Please understand the emotion cause for the target utterance in a given conversation.</p> <p>In a given conversation, each sentence contains its index number, speaker, emotion, and its content, written in this format: #[index number]: [speaker] ([emotion]): "[content]". Besides, a target non-neutral utterance will be specified, where the emotion of the target non-neutral utterance may be one of the types of [happiness, sadness, anger, fear, surprise, disgust]. You should predict which particular utterances in the conversation history (including the target itself) are causal utterances that are responsible for the non-neutral emotion in the target utterance.</p> <p>Specifically, you should perform a 5-step reasoning process:</p> <ol style="list-style-type: none"> 1. Output the "Theme": Summarize the conversation theme with a few words. 2. List "Reactions" items of target utterance: Describe the target speaker's reaction or behavior with his/her target emotion when the target utterance is saying according to the conversation theme and the context. List them by items. 3. List "Appraisals" items of target utterance: Infer and describe the inner thoughts of the target speaker when the target utterance is saying based on the context and the "Reactions" items listed above. List each "Appraisals" item with its corresponding "Reactions" item ids. 4. List "Stimuli" items with their corresponding utterance id: The "Stimuli" items can be events, situations, facts, or experiences in the conversational context that are primarily responsible for the "Appraisals" items listed above. Apart from them, the "Stimuli" items could also be the other speaker's counterpart reactions toward events that are cared for by the target speaker and can cause these "Appraisals" items of the target speaker. List each "Stimuli" item with its corresponding "Appraisals" item ids. 5. Based on these "Stimuli" items, output the index number of the causal utterances in the form of Python list without any other content. <p>I will show you some examples:</p> <p>[Example 1]</p> <p>...</p> <p>— To be solved —</p> <p>[Question]</p> |
|--|

Table 6: Prompt template of <reasoning>.

| |
|---|
| <p><rationalization></p> <p>Please understand the emotion cause for the target utterance in a given conversation.</p> <p>In a given conversation, each sentence contains its index number, speaker, emotion, and its content, written in this format: #[index number]: [speaker] ([emotion]): "[content]". A target non-neutral utterance will be specified, where the emotion of the target non-neutral utterance may be one of the types of [happiness, sadness, anger, fear, surprise, disgust]. The target speaker is the corresponding speaker of the target utterance. Besides, a causal utterance will be given. A causal utterance is a particular utterance in the conversation history (can be the target utterance itself) responsible for the non-neutral emotion in the target utterance. You should find what the exact cause factor is from the causal utterance and understand how the factor causes the target speaker's emotional response reflected in the target utterance.</p> <p>Specifically, you should perform a 4-step reasoning process:</p> <ol style="list-style-type: none"> 1. Give an "Explanation": First, understand the semantics expressed by the cause utterance based on the conversation context. Then, consider what exact factor conveyed by the cause utterance might be responsible for the non-neutral emotion in the target utterance. You need to consider the factor in conjunction with the target speaker's inner thoughts and his/her reaction or behavior corresponding to the emotion in the target utterance. But note that if the speaker of the causal utterance is also the target speaker, the exact factor is more likely to be some objective events he/she described or his/her own subjective opinions, rather than his/her own actions. Because generally speaking, one's actions would not cause his/her own subsequent emotions, only his/her opinions or other events (including the other speaker's actions) can do. Use a few sentences to give your "Explanation" of the target speaker's emotion. 2. Extract the "Stimulus" from the causal utterance: A "Stimulus" can be an event, situation, opinion, or experience in the conversational context that is primarily responsible for the elicited emotion in the target utterance. Apart from them, the "Stimulus" could also be the other speaker's counterpart reaction towards an event cared for by the target speaker. Write the "Stimulus" in one brief sentence. Note that it should correspond to the exact factors you found in the "Explanation". 3. Infer the "Appraisal" of the target speaker in the target utterance: Infer and describe the inner thoughts of the target speaker when the target utterance is saying based on the conversation context, the "Explanation" and the "Stimulus" you have written before. Write the "Appraisal" in one brief sentence. 4. Describe the "Reaction" of the target speaker in the target utterance: Describe the target speaker's reaction or behavior with his/her target emotion when the target utterance is saying according to the context, the "Explanation" and the "Appraisal" you have written before. Write the "Reaction" in one brief sentence. <p>I will show you some examples:</p> <p>[Example 1]</p> <p>...</p> <p>— To be solved —</p> <p>[Question]</p> |
|---|

Table 7: Prompt template of <rationalization>.

| |
|---|
| <p><consolidation></p> <p>Please simplify the given reasoning chains for the conversational emotion cause understanding task.</p> <p>In a given conversation, each sentence contains its index number, speaker, emotion, and its content, written in this format: #[index number]: [speaker] ([emotion]): "[content]". A target utterance and some causal utterances will be given. The target utterance always has a non-neutral emotion. The target speaker is the corresponding speaker of the target utterance. The causal utterances are those in the conversation history (including the target itself) that contain causes responsible for the non-neutral emotion in the target utterance.</p> <p>Besides, some reasoning chains for understanding the causal relationships between the target utterance and causal utterances are also given. The reasoning chains consist of 3 parts: "Reactions", "Appraisals", and "Stimuli". "Reactions" are the target speaker's reaction or behavior with his/her target emotion when the target utterance is saying. "Appraisals" are the inner thoughts of the target speaker when the target utterance is saying. "Stimuli" can be events, situations, opinions, or experiences in the conversational context, and can also be the other speaker's counterpart reactions toward events. The given reasoning chains may have some redundant items in each of these 3 parts, so they may need to be simplified.</p> <p>Specifically, you need to note the following points:</p> <ul style="list-style-type: none"> - The simplified version still needs to have 3 parts: "Reactions", "Appraisals", and "Stimuli". - You can, based on semantics, re-summarize the content of each part into some new items. Ensure that different items should emphasize different aspects. - Unless necessary, please refrain from making excessive modifications to the original expression. However, you are allowed to make adjustments to merge several items, especially when these items share semantic similarities or when some of them are encompassed by others, resulting in a single item. - Each "Stimuli" item should be accompanied by its corresponding causal utterances ids, using the format like <stimulus description> (#<id[, id, ...]>). Ensure that all causal utterances ids are covered in the "Stimuli" part. <p>I will show you some examples:</p> <p>[Example 1]</p> <p>...</p> <p>— To be solved —</p> <p>[Question]</p> |
|---|

Table 8: Prompt template of <consolidation>.

| |
|--|
| <p><answer></p> <p>Please understand the emotion cause for the target utterance in a given conversation.</p> <p>In a given conversation, each sentence contains its index number, speaker, emotion, and its content, written in this format: #[index number]: [speaker] ([emotion]): "[content]". Besides, a target non-neutral utterance will be specified, where the emotion of the target non-neutral utterance may be one of the types of [happiness, sadness, anger, fear, surprise, disgust]. You should predict which particular utterances in the conversation history (including the target itself) are causal utterances that are responsible for the non-neutral emotion in the target utterance. Output the index number of the causal utterances in the form of Python list without any other content.</p> <p>I will show you some examples:</p> <p>[Example 1]</p> <p>...</p> <p>— To be solved —</p> <p>[Question]</p> |
|--|

Table 9: Prompt template of <answer>.

| |
|---|
| <GPT-4 evaluation> |
| <p><i>System:</i> You are an expert in understanding emotion and can assist in assessing the quality of the emotional cause reasoning chains.</p> |
| <p><i>User:</i> [Question] Question: Please give a “theme->reactions->appraisals->stimuli” chain to analyze the cause of the target speaker’s emotion expressed in the target utterance. Note: “Theme” means a summarization of the conversation in a few words and can be “none” if the conversation does not contain much information. “Reactions” means target speaker’s reaction or behavior with his/her target emotion. “Appraisals” means the inner thoughts of the target speaker. “Stimuli” can be events, situations, opinions, or experiences of the target speaker in the conversational context, and they can also be the other speaker’s counterpart reactions, but they should be primarily responsible for the target speaker’s emotion in the target utterance. [Answer of Assistant 1] [Answer of Assistant 2] We would like to request your feedback on the performance of two AI assistants in response to the emotional cause reasoning question displayed above. Please consider the logical coherence within each part of the chain, the relevance of the chain’s content to the original conversation, and the accuracy of the final inferred stimulus factors, leading to an overall evaluation. Each assistant receives an overall score on a scale of 1 to 10 (should be an integer), where a higher score indicates better overall performance. Please first provide a comprehensive explanation of your evaluation, avoiding any potential bias and ensuring that the order in which the responses were presented does not affect your judgment. Then, output two lines indicating the scores for Assistant 1 and 2, respectively. Output with the following format: Evaluation evidence: <your evaluation explanation here> The score of Assistant 1: <score> The score of Assistant 2: <score></p> |

Table 10: Prompt template of <GPT-4 evaluation>.

| No. | Question | Answer |
|-------|---|---|
| Ex. 1 | <p>Conversation: #1: A (neutral): "What a hot day !" #2: B (neutral): "Yes , summer is coming ." #3: A (neutral): "Where are you going this summer vacation ?" #4: B (neutral): "Td like to go to Tianjin ." #5: A (happiness): "Do you want to go to the beach there ?" #6: B (happiness): "Yes , of course . And what's your plan ?" #7: A (happiness): "The same to you ." #8: B (happiness): "That's great ."</p> <p>Target Utterance: #8</p> | <p>Theme: Summer vacation plan.</p> <p>Reactions of B in #8: 1. B is happy to hear that A's plan is the same as B's one.</p> <p>Appraisals of B in #8: 1. For Reaction 1: B thinks it is great that A's plan is the same as his own.</p> <p>Stimuli: 1. For Appraisal 1: A's plan is the same as B's one (#7).</p> <p>Causal utterances: [7]</p> |
| Ex. 2 | <p>Conversation: #1: A (anger): "Shut up ... Billy ... BILLY !" #2: B (surprise): "What !" #3: A (neutral): "Wow ! What's that ?" #4: B (happiness): "It's my new radio earphone ."</p> <p>Target Utterance: #4</p> | <p>Theme: Find the earphone.</p> <p>Reactions of B in #4: 1. B is happy for seeing his new radio earphone.</p> <p>Appraisals of B in #4: 1. For Reaction 1: B thinks he is so like this earphone.</p> <p>Stimuli: 1. For Appraisal 1: B finds his new radio earphone (#4).</p> <p>Causal utterances: [4]</p> |
| Ex. 3 | <p>Conversation: #1: A (neutral): "Hey , Mary , can you cut that out ?" #2: B (neutral): "Cut what out I ' m not doing anything ." #3: A (anger): "The tapping of your pen on your desk . It ' s driving me crazy ." #4: B (anger): "Fine ! By the way would you mind not slurping your coffee every time you have a cup !" #5: A (anger): "I don ' t slurp my coffee . And plus , how can you hear it when you ' re shouting into your phone all the time ?" #6: B (anger): "You ' Ve got to be kidding me ! You ' re complaining about me talking on the phone when you go out for a cigarette break ten times a day to shoot the breeze ?"</p> <p>Target Utterance: #6</p> | <p>Theme: Complain about each other's noise.</p> <p>Reactions of B in #6: 1. B expresses his anger with A's complaint to B. 2. B angrily points out that A chats frequently which also makes much noise.</p> <p>Appraisals of B in #6: 1. For Reaction 2: B thinks the noise made by A is annoying. 2. For Reaction 1 & 2: B thinks that A has no right to blame B for making noise as A is also used to making noise.</p> <p>Stimuli: 1. For Appraisal 1: A is used to slurping his coffee which makes some annoying noise (#4). 2. For Appraisal 2: A accused B of tapping the pen on the desk (#3). 3. For Appraisal 2: A accused B of making noise during the phone call (#5, #6).</p> <p>Causal utterances: [3, 4, 5, 6]</p> |
| Ex. 4 | <p>Conversation: #1: A (happiness): "Nice and warm again , isn ' t it ?" #2: B (happiness): "Oh , it ' s lovely . Think of the nasty weather we had last week ." #3: A (happiness): "How are you these days ?"</p> <p>Target Utterance: #3</p> | <p>Theme: Weather, greeting.</p> <p>Reactions of A in #3: 1. A has been happy about the good weather till now.</p> <p>Appraisals of A in #3: 1. For Reaction 1: A thinks the weather is comfortable for him.</p> <p>Stimuli: 1. For Appraisal 1: The weather is nice and warm (#1).</p> <p>Causal utterances: [1]</p> |

Figure 5: 4-shot exemplars of <reasoning> prompt.

| No. | Question | Answer |
|-------|--|--|
| Ex. 1 | <p>Conversation:</p> <p>#1: A (neutral): "How's your business ?"</p> <p>#2: B (disgust): "Everything seems to be getting worse . I don't know what to do with it ."</p> <p>#3: A (neutral): "Things are tough all over . You should never give up !"</p> <p>#4: B (neutral): "But how can I prevent a deficit ?"</p> <p>#5: A (happiness): "I think you should make a thorough market investigation now ."</p> <p>#6: B (happiness): "OK , OK . I ' ll do my best ."</p> <p>#7: A (happiness): "I always consider you are such a guy who never gives up ."</p> <p>#8: B (happiness): "I know life is not all roses , and I believe I will pull this through ."</p> <p>#9: A (happiness): "Yeah , you can make it ."</p> <p>Target Utterance: #9</p> <p>Causal Utterance: #6</p> | <p>Explanation:</p> <p>The two speakers are talking about B's business. A advised B to make a market investigation to prevent a deficit. In the cause utterance #6, B accepted A's suggestion and promised to do his best. B's performance of not being afraid of difficulties moved A and made A willing to believe that he could succeed. Therefore, in the target utterance #9, A happily encourages B that he can make it.</p> <p>Stimulus in #6:</p> <ul style="list-style-type: none"> - B promises to do his best. <p>Appraisal of A in #9:</p> <ul style="list-style-type: none"> - A thinks B is not afraid of difficulties and will succeed in the end. <p>Reaction of A in #9:</p> <ul style="list-style-type: none"> - A expresses his happiness and encouragement to B. |
| Ex. 2 | <p>Conversation:</p> <p>#1: A (neutral): "Oh , God . It ' s late . I ' m afraid I have to leave ."</p> <p>#2: B (neutral): "But you just got here . Can ' t you stay a little longer and have some tea ?"</p> <p>#3: A (happiness): "That ' s very kind of you . But if I don ' t go now I ' ll miss the last bus ."</p> <p>#4: B (neutral): "That ' s too bad ."</p> <p>#5: A (happiness): "Thank you for your hospitality ."</p> <p>Target Utterance: #5</p> <p>Causal Utterance: #3</p> | <p>Explanation:</p> <p>In this conversation, B is trying to persuade A to stay a while. In the cause utterance #3, A was pleased by B's effort and thought that B is very kind. It's A's view of B that made A want to thank B. So in the target utterance #5, A expresses his happiness and gratitude for B's hospitality.</p> <p>Stimulus in #3:</p> <ul style="list-style-type: none"> - A's view that B is very kind. <p>Appraisal of A in #5:</p> <ul style="list-style-type: none"> - A thinks B is very kind. <p>Reaction of A in #5:</p> <ul style="list-style-type: none"> - A expresses his happiness and gratitude for B's hospitality. |
| Ex. 3 | <p>Conversation:</p> <p>#1: A (neutral): "Here ' s your hot dog and beer . What happened ? Did I miss anything ?"</p> <p>#2: B (neutral): "Yeah , Cal Ripen just hit a home run ."</p> <p>#3: A (neutral): "What ' s the score ?"</p> <p>#4: B (neutral): "Well it was 3 to 4 , but Ripen ' s home run made it 5 to 4 since another player was on first base ."</p> <p>#5: A (neutral): "So Baltimore is winning ?"</p> <p>#6: B (happiness): "Right ."</p> <p>#7: A (happiness): "This is a really great place to watch a baseball game ."</p> <p>#8: B (happiness): "Yeah , there isn ' t a bad seat in the place ."</p> <p>Target Utterance: #8</p> <p>Causal Utterance: #8</p> | <p>Explanation:</p> <p>The two speakers are chatting while watching a baseball game. A said that this is a good place to watch a baseball game. In the cause utterance #8, B agrees with A's opinion because he thinks that there isn't a bad seat in the place. It's B's view of the watching place that makes B feel happy and praise for this watching place in the target utterance #8.</p> <p>Stimulus in #8:</p> <ul style="list-style-type: none"> - B's view that there isn't a bad seat in the place they sit in. <p>Appraisal of B in #8:</p> <ul style="list-style-type: none"> - B thinks that the place they sit in is good. <p>Reaction of B in #8:</p> <ul style="list-style-type: none"> - B happily praises for their watching place. |
| Ex. 4 | <p>Conversation:</p> <p>#1: A (neutral): "Jane , I was wondering if you had any plans for Saturday afternoon ."</p> <p>#2: B (neutral): "A friend and I are planning to go out , why ? what's up ?"</p> <p>#3: A (neutral): "There is a special exhibition of French sculptures at the museum , I was hoping you ' d like to come with me ."</p> <p>#4: B (sadness): "I am afraid I can't I am going to be out all day ."</p> <p>Target Utterance: #4</p> <p>Causal Utterance: #3</p> | <p>Explanation:</p> <p>The two speakers are talking about plans for Saturday afternoon. In the cause utterance #3, A told B that there was a special exhibition and invited B to go with him. However, we know that B will have other arrangements on that day according to the target utterance #4. So A's invitation may make B feel regretful, which leads B to express his disappointments and regrets for not going with A in the target utterance #4.</p> <p>Stimulus in #3:</p> <ul style="list-style-type: none"> - A invites B to go to the exhibition together. <p>Appraisal of B in #4:</p> <ul style="list-style-type: none"> - B thinks it is regretful to decline A's invitation. <p>Reaction of B in #4:</p> <ul style="list-style-type: none"> - B regretfully explains the reason why he cannot go with A. |

Figure 6: 4-shot exemplars of <rationalization> prompt.

| No. | Question | Answer |
|-------|--|---|
| Ex. 1 | <p>Conversation: #1: A (happiness): "It's a lovely day , isn't it ?" #2: B (happiness): "Yes , the weather sure is nice today ." #3: A (happiness): "I love it when the weather is like this . Why don't we sit down outside and have our lunch ?" #4: B (happiness): "That sounds like a good idea . The sky is clear , the sun is shining and there's a nice cool breeze , so it is not too hot ."</p> <p>Target Utterance: #4 Causal Utterances: [#1, #2, #3, #4]</p> | <p>Original Reasoning Chains: <u>Reactions of B in #4:</u> 1. B agrees with A's suggestion of having lunch outside. 2. B describes the pleasant weather conditions. 3. B expresses his happiness and agrees that having lunch outside is a good idea. 4. B enthusiastically agrees with A's suggestion and provides additional reasons why it's a good idea. <u>Appraisals of B in #4:</u> 1. B thinks it is a good idea to have lunch outside. 2. B thinks the weather is perfect for outdoor activities. 3. B thinks that the weather conditions described by A are favorable for having lunch outside. 4. B thinks it's a good idea to have lunch outside because the sky is clear, the sun is shining, and there's a nice cool breeze. <u>Stimuli:</u> 1. A suggests having lunch outside (#3). 2. The sky is clear, the sun is shining, and there's a nice cool breeze (#4). 3. A's statement about the lovely day (#1). 4. B's agreement with A's statement about the nice weather (#2).</p> <p>Reactions of B in #4: 1. B enthusiastically agrees with A's suggestion of having lunch outside. 2. B describes the pleasant weather conditions. 3. B expresses his happiness and says that having lunch outside is a good idea. Appraisals of B in #4: 1. B thinks it is a good idea to have lunch outside. 2. B thinks the weather is perfect for outdoor activities. Stimuli: 1. A's statement about the lovely day (#1). 2. B's agreement with A's statement about the nice weather (#2). 3. A suggests having lunch outside (#3). 4. The sky is clear, the sun is shining, and there's a nice cool breeze (#4).</p> |
| Ex. 2 | <p>Conversation: #1: A (happiness): "Happy Mother's Day , Mom !" #2: B (happiness): "Thanks ! These are beautiful . You shouldn't have !"</p> <p>Target Utterance: #2 Causal Utterances: [#1, #2]</p> | <p>Original Reasoning Chains: <u>Reactions of B in #2:</u> 1. B expresses happiness and gratitude for the beautiful gift. 2. B mentions that A shouldn't have gotten the gift. 3. B happily thanks A and expresses his appreciation for the beautiful gift. <u>Appraisals of B in #2:</u> 1. B thinks the gift is beautiful and appreciates it. 2. B thinks A shouldn't have spent money on the gift. 3. B is pleasantly surprised and feels grateful for the unexpected gift. <u>Stimuli:</u> 1. A gives B a beautiful gift for Mother's Day (#1). 2. A spends money on the gift for B (#1). 3. A gives B a beautiful gift (#2).</p> <p>Reactions of B in #2: 1. B expresses happiness and gratitude for the beautiful gift. 2. B mentions that A shouldn't have gotten the gift. Appraisals of B in #2: 1. B thinks the gift is beautiful and appreciates it. 2. B thinks A shouldn't have spent money on the gift. 3. B is pleasantly surprised and feels grateful for the unexpected gift. Stimuli: 1. A gives B a beautiful gift for Mother's Day (#1, #2). 2. A spends money on the gift for B (#1).</p> |
| Ex. 3 | <p>Conversation: #1: A (happiness): "Do you have a park in the city ?" #2: B (happiness): "Yes , there is a lake in the park . It is very beautiful ." #3: A (happiness): "That 's great . We can go boating on the lake in the afternoon ."</p> <p>Target Utterance: #3 Causal Utterances: [#2, #3]</p> | <p>Original Reasoning Chains: <u>Reactions of A in #3:</u> 1. A expresses happiness about the plan to go boating on the lake in the afternoon. 2. A expresses his happiness and excitement about the opportunity to go boating on the lake in the afternoon. <u>Appraisals of A in #3:</u> 1. A thinks it is great to have the opportunity to go boating on the lake in the afternoon. 2. A thinks it is great that there is a lake in the park and they can go boating in the afternoon. <u>Stimuli:</u> 1. B mentioned that there is a lake in the park (#2). 2. A's positive view of the park and the opportunity to go boating on the lake in the afternoon (#3).</p> <p>Reactions of A in #3: 1. A expresses happiness and excitement about the plan to go boating on the lake in the afternoon. Appraisals of A in #3: 1. A thinks it is great to have the opportunity to go boating on the lake in the afternoon. Stimuli: 1. B mentioned that there is a lake in the park (#2). 2. A's positive view of the park and the opportunity to go boating on the lake in the afternoon (#3).</p> |

Figure 7: 3-shot exemplars of <consolidation> prompt.

| | | | |
|---|---|---|---|
| <p>Conversation: #1: A (neutral): "Is there anything I can do for you ?" #2: B (neutral): "Would you please show me that mural ?" #3: A (happiness): "Certainly , Here you are ." #4: B (happiness): "It looks so beautiful . How much is it ?" #5: A (neutral): "Four hundred yuan ." #6: B (neutral): "That's a bit more than I wanted to pay . Can you cut the price down a bit ?" #7: A (neutral): "I'm sorry . Ours is a fixed price shop and I can assure you this is the best price you can get for a mural like this ." #8: B (neutral): "Maybe you are right . OK , I'll take it ." #9: A (happiness): "Well , I'll have it wrapped for you ." #10: B (happiness): "Thank you . Here is four hundred yuan ." #11: A (happiness): "Thank you . Please come again ."</p> <p>Question: Which utterances cause A's happiness in #11?</p> <p>Label: #10</p> | <p>[Vanilla-Vicuna]</p> <p>Theme: The conversation is about buying a mural.</p> <p>Reactions of A in #11: 1. A is happy to help B and show the mural. 2. A is happy that B likes the mural and wants to buy it.</p> <p>Appraisals of A in #11: 1. A thinks B is a nice customer. 2. A thinks B is willing to pay the fixed price for the mural. 3. A thinks B is a nice customer (#11).</p> <p>Stimuli: 1. B asks A to show the mural (#2). 2. B is interested in buying the mural (#4, #6). 3. A thinks B is a nice customer (#11).</p> <p>GPT4 Score: 6.5</p> | <p>[MT-Vicuna]</p> <p>Theme: Selling a mural and expressing gratitude.</p> <p>Reactions of A in #11: 1. A expresses happiness and gratitude towards B for making a purchase.</p> <p>Appraisals of A in #11: 1. A thinks it is a good thing that B made a purchase and is satisfied with the service.</p> <p>Stimuli: 1. B made a purchase and gave A four hundred yuan (#10).</p> <p>GPT4 Score: 8.0</p> | <p>[ChatGPT]</p> <p>Theme: Buying a mural in a shop.</p> <p>Reactions of A in #11: 1. A is happy to receive the payment from B.</p> <p>Appraisals of A in #11: 1. A thinks it is a positive interaction with B. 2. A feels satisfied with the transaction.</p> <p>Stimuli: 1. B thanked A and paid for the mural (#10). 2. B agreed to purchase the mural (#8). 3. A wrapped the mural for B (#9).</p> <p>GPT4 Score: 9.0</p> |
| <p>[Vanilla-Vicuna] vs. [MT-Vicuna]</p> <p>[Vanilla-Vicuna]'s response provides a more detailed analysis of the emotional cause reasoning chain, including a broader range of reactions and appraisals. It correctly identifies the theme of the conversation and includes multiple stimuli that contribute to A's happiness. However, it inaccurately states that A is happy to help B show the mural in #11, which is not directly supported by the conversation at that point. The happiness in #11 is more directly related to the completion of the sale.</p> <p>[MT-Vicuna]'s response is more concise and directly relevant to the target utterance #11. It correctly identifies the theme as selling a mural and expressing gratitude, and it accurately captures the stimulus leading to A's happiness, which is B's purchase and payment for the mural. The appraisal is also relevant, as it reflects A's satisfaction with the transaction. However, [MT-Vicuna]'s response lacks the depth and detail found in [Vanilla-Vicuna]'s response, which could provide a more nuanced understanding of A's emotional state throughout the conversation.</p> <p>Both assistants could improve by focusing more on the specific utterance #11 and providing a more precise appraisal that directly leads to the happiness expressed in that moment. The appraisal should reflect A's satisfaction with the successful sale and the positive interaction with the customer, rather than a general sense of B being a nice customer or the service being satisfactory.</p> <p>The score of [Vanilla-Vicuna]: 7 The score of [MT-Vicuna]: 8</p> | | <p>[MT-Vicuna] vs. [ChatGPT]</p> <p>Both [MT-Vicuna] and [ChatGPT] provided coherent and relevant emotional cause reasoning chains for Speaker A's happiness in utterance #11. [MT-Vicuna]'s response was succinct and directly linked the happiness to the purchase made by Speaker B, which is a straightforward and accurate appraisal of the situation. [MT-Vicuna]'s theme of "Selling a mural and expressing gratitude" is a concise summary of the conversation's context.</p> <p>[ChatGPT], on the other hand, expanded on the stimuli by including the fact that B thanked A and paid for the mural, B agreed to purchase the mural, and A wrapped the mural for B. This response provides a more detailed account of the stimuli leading to A's happiness, which shows a deeper analysis of the conversation. However, the inclusion of "A is happy that B expressed gratitude" as a reaction seems to be an assumption, as there is no explicit indication in the conversation that A's happiness is due to B's expression of gratitude. The theme provided by [ChatGPT], "Buying a mural in a shop," is accurate but does not capture the aspect of gratitude which is present in the conversation.</p> <p>Both assistants have accurately identified the stimuli related to the transaction and the payment, which are the primary causes of A's happiness. However, [ChatGPT]'s more detailed analysis of the stimuli provides a slightly more comprehensive understanding of the situation, despite the minor assumption made about A's reaction to B's gratitude.</p> <p>The score of [MT-Vicuna]: 8 The score of [ChatGPT]: 9</p> | |

Figure 8: Case 1 of GPT-4 evaluation, which is a detailed version of Figure 4 that contains GPT-4's specific analysis of two comparisons. According to the analysis, Vanilla-Vicuna's description of the target speaker's reaction is inaccurate (blue background); MT-Vicuna's response is accurate and direct but somewhat lacks depth (red text); ChatGPT provides a more detailed analysis of stimuli (light green background), although its reaction part may include some assumptions (dark green background).

| | | | |
|--|--|--|--|
| <p>Conversation: #1: A (neutral): "Hello , ma ' am , can I help you find something ?" #2: B (neutral): "Yes , actually I ' m looking to buy a camera ." #3: A (neutral): "We ' Ve got a wide selection do you know if you ' d like a point-and-shoot , or something a little fancier ? Are you shopping for yourself or for someone else ?" #4: B (neutral): "Actually I ' m buying a camera for my husband ." #5: A (neutral): "Ah , well then I ' d recommend a nice entry-level digital SLR ." #6: B (neutral): "Yeah ? Can I take a look at the Sirs you carry ?" #7: A (happiness): "Sure thing , follow me . This here is the ."</p> <p>Question: Which utterances cause A's happiness in #7?</p> <p>Label: #6</p> | <p>[Vanilla-Vicuna]</p> <p>Theme: Help customers find products.</p> <p>Reactions of A in #7: 1. A is happy to help the customer find the camera. 2. A is proud of having a wide selection of cameras.</p> <p>Appraisals of A in #7: 1. A thinks it's a pleasure to help customers find what they need. 2. A thinks it's important to have a wide selection of cameras to cater to different needs.</p> <p>Stimuli: 1. A is happy to help the customer find the camera (#7). 2. A is proud of having a wide selection of cameras (#3).</p> <p>GPT4 Score: 5.5</p> | <p>[MT-Vicuna]</p> <p>Theme: Selling cameras.</p> <p>Reactions of A in #7: 1. A happily agrees to show B the cameras they carry.</p> <p>Appraisals of A in #7: 1. A thinks it is a good opportunity to show B the cameras they carry.</p> <p>Stimuli: 1. B asks A to show him the cameras they carry (#6).</p> <p>GPT4 Score: 8.0</p> | <p>[ChatGPT]</p> <p>Theme: Camera shopping.</p> <p>Reactions of A in #7: 1. A is happy to assist B in finding a camera.</p> <p>Appraisals of A in #7: 1. A thinks it is a good thing to help B find a camera.</p> <p>Stimuli: 1. B is looking to buy a camera (#2). 2. B is buying a camera for her husband (#4).</p> <p>GPT4 Score: 7.5</p> |
| <p>[Vanilla-Vicuna] vs. [MT-Vicuna]</p> <p>[Vanilla-Vicuna]'s response provides a more detailed analysis of the emotional cause reasoning chain, including both reactions and appraisals, which are relevant to the conversation. The theme identified as "Help customers find products" is accurate and reflects the context of the conversation. The reactions listed are plausible, suggesting that A is happy to help and proud of the camera selection. The appraisals are also relevant, indicating that A finds pleasure in helping customers and values the importance of a wide selection. However, the stimuli section is redundant and merely repeats the reactions without identifying the specific event that triggered A's happiness.</p> <p>[MT-Vicuna]'s response is more concise and directly connects the stimulus to the target emotion. The theme "Selling cameras" is a bit narrower but still relevant. The reaction identified is appropriate, showing A's happiness in agreeing to show the cameras. The appraisal is straightforward, suggesting that A sees it as a good opportunity to showcase their products. The stimulus is correctly identified as B's request to see the cameras, which is the direct cause of A's happiness in the target utterance.</p> <p>Both assistants have room for improvement. [Vanilla-Vicuna] could refine the stimuli section to be more specific and avoid repetition, while [MT-Vicuna] could provide a more detailed appraisal to enrich the emotional cause reasoning chain.</p> <p>The score of [Vanilla-Vicuna]: 6 The score of [MT-Vicuna]: 7</p> | | <p>[MT-Vicuna] vs. [ChatGPT]</p> <p>Both [MT-Vicuna] and [ChatGPT] provided coherent emotional cause reasoning chains for the target speaker A's happiness in utterance #7. However, there are subtle differences in their responses that affect the quality of their analysis.</p> <p>[MT-Vicuna]'s theme of "Selling cameras" is a straightforward and accurate summary of the conversation's context. The reaction identified is also directly related to the target utterance, where A happily agrees to show B the cameras. The appraisal is logical, as A likely sees B's request as a good opportunity to showcase their products, which aligns with A's role as a salesperson. The stimulus identified by [MT-Vicuna] is specific and directly related to the previous utterance (#6), where B asks to see the cameras, which is a clear trigger for A's happiness.</p> <p>[ChatGPT]'s theme of "Camera shopping" is also accurate and captures the essence of the conversation. The reaction is similar to [MT-Vicuna]'s, noting A's happiness in assisting B. However, the appraisals and stimuli identified by [ChatGPT] are broader and less specific. [ChatGPT] includes B's intent to buy a camera (#2) and the fact that it is for her husband (#4) as stimuli. While these are relevant to the overall conversation, they are not the immediate triggers for A's happiness in utterance #7. The most direct stimulus for A's happiness is B's request to see the cameras, which is only indirectly implied in [ChatGPT]'s response.</p> <p>Based on these observations, [MT-Vicuna]'s response is slightly more accurate in pinpointing the direct stimulus for A's happiness, while [ChatGPT]'s response, although relevant, is less precise in identifying the immediate cause of the emotion.</p> <p>The score of [MT-Vicuna]: 9 The score of [ChatGPT]: 8</p> | |

Figure 9: Case 2 of GPT-4 evaluation. According to the analysis, Vanilla-Vicuna incorrectly treats A's reactions as stimuli and fails to identify the real triggers (blue background); ChatGPT's provided appraisals and stimuli are broader and less specific (green background); only MT-Vicuna accurately pinpoints the direct stimulus (red text), although its theme appears a bit narrower (red background).

| | | | |
|---|---|--|---|
| <p>Conversation: #1: A (neutral): "Do you have any shoes like these ?" #2: B (neutral): "What size ?" #3: A (neutral): "Size five ." #4: B (neutral): "What colour ?" #5: A (neutral): "Black ." #6: B (sadness): "I ' m sorry . We don't have any ."</p> <p>Question: Which utterances cause B's sadness in #6?</p> <p>Label: [#6]</p> | <p>[Vanilla-Vicuna]</p> <p>Theme: Inquiring about shoes.</p> <p>Reactions of B in #6: 1. B expresses sadness for not having the shoes in stock. 2. B apologizes for not having the shoes in stock.</p> <p>Appraisals of B in #6: 1. B thinks that the shoes are not available in the store. 2. B feels sorry for not being able to provide the shoes to the customer.</p> <p>Stimuli: 1. A asked for a size and color of shoes (#3, #5). 2. B asked for the size of the shoes (#2). 3. B asked for the color of the shoes (#4).</p> <p>GPT4 Score: 6.5</p> | <p>[MT-Vicuna]</p> <p>Theme: Shoe purchase.</p> <p>Reactions of B in #6: 1. B expresses sadness and apologizes to A for not having any shoes like the ones A is looking for.</p> <p>Appraisals of B in #6: 1. B feels sorry for not being able to fulfill A's request.</p> <p>Stimuli: 1. B realizes that they don't have any shoes like A is looking for (#6).</p> <p>GPT4 Score: 8.5</p> | <p>[ChatGPT]</p> <p>Theme: Shopping for shoes.</p> <p>Reactions of B in #6: 1. B is sad because they don't have the shoes A is looking for.</p> <p>Appraisals of B in #6: 1. B feels sorry for not being able to fulfill A's request. 2. B is disappointed that they don't have the shoes A wants.</p> <p>Stimuli: 1. A asks if they have shoes like the ones A is describing (#1). 2. A specifies the size of the shoes (#3). 3. A specifies the color of the shoes (#5).</p> <p>GPT4 Score: 8.0</p> |
| <p>[Vanilla-Vicuna] vs. [MT-Vicuna]</p> <p>Both [Vanilla-Vicuna] and [MT-Vicuna] provided coherent emotional cause reasoning chains for the sadness expressed by Speaker B in utterance #6. However, there are differences in the detail and accuracy of their responses.</p> <p>[Vanilla-Vicuna]'s response includes a more detailed appraisal and stimuli section. The appraisal correctly identifies that B feels sorry for not being able to provide the shoes to the customer, which is a reasonable internal thought process that could lead to the expression of sadness. The stimuli section, however, includes unnecessary details such as "A asked for a size and color of shoes" and "B asked for the size of the shoes," which are not directly responsible for B's emotion in the target utterance. The actual stimulus for B's sadness is the realization that the store does not have the shoes in the requested size and color, which is only implied in [Vanilla-Vicuna]'s response.</p> <p>[MT-Vicuna]'s response is more concise and directly addresses the stimulus that leads to B's sadness, which is the realization that they do not have the shoes A is looking for. This is a more accurate and relevant stimulus than the ones provided by [Vanilla-Vicuna]. However, [MT-Vicuna]'s appraisal section could have been slightly more detailed by including the thought process behind B's apology, such as a sense of responsibility or a desire to provide good customer service.</p> <p>Overall, [MT-Vicuna]'s response is more focused and relevant to the target utterance, despite being less detailed in the appraisal section. [Vanilla-Vicuna]'s response, while more detailed, includes stimuli that are not directly responsible for B's sadness.</p> <p>The score of [Vanilla-Vicuna]: 6 The score of [MT-Vicuna]: 8</p> | | <p>[MT-Vicuna] vs. [ChatGPT]</p> <p>Both [MT-Vicuna] and [ChatGPT] provided coherent and relevant emotional cause reasoning chains for the target speaker B's sadness in utterance #6. They both identified the theme correctly as related to a shoe purchase or shopping for shoes. Their reactions and appraisals are also consistent with the conversation, recognizing B's sadness and apology for not having the shoes A is looking for.</p> <p>[MT-Vicuna]'s appraisal is concise, focusing on B's feeling of sorrow for not being able to fulfill A's request. The stimulus identified is directly related to the target utterance, which is B's realization that they do not have the shoes A wants.</p> <p>[ChatGPT]'s appraisal adds an additional point, suggesting that B is disappointed, which is a reasonable inference but not explicitly stated in the conversation. The stimuli identified by [ChatGPT] include the entire interaction leading up to the realization that the shoes are not available. This is a more comprehensive approach, considering the buildup of the conversation, but it could be argued that only the final realization (the lack of shoes) is the direct stimulus for B's sadness.</p> <p>Both assistants have provided accurate and relevant chains, but [ChatGPT]'s inclusion of the entire interaction as stimuli might be seen as slightly less focused on the direct cause of B's sadness. However, this does not significantly detract from the quality of the response.</p> <p>The score of [MT-Vicuna]: 9 The score of [ChatGPT]: 8</p> | |

Figure 10: Case 3 of GPT-4 evaluation. According to the analysis, Vanilla-Vicuna’s stimuli part includes unnecessary details and lacks clarity (blue background); ChatGPT considers the entire interaction as the stimuli, failing to capture the direct trigger (dark green background), and its appraisal contains some assumptions (light green background); MT-Vicuna address the stimulus factors of B’s sadness accurately (red text), though a more in-depth appraisal part would be preferable (red background).