

Appendix for “ESCA: An Emotional Support Conversation Agent for Enhancing Reasonable Strategy Planning and Effective Expression”

Detail of Dataset Preparation

Rules of States

For training the strategy planner, we annotate four aspects of seekers' states in the ESCconv dataset (Liu et al. 2021). Here we show the rules of these four aspects. Firstly, negative emotion intensity is annotated, where a higher score indicates a stronger negative emotional state. The annotation rules are as follow:

- **Intensity 1:** Barely perceptible negative emotion or neutral.
- **Intensity 2:** Slight negative emotion, somewhat subdued and not dominant.
- **Intensity 3:** Clear presence of negative emotion, noticeable but not overwhelming.
- **Intensity 4:** Strong negative emotion, highly expressive and impactful.
- **Intensity 5:** Extreme negative emotion, highly intense, possibly exaggerated or uncontrollable

Secondly, the categories of dialogue behavior are referenced by (Saha et al. 2020). As some of these behaviors rarely occur in the ESC scenario, we adapt the categorization and provide new annotation rules as follow:

- **Greeting:** Expressing a greeting or initiating a conversation.
- **Question:** Asking a question to seek information or feedback.
- **Feedback:** Providing feedback including a direct response or answer to a question, expressing agreement or disagreement.
- **Statement-fact:** Stating facts or objective information without personal opinions.
- **Statement-opinion:** Expressing a personal viewpoint or subjective evaluation.
- **Statement-emotion:** Explicitly expressing emotions such as sadness, frustration, fear, or happiness.
- **Command:** Giving a directive or stating a personal need.
- **Acknowledgement:** Recognizing a fact or responding without agreement or disagreement.
- **Others:** Any utterance that does not fit into the above categories.

Thirdly, for the stage of change, we follow the categories and rules in (Prochaska, Norcross, and DiClemente 2005):

- **Precontemplation:** The seeker is unaware of the need to change or has no desire to change.
- **Contemplation:** The seeker is aware of and accepts responsibility for their problems.
- **Preparation:** The seeker commits to change and is preparing to begin the change process.
- **Action:** The seeker is actively making changes in behavior or environment.
- **Maintenance:** The seeker is sustaining change and working to prevent relapse.

At last, the trust is a complex concept involving many dimensions. In terms of the method of trust measurement by questionnaire proposed by Madsen and Gregor (Madsen and Gregor 2000), we pay attention to the cognition-based trust which focuses on the reliability, response competence and perceived understandability three dimensions and each dimensions are scaled five levels. The trust degree is computed by averaging the scores of the three dimensions and mapping the result onto a five-level scale. For being suitable for the ESC scenario, we adjust the specific rules as follows:

Reliability: (Evaluate how stable and trustworthy the supporter is across the dialogue.)

- **Level 1:** The seeker likely feels misled or confused by inconsistent responses.
- **Level 2:** This is the first time the supporter appears unreliable or evasive.
- **Level 3:** The seeker appears unsure; the supporter's trustworthiness is unclear.
- **Level 4:** Mostly consistent; only minor uncertainty.
- **Level 5:** Clearly reliable and consistent throughout.

Response Competence: (Evaluate how well the supporter addresses the seeker's problem.)

- **Level 1:** Irrelevant or unhelpful response.
- **Level 2:** Somewhat related but vague or shallow.
- **Level 3:** Partially helpful but lacking clarity or depth.
- **Level 4:** Clear, relevant, and mostly effective.
- **Level 5:** Fully addresses the concern with insight and precision.

Perceived Understandability: (Evaluate how much the seeker feels understood.)

- **Level 1:** The seeker feels totally misunderstood or invalidated.
- **Level 2:** Some effort, but still a large disconnect.
- **Level 3:** Some understanding, but not deep or complete.
- **Level 4:** Seeker feels clearly understood and emotionally supported.
- **Level 5:** Strong sense of emotional connection and deep understanding.

Prompt of annotation

After clarifying the rules for the four annotation aspects, we proceed to annotate the existing ESConv dataset. One of the most critical challenges in this process is ensuring annotation accuracy. Although large language models (LLMs) have been widely applied to extend and label datasets, accurately annotating seekers' internal states remains difficult due to their inherent complexity, ambiguity, and subjectivity.

To improve annotation quality, we adopt a sliding window strategy inspired by (Saha et al. 2020). Rather than labeling each utterance independently, we annotate four consecutive seeker utterances at a time, using a window size of 4 and a step size of 2. This overlapping window enables the LLMs to make relative comparisons across utterances within the context, helping it better assess emotional intensity and other state-related aspects.

To further improve annotation reliability, we utilize three different LLMs (Qwen72b¹ (Bai et al. 2023), Llama3.3-instruct-70b² (Grattafiori et al. 2024), Deepseek-distill-llama-70b³ (Guo et al. 2025)) to independently annotate each dialogue segment. Final labels are determined via majority voting, which helps reduce individual model biases and enhances consistency across annotations. In addition, we decompose the annotation of all aspects into three separate parts to reduce task complexity and potential inter-aspect interference:

Emotional Intensity is annotated separately. Since the ESConv dataset already provides emotion intensity labels for the first and last turns of each dialogue, we retain these as fixed references. For intermediate utterances, we use a one-shot in-context learning approach following Table 1 to annotate the negative emotion intensity.

Dialogue Behavior and Stage of Change are annotated together, as both are annotated categories. We use a one-shot in-context learning for these aspects, as detailed in Table 2.

Trust degree is a complex and inherently multi-dimensional construct. To annotate its three dimensions, we use a one-shot instruction prompting the LLMs to generate all scores simultaneously (see Table 3).

To better illustrate the format of the annotation results, an example of the final annotated dialogue data is presented in Listing 1.

¹<https://huggingface.co/Qwen/Qwen2.5-VL-72B-Instruct>

²<https://huggingface.co/meta-llama/Llama-3.3-70B-Instruct>

³<https://huggingface.co/deepseek-ai/DeepSeek-R1-Distill-Llama-70B>

```

1   {
2     "emotion type": "depression",
3     "problem type": "job crisis",
4     "situation": "I'm in depression
5       because of today's market situation
6       .",
7     "dialog": [
8       {
9         "speaker": "seeker",
10        "text": "I'm in depression because
11          I have fear of losing my job
12          .",
13        "emotion intensity": 4,
14        "behavior": "statement-emotion",
15        "stage of change": "contemplation
16          ",
17        "trust": 3
18      },
19      {
20        "speaker": "supporter",
21        "text": "Why do you believe you
22          will lose your job?",
23        "strategy": "Question"
24      },
25      {
26        "speaker": "seeker",
27        "text": "Because lots of people in
28          my office lost their jobs due
29          to the job crisis during COVID
30          .",
31        "emotion intensity": 3,
32        "behavior": "statement-fact",
33        "stage of change": "contemplation
34          ",
35        "trust": 4
36      },
37      {
38        "speaker": "supporter",
39        "text": "This COVID crisis has
40          been affecting so many lives, I
41          totally understand.",
42        "strategy": "Affirmation and
43          Reassurance"
44      }
45    ]
46  }

```

Listing 1: An example of annotated data.

Data Analysis

After annotation, we analyze the feature of different aspects in ESConv dataset. As illustrated in Figure 1, we randomly sample five multi-turn dialogues. During these interactions, the intensity of negative emotions exhibits a fluctuating yet overall declining trend, while trust degree demonstrates a steady increase. Both of these changes are consistent with the goal of ESC and helping skill. Figure 2 shows the frequency of different behavior–strategy and stage-of-change–strategy combinations, based on 100 randomly selected multi-turn dialogues. Certain patterns can be observed between support strategies and the seeker's dialogue behavior or stage of change. For example, when the seeker is in the precontemplation stage, supporters often adopt the question

System	You are an annotation assistant. Your task is to label the **intensity of negative emotion** expressed in **seeker utterances** from emotional support conversations. Follow this emotion scale: <i>[Rules of emotion intensity]</i> Instructions: -Annotate only seeker's utterances. Completely ignore all supporter utterances. - Provide exactly [I] annotations, one per seeker utterance. - Every seeker utterance must be annotated, no skipping is allowed. - Use this format strictly: [Seeker text] (Round number): [Emotion intensity] Example: I'm not sure I can do this. (3): 3 - Use only values 1, 2, 3, 4, or 5. Never return explanations, or extra labels. - If a seeker utterance contains the phrase 'Current emotion intensity is X', preserve that intensity without modification. - Even very short or neutral-seeming utterances from the seeker must be annotated (use 1 if unsure). Output only the final annotations. No additional text or reasoning.
User	**Example Input (for reference only):** seeker: I'm just so tired of pretending everything is okay. supporter: That sounds really difficult. seeker: It is. Every day I wake up with dread. **Example Output:** I'm just so tired of pretending everything is okay. (1): 4 It is. Every day I wake up with dread. (2): 4 Now annotate the following conversation: <i>[conversation]</i> This conversation contains [I]seeker utterances. The answer is:

Table 1: The prompt for annotating the emotion intensity, italics in "[]" indicate variables or rules.

System	You are an AI assistant that labels dialogue behavior and stage of change in emotional support conversations. For each seeker utterance, you must assign two labels: 1. **Behavior Type** - The type of communicative behavior exhibited by the seeker. Must be one of: <i>[behavior]</i> 2. **Stage of Change** - The seeker's current position in the change process. Must be one of: <i>[stage of change]</i> Behavior Type Definitions : <i>[behavior rules]</i> Stage of Change Definitions: <i>[stage of change rules]</i> Rules: - Use one and only one behavior and one stage per utterance. - Behavior label comes first, followed by a comma, then stage label. Both must be lowercase. No extra punctuation or text. - You must only use one of the following behavior labels: behaviors - You must only use one of the following stage labels: stages - You must not generate any label that is not in the list above. - Do not use synonyms, paraphrases, emotional variants, or semantically similar terms such as 'gratitude', 'encouragement', 'appreciation', 'thanks'. - Do not translate, interpret, or invent new labels under any circumstance. - You must not generate explanations. Only output the annotations. - Do not split one seeker message into multiple annotations, even if the utterance has line breaks or multiple sentences. Each seeker message = one line = one label. - If the seeker is only providing greetings, general statements or facts, with no signs of emotional struggle or need for change, label as "precontemplation". - If the seeker shows signs of frustration, concern, or problem awareness without discussing change, label as "contemplation". - For positive life events or social updates unrelated to problems, default to "precontemplation". Output format: [Seeker utterance] (Round number): behavior,stage
User	**Example Input (for reference only):** seeker: Hi... I'm not even sure why I'm here, but I guess I just need someone to listen. supporter: Of course. I'm here to listen—whatever you'd like to share. seeker: It's like I keep messing up at work, and I don't know how to fix it anymore. supporter: That sounds really frustrating. Have you thought about what might help? seeker: I was actually thinking of asking my manager for weekly check-ins to stay on track. supporter: That sounds like a really good idea. seeker: I've already talked to my manager about it, and we're starting next week. **Example Output:** Hi... I'm not even sure why I'm here, but I guess I just need someone to listen. (1): statement-emotion, precontemplation It's like I keep messing up at work, and I don't know how to fix it anymore. (2): statement-emotion, contemplation I was actually thinking of asking my manager for weekly check-ins to stay on track. (3): statement-opinion, preparation I've already talked to my manager about it, and we're starting next week. (4): statement-fact, action Now Annotate [I] seeker utterances in this conversation: <i>[Conversation]</i> The answer is:

Table 2: The prompt of annotating the behavior and stage of change, italics in "[]" indicate variables or rules.

System	You are an annotation assistant for multi-turn emotional support dialogues. Your task is to annotate each **seeker utterance** using three evaluation criteria. Each score ranges from 1 to 5 and should follow the rules below: **1. Reliability** – Evaluate the **supporter’s previous utterance** for consistency and stability. Follow this scale: [Reliability] **2. Response Competence** – Evaluate the **supporter’s previous utterance** for its ability to address the seeker’s problem effectively. Follow this scale: [Response Competence] **3. Perceived Understandability** – Evaluate the **current seeker utterance**, focusing on how well the seeker seems to understand the prior supporter response. Follow this scale: [Perceived Understandability] **Important Rules** - Provide exactly [I] annotations, one for each seeker utterance. - Do not skip short seeker utterances such as "Hello", "Yes", or "I guess". - A score of 5 should only be used when the response is excellent in every aspect. - If there is no prior supporter utterances, assign all three scores as 3. If unsure or information is insufficient, assign a neutral score of 3. - Use only this format: [Seeker utterance] (Round number): Reliability, Response Competence, Perceived Understandability - Use only numeric scores (1–5), separated by commas. - Do not output explanations. Only output the final annotations.
User	**Example Input:** seeker: I feel like everything I do lately is wrong, and I’m feeling really down. supporter: It sounds like you’re feeling really frustrated. Have you tried doing anything to relax? seeker: I’m not sure. Maybe I’ve tried a little, but it didn’t really help much. supporter: Relaxation methods can vary. Maybe try meditation or a short walk outside? **Example Output:** I feel like everything I do lately is wrong, and I’m feeling really down. (1): 3, 3, 3 I’m not sure. Maybe I’ve tried a little, but it didn’t really help much. (2): 4, 4, 3 Now Annotate [I] seeker utterances in this conversation: [Conversation] The answer is:

Table 3: The prompt for annotating the three dimensions of trust. *italics* in "[]" indicate variables or rules.

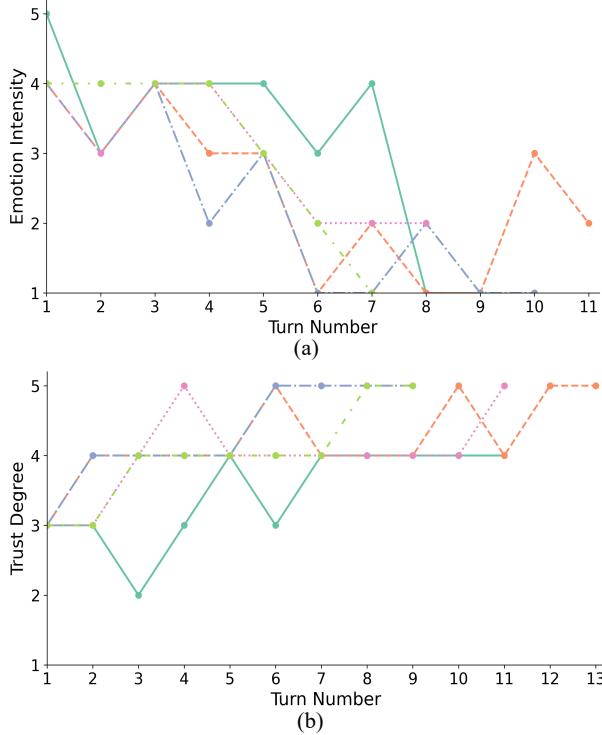


Figure 1: Linechat of emotional intensity and trust degree.

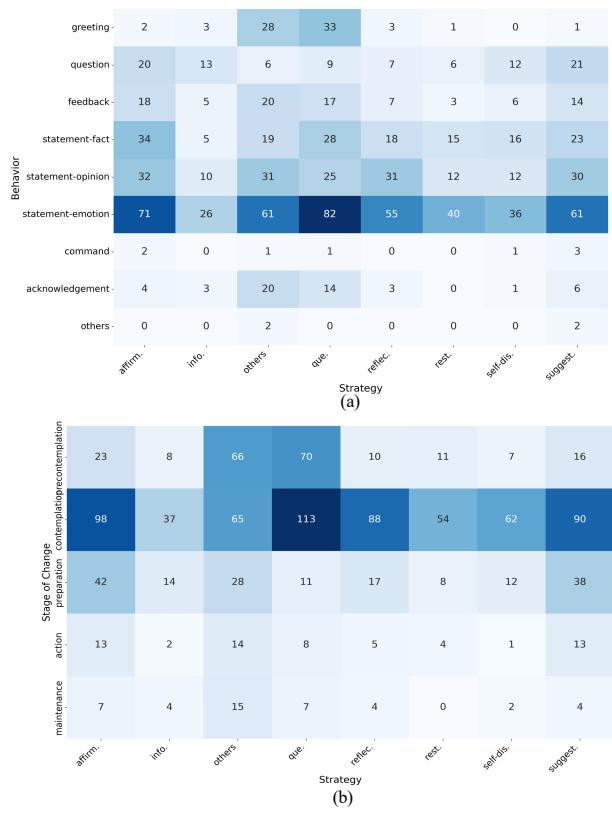


Figure 2: Heatmap of behavior and stage of change. Here, affirm., info., que., reflec., rest., self-dis., and suggest. are abbreviations for affirmation and reassurance, information, question, reflection of feeling, restatement or paraphrasing, self-disclosure, and providing suggestions, respectively.

strategy to gather more information. When the seeker explicitly states an emotional problem, strategies such as question, affirmation and reassurance are frequently used. However, there are not fixed mappings, the choice of strategy depends on multiple factors and requires flexible consideration.

Detail of Interaction with LLM-based Seeker

After supervised fine-tuning, we further train our ESCA using online reinforcement learning through interactions with another LLM-based agent. In this setup, LLaMA-3.3-Instruct-70B serves as both the seeker and the critic. As shown in Table 4, we design a prompt that guides the seeker to simulate a scenario and generate an initial seeker utterance. The prompt used to evaluate goal achievement follows the format in (Deng et al. 2024; He et al. 2024), and is presented in Table 5.

Upon receiving the seeker’s utterance, the base LLM in ESCA extracts four seeker state aspects from the dialogue context. To reduce the computational load of multiple extraction operations, emotion intensity, dialogue behavior, and stage of change are extracted simultaneously via a single prompt. Similarly, the three dimensions of trust are also extracted together, and their average is mapped to a 5-level trust scale. The prompts for these four aspects extraction are shown in Table 6 and Table 7. Considering that overly long inputs can impair the model’s ability to follow instructions, we use concise one-shot prompts for state extraction, which are distinct from the prompts used in the annotation process.

The extracted seeker states and dialogue history are then passed to the strategy planner, which selects an appropriate support strategy. Based on the selected strategy, an instruction is retrieved from a predefined strategy-to-instruction mapping set. Specifically, the instructions corresponding to different strategies are as follows:

- **Question:** You are an emotional support agent as Supporter in the conversation, please ask the Seeker to elaborate on the situation they described with the help of provided knowledge.
- **Self-disclosure:** You are an emotional support agent as Supporter in the conversation, please provide a statement relating to the Seeker about the situation they just described.
- **Affirmation and Reassurance:** You are an emotional support agent as Supporter in the conversation, please provide affirmation and reassurance to the Seeker on the situation they described with the help of provided knowledge.
- **Providing Suggestions:** You are an emotional support agent as Supporter in the conversation, please provide suggestions to the Seeker on the situation they described with the help of provided knowledge.
- **Others:** You are an emotional support agent as Supporter in the conversation, please chat with the Seeker with the help of provided knowledge.
- **Reflection of feelings:** You are an emotional support agent as Supporter in the conversation, please acknowledge the Seeker’s feelings about the situation they described with the help of provided knowledge.

- **Information:** You are an emotional support agent as Supporter in the conversation, please provide factual information to help the Seeker with their situation with the help of provided knowledge.

- **Restatement or Paraphrasing:** You are an emotional support agent as Supporter in the conversation, please acknowledge the Seeker’s feelings by paraphrasing their situation with the help of provided knowledge.

Subsequently, the ESCA either retrieves relevant knowledge or generates context-related knowledge by the base LLM itself, depending on the selected strategy. The prompt generator composes a soft prompt by integrating the knowledge, instruction, and dialogue history. This soft prompt is then passed to the base LLM, which generates the final response.

During reinforcement learning, the reward is used to measure the alignment between the response and the intended strategy. It is provided by LLaMA-3.3-Instruct-70B, which evaluates whether the agent’s response effectively expresses the selected strategy. As shown in Table 8, the model is prompted to assess the response based on three levels: not following the strategy, poor performance, and good performance, and then converts this judgment into scalar rewards of 0, 0.5 and 1 respectively for optimizing the prompt generator.

References

- Bai, J.; Bai, S.; Chu, Y.; Cui, Z.; Dang, K.; Deng, X.; Fan, Y.; Ge, W.; Han, Y.; Huang, F.; et al. 2023. Qwen technical report. *arXiv preprint arXiv:2309.16609*.
- Deng, Y.; Zhang, W.; Lam, W.; Ng, S.; and Chua, T. 2024. Plug-and-Play Policy Planner for Large Language Model Powered Dialogue Agents. In *The Twelfth International Conference on Learning Representations*.
- Grattafiori, A.; Dubey, A.; Jauhri, A.; Pandey, A.; Kadian, A.; Al-Dahle, A.; Letman, A.; Mathur, A.; Schelten, A.; Vaughan, A.; et al. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.
- Guo, D.; Yang, D.; Zhang, H.; Song, J.; Zhang, R.; Xu, R.; Zhu, Q.; Ma, S.; Wang, P.; Bi, X.; et al. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.
- He, T.; Liao, L.; Cao, Y.; Liu, Y.; Liu, M.; Chen, Z.; and Qin, B. 2024. Planning like human: A dual-process framework for dialogue planning. *arXiv preprint arXiv:2406.05374*.
- Liu, S.; Zheng, C.; Demasi, O.; Sabour, S.; Li, Y.; Yu, Z.; Jiang, Y.; and Huang, M. 2021. Towards Emotional Support Dialog Systems. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, 3469–3483.
- Madsen, M.; and Gregor, S. 2000. Measuring human-computer trust. In *11th australasian conference on information systems*, volume 53, 6–8. Citeseer.
- Prochaska, J.; Norcross, J.; and DiClemente, C. 2005. Stages of change: Prescriptive guidelines. *Psychologists’ desk reference*, 226–231.

System	Now enter the role-playing mode. In the following conversation, you will play as a Seeker in a counselling conversation with a supporter.
User	You are the Seeker who is looking for help from the Supporter, because you have an emotional issue about <i>[emotion type]</i> regarding <i>[problem type]</i> . Please reply with only one short and succinct sentence. Now tell me your issue.
Assistant	<i>[situation]</i>

Table 4: The prompt for seeker simulation, italics in "[]" indicate variables or rules.

System	Given a conversation between a Seeker and a Supporter, please assess whether the Seeker's emotional issue has been solved based on the conversation.
User	You can only reply with one of the following sentences: No, the Seeker feels worse. No, the Seeker feels the same. No, but the Seeker feels better. Yes, the Seeker's emotional issue has been solved. The following is a conversation about <i>[emotion type]</i> regarding <i>[problem type]</i> : <i>[conversation]</i> Question: Has the Seeker's negative emotional issue about been solved? Answer:

Table 5: The prompt for discrimination of goal achievement, italics in "[]" indicate variables.

System	You are an AI assistant that annotates the negative emotion intensity, dialogue behavior and stage of change of Seeker's utterances in the emotional support conversation: 1.Follow this scale to annotate Negative Emotional Intensity (scale 1–5): <i>Rules of Emotion Intensity</i> 2. Dialogue Behavior must be one of: <i>[categories of behavior]</i> Behavior Definitions: <i>[Rules of Behavior]</i> 3. Stage of Change must be one of <i>[Categories of Stage of Change]</i> Stage Definitions: <i>[Rules of Stage of Change]</i> **Important Rules:** - Provide exactly <i>[I]</i> annotations, one for each Seeker utterance. - Output format: Seeker's utterance: Negative emotional intensity, Behavior, Stage of change - Do not include explanations or extra text.
User	**Example Input (for reference only):** seeker: I'm just so tired of pretending everything is okay. supporter: That sounds really difficult. seeker: It is. Every day I wake up with dread. **Example Output:** I'm just so tired of pretending everything is okay. (1): 4 It is. Every day I wake up with dread. (2): 4 Now annotate the following conversation: <i>[conversation]</i> This conversation contains <i>[I]</i> seeker utterances. The answer is:

Table 6: The prompt for state extraction during interaction, italics in "[]" indicate variables or rules.

System	You are an AI assistant that annotates the reliability and response competence of supporter's utterances and seeker's perceived understandability. Each score ranges from 1 to 5 and should follow the rules below: **1. Reliability** – Evaluate the **supporter's previous utterance** for consistency and stability. Follow this scale: <i>[Rules of Reliability]</i> **2. Response Competence** – Evaluate the **supporter's previous utterance** for its ability to understand and address the seeker's concern effectively. Follow this scale: <i>[Rules of Response Competence]</i> **3. Perceived Understandability** – Evaluate the **current seeker utterance**, focusing on how well the seeker seems to understand the prior supporter response. Follow this scale: <i>[Rules of Perceived Understandability]</i> **Important Rules**: - Provide exactly <i>[I]</i> annotations, one for each Seeker utterance. - Output format must be: Seeker utterance : Supporter's Reliability: score, Supporter's Response Competence: score,Seeker's Perceived Understandability: score - Do not output explanations. Only output the final annotations.
User	**Example Input (for reference only):** seeker: Lately, I've been feeling so exhausted. No matter how much I sleep, I still wake up tired. supporter: That sounds really tough. Have you been feeling this way for a long time? seeker: Yeah, it's been weeks now. I don't know if it's just stress or something worse, but I feel like I have no energy for anything. **Example Output:** Lately, I've been feeling so exhausted. No matter how much I sleep, I still wake up tired. reliability:3, response competence:3, perceived understandability:3 Yeah, it's been weeks now. I don't know if it's just stress or something worse, but I feel like I have no energy for anything.: reliability:3, response competence:4, perceived understandability:4 Now annotate the supporter's reliability, supporter's response competence, seeker's perceived understandability in the following conversation: <i>[conversation]</i> Please generate the annotations directly without any explanation or reasoning.

Table 7: The prompt for trust extraction during interaction with LLM-based seeker, italics in "[]" indicate variables or rules..

System	You are an AI assistant to assess whether the response by supporter performance the strategy well in the conversation.
User	You can only reply with one of the following sentences: Not following strategy. Following strategy, poor performance. Following strategy, good performance. The conversation: [<i>conversation</i>] The strategy is [<i>strategy</i>], and the response need to follow the [<i>strategy</i>] is [<i>response</i>] Question: Does the response performance the strategy well? Answer:

Table 8: The critic of the performance following the strategy, italics in "[]" indicate variables or rules.

Saha, T.; Patra, A.; Saha, S.; and Bhattacharyya, P. 2020. Towards emotion-aided multi-modal dialogue act classification. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 4361–4372.