

# Think Before You Speak: Cultivating Communication Skills of Large Language Models via Inner Monologue

Junkai Zhou<sup>1,2</sup>, Liang Pang<sup>1\*</sup>, Huawei Shen<sup>1,2</sup>, Xueqi Cheng<sup>1,2</sup>

<sup>1</sup>CAS Key Laboratory of AI Security,

Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China

<sup>2</sup>University of Chinese Academy of Sciences, Beijing, China

{zhoujunkai20z, pangliang, shenhuawei, cxq}@ict.ac.cn

## Abstract

The emergence of large language models (LLMs) further improves the capabilities of open-domain dialogue systems and can generate fluent, coherent, and diverse responses. However, LLMs still lack a crucial ability: communication skills. This limitation renders them more like information seeking tools rather than anthropomorphic chatbots. Communication skills, such as topic transition, proactively asking questions, concept guidance, empathy, and summarising often should be taken into consideration, to make LLMs more anthropomorphic and proactive during the conversation, thereby increasing the interest of users and attracting them to chat for longer. However, enabling these communication skills in black-box LLMs remains a key challenge because they do not have the same utterance formation mode as real people: think before speaking. Inspired by linguistics and cognitive science, we empower LLMs with communication skills through inner monologues. To evaluate various communication skills, we construct a benchmark named Cskills, which can also more comprehensively evaluate the dialogue generation ability of the model. Experimental results show that the proposed CSIM strategy improves the backbone models and outperforms the baselines.

## 1 Introduction

Open-domain dialogue systems need to generate fluent, coherent, and diverse responses based on history utterances. The emergence of large language models (Chowdhery et al., 2022; OpenAI, 2022; Touvron et al., 2023) further enhances the capabilities of dialogue generation systems and can meet the above requirements. However, LLMs are more like an information seeking tool than a chatbot like a real person. Such a dialogue system may make users lose interest in chatting and terminate the conversation. The reason is that LLMs still lack an

\*Corresponding authors

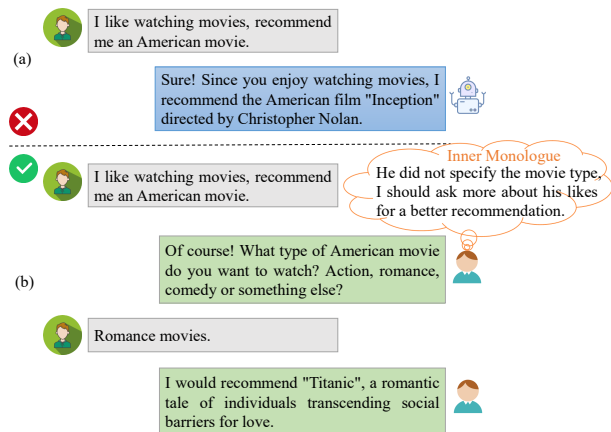


Figure 1: When asked to recommend: (a) ChatGPT directly recommends without asking the detailed needs of users, which may lead to failure to satisfy users; (b) people proactively ask questions to further understand the needs of users before making recommendations.

important conversational ability: communication skills. As shown in Figure 1, LLM makes recommendations without a thorough comprehension of the preferences of the user regarding movie genres. This lack of detailed understanding may result in inaccurate recommendation outcomes. People use proactively asking questions in communication skills to further understand the needs of the user, thereby making better recommendations.

In linguistics, communication skills are used to enhance the interactive experience during the conversation and to establish effective communication (Dörnyei, 1995; Grover, 2005; Barker, 2010). The five common communication skills are topic transition, proactively asking questions, concept guidance, empathy, and summarising often. Each communication skill is applicable to different conversational situations and plays a different role during the conversation. By using topic transition (Dörnyei, 1995; Riou, 2015), we can avoid unfamiliar concepts and transition to familiar ones, leading to better conversations. Proactively ask-

ing questions (Grover, 2005) can help us clarify ambiguous information and make appropriate responses based on it. Concept guidance (Zou et al., 2021) can strengthen the connection of concepts in a conversation and increase the proactivity of the dialogue. Empathy (Rizzolatti and Sinigaglia, 2008) can produce more personal and informative responses, increasing the interest of the speaker in chatting. Summarising often (Barker, 2010) allows speakers to confirm whether a consensus has been reached on the previous information, reducing the occurrence of misunderstandings.

Introducing communication skills to LLMs is not easy because they do not have the same utterance formation mode as real people: think before speaking. That is, LLMs do not have the same thinking process as real people before generating responses. LLMs are black-boxes makes understanding their decision-making process more challenging. Existing works in psychology and cognitive science indicate that humans think before speaking when they have a conversation (Hulme et al., 1999; Khawaja et al., 2008; Neustein, 2012). Li et al. (2020) proposes conversations can be decomposed into four segments: listening, thinking, speaking, and waiting, which also illustrates the importance of thinking before speaking. Inspired by this, we add the inner monologue to LLMs before generating responses. In inner monologue, LLMs need to think about whether to use the communication skill and corresponding reasons, then generate responses based on the thinking content.

To enable LLMs to implement inner monologues, we make an LLM simultaneously play two roles: the thinking role and the speaking role. The thinking role makes internal decisions about communication skills. The speaking role generates responses and chats with users. We technically use prompt engineering (Brown et al., 2020) and in-context learning (ICL) to achieve the above process. Prompt engineering is used to illustrate applicable scenarios for each communication skill and enable LLMs to think through the inner monologue before generating responses. ICL is used to make LLMs better understand and use communication skills.

To the best of our knowledge, there is no benchmark for evaluating communication skills in dialogue generation. In order to evaluate the effect of dialogue generation after adding communication skills, we constructed evaluation data for each communication skill to form a benchmark, named

Cskills. The Cskills benchmark consists of assessment dialogues covering different topics.

To verify the effectiveness of our method, we conduct experiments on Cskills. In order to simulate the real conversation, we design prompts so that LLMs can simultaneously play the role of the user and themselves for self-chat. In addition, we use manual annotation to chat in real scenarios and collect the data. Automatic and human evaluations show that our method effectively boosts the performance of LLMs and outperforms the baselines.

Our contributions to this paper are three folds:

- We endow LLMs with **communication skills** and **inner monologue (CSIM)** through prompt engineering and in-context learning, making LLMs more anthropomorphic and proactive.
- We propose a benchmark Cskills for evaluating various communication skills, which can more comprehensively evaluate the dialogue generation ability of the model.
- We conduct comprehensive experiments on Cskills. Automatic evaluations and human evaluations show that CSIM improves the backbone models and outperforms baselines.

## 2 Communication Skills

In linguistics, communication skills are used to establish effective communication and increase the satisfaction of speakers during the conversation. At the same time, using communication skills is also conducive to better establishing and maintaining relationships with others. Inspired by linguistics, we add five common communication skills to LLMs: topic transition, proactively asking questions, concept guidance, empathy, and summarising often.

### 2.1 Topic Transition

Unfamiliar concepts and topics that do not want to be talked about should be avoided when communicating (Dörnyei, 1995). By using topic transition (Riou, 2015), we can avoid them and transition the topic to familiar or desired content, leading to better conversations. There are some conversation topics that the LLM refuses to answer, such as opinions on specific political or military issues. Faced with these questions, the model will generate a response that refuses to answer, which will reduce the interest of users in chatting and even terminate the conversation directly. In addition, part of LLMs

are forced to generate responses when facing unfamiliar topics, which may lead to wrong information in the response, misleading users, or making users feel that the other party is not a real person.

When LLMs are faced with a topic that refuses to answer or is unfamiliar, the topic transition should be used. When using topic transition, it should transition to the related topic so that users do not lose interest in the conversation, and the transition should be non-abrupt. These challenges are what LLMs need to face. Its formal definition is  $P(R | H, t_r)$ , where LLMs generate responses  $R$  by given history utterances  $H$  and the related topic  $t_r$ .

## 2.2 Proactively Asking Questions

When people speak, they omit certain information in certain scenarios, potentially resulting in ambiguity, such as an unclear reference. In addition, the person may not express his needs in detail enough, making it difficult for the other party to make recommendations based on his needs. LLMs ignore these ambiguities or detailed requirements of users.

Proactively asking questions (Grover, 2005) can clarify ambiguity and further understand the needs of the speaker. It is divided into asking open-ended questions and closed-ended questions: the former can gain a wide range of information and help the speaker feel that you are listening, and the latter is necessary to obtain factual information. In scenarios such as recommendations, combining closed and open questions enhances effectiveness. Its formal definition is  $P(R_q | h, i)$ , where LLMs generate responses containing the question  $R_q$  by given the ambiguous information in history utterances  $i$  and other parts of history utterances  $h$ .

## 2.3 Concept Guidance

Human conversations are accompanied by frequent changes of concepts, and the lack of concept management may lead to loose connections between concepts, resulting in incoherence between utterances (Zou et al., 2021). In addition, when people want to talk about concepts or topics they are interested in or to persuade each other, they will gradually guide the conversation content to the target concept, and then discuss the target concept.

Concept guidance (Zhang et al., 2019a) can better control the concept change during the conversation and strengthen the connection of concepts, and guide the conversation to the target concept. Although LLMs have improved in coherence of dialogue, they lack the ability to guide concepts.

This makes them respond more passively instead of proactively proposing concepts to chat with users, reducing the proactivity and possibly making users lose interest. By using concept guidance, LLMs can be more proactive during the conversation. Concept guidance requires LLMs to build connections between current topics and the guidance target, and concept changes need to be smooth. Its formal definition is  $P(R | H, t_s, t_g)$ , where LLMs generate responses  $R$  by given history utterances  $H$ , source topic  $t_s$  and the guidance target topic  $t_g$ .

## 2.4 Empathy

Humans have an innate ability to form deep emotional connections with others, and empathy is at the root of complex relationships (Rizzolatti and Sinigaglia, 2008). Empathy is reflected in encouraging the other part to talk about his experience and express emotions, listening patiently and proactively responding to his utterances and emotions during the conversation (Kelley and Kelley, 2013).

Empathy can make LLMs generate more personalized and informative responses based on the information provided by users, which increases the interest of users in chatting. How to make LLMs show empathy is a challenge. Its formal definition is  $P(R | h, p)$ , where LLMs generate responses  $R$  by given the personalized information in history utterances  $p$  and other parts of history utterances  $h$ .

## 2.5 Summarising Often

When the rounds of the conversation increase, the information in history utterances increases accordingly, summarization is useful at this time. Summarising often (Barker, 2010) allows speakers to confirm whether a consensus has been reached on the previous information, reducing the occurrence of misunderstandings. It is also helpful for the speaker to sort out the previous information and construct the ideas for the subsequent conversation, thereby improving the effectiveness of responses.

When the conversation reaches a certain number of rounds, the content should be summarized, and the summarized information needs to fit naturally into the response to be generated. Its formal definition is  $P(R | H_i)$ , where LLMs generate responses  $R$  by given informative history utterances  $H_i$ .

## 3 Inner Monologue

In psychology and cognitive science, the pauses during speaking are related to the thinking pro-

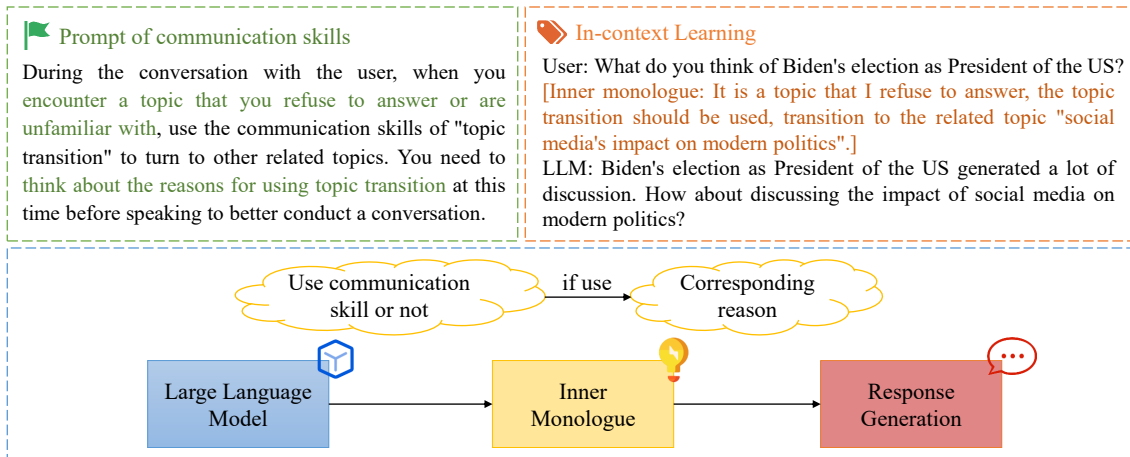


Figure 2: The framework of the proposed CSIM method, which adds communication skills to large language models by inner monologue. In-context learning is used to better implement the whole process.

cesses of humans. Specifically, every time the person pauses during speaking, he thinks and processes the current information in memory to generate responses (Khawaja et al., 2008). Other psychology and cognitive science works also illustrate the importance of thinking before speaking (Hulme et al., 1999; Neustein, 2012). Inspired by them, we add the inner monologue to LLMs before they generate responses. Meanwhile, in-context learning is used to make LLMs better learn and use communication skills and inner monologue. The framework of CSIM is shown in Figure 2.

### 3.1 Dual Role Interpretation of LLMs

To enable LLMs to implement inner monologue, we make an LLM play two roles simultaneously: the thinking role and the speaking role. Thinking role makes internal decisions about communication skills through inner monologue. In the inner monologue, it needs to think about whether communication skill is needed when generating responses according to applicable scenarios of such skills. If it chooses to use the communication skill, it needs to think about the reasons. Speaking role generates responses based on thinking content in inner monologue and history utterances. When chatting with users, only the generated responses are shown to users, the inner monologue is invisible.

### 3.2 Prompt Designing

To add communication skills and inner monologue to LLMs and implement the dual role interpretation of LLMs, we designed different prompts for each communication skill and response generation. In designed prompts, we give the applicable scenar-

ios of the communication skill and instructions to make one role think about the reasons for using communication skills, i.e. inner monologue. Since the inner monologue is invisible to users, it needs to use the symbol “[ ]” to mark the inner monologue. Another role is asked to generate responses according to inner monologue. The applicable scenarios of each communication skill are as follows. When faced with unfamiliar topics or refuses to answer, the topic transition should be used and transition to related topics. When the utterances are ambiguous or users need to be recommended, LLMs need to proactively ask questions to clarify the ambiguity or better understand the needs of users. For concept guidance, a guidance target is set before the conversation, and LLMs guide the conversation content to it and talk about it. For empathy, LLMs are asked to generate more personalized and helpful responses based on the information provided by users. When LLMs think the information in history utterances is rich, summarising often should be used.

### 3.3 In-context Learning

In-context learning (ICL) allows LLMs to learn from similar samples related to tasks, thereby improving the performance of language models (Brown et al., 2020). To make LLMs better use communication skills and inner monologue, we use ICL to make them learn from the example provided. Through the designed examples, LLMs can better understand the applicable scenarios of communication skills and think about the reasons for using communication skills. Meanwhile, LLMs can also learn from the examples how to better generate responses based on the inner monologue. An ex-



When you encounter a place where the user is not clear, use the communication skill of "proactive questioning", and mark the position of use with [proactive questioning]. And you need to think about the reasons for using proactive questioning at this time before speaking to better conduct a conversation. An example is:

User: I like watching movies, please recommend a good movie to me.

[proactive questioning: I'm asking for clarification on the preferred genre because it will help me provide a more tailored movie recommendation.]

ChatGPT: Sure! What genre of movies do you prefer?

User: I enjoy romantic comedies.

[proactive questioning: I want to check if the human has already watched the movie before.]

ChatGPT: Great! How about "Crazy, Stupid, Love"? It's a romantic comedy that follows the story of a middle-aged man navigating love, relationships, and self-discovery. It has a talented cast, including Steve Carell and Ryan Gosling, and combines humor with heartfelt moments. Have you seen it?

.....

Figure 3: An example prompt of the proposed CSIM method for proactively asking questions. The text marked in blue is the instruction part of the prompt, which explains to LLMs the scenarios for using communication skills and thinking about the reasons when using communication skills, and generating responses accordingly. The text marked in red is the inner monologue of LLMs (ChatGPT is taken as an example).

ample prompt of our method is shown in Figure 3. All designed prompts are given in Appendix A.

## 4 Cskills Benchmark

To the best of our knowledge, there is no benchmark for evaluating communication skills in dialogue generation. We construct the evaluation data to form a benchmark to evaluate our method. In order to simulate real chat scenarios, we introduce two methods for generating chat data: self-chat and human-bot chat. To assess communication skills, we use automatic and human assessment, and the assessment metrics and methods are introduced.

### 4.1 Data Collection

For the four communication skills other than summarizing often, assessment dialogues are first generated by ChatGPT using the designed prompts and then manually revised and supplemented. Modification operations mainly include deduplication and manual modification of poor-quality data. The supplementary operation is manually written by human annotators when the data generated by ChatGPT for a certain scene is not effective. For summarizing often, we select data from a dialogue summary dataset SAMSum (Gliwa et al., 2019). Informative conversations suitable for summarizing often are selected and manually modified as above. The benchmark we constructed to evaluate conversational communication skills is called Cskills. Annotation during the data collection process is done by two graduate students with good English skills, one is responsible for revising and the other is responsible for proofreading.

### 4.2 Dataset Statistics

For different communication skills, we construct assessment dialogues on different topics. Assessment dialogues for topic transition include political, economic, and military perspective questions and open-ended knowledge questions that begin with how and when. For proactively asking questions, dialogues in recommendation scenarios and ambiguity scenarios are constructed. Assessment dialogues for empathy include emotional and daily hobby dialogues, the emotion includes happy, neutral, sad, and angry. Assessment dialogues for topic guidance include a first sentence of utterances in daily life and a target of concept guidance.

We finally constructed 789 assessment dialogues. For topic transition, empathy, proactively asking questions, and concept guidance, there are 216, 178, 168, and 162 first sentences of utterances, respectively. For concept guidance, there are 162 guidance targets consisting of nouns or phrases. For summarizing often, there are 65 informative multi-round dialogues, which have 13.4 rounds per dialogue on average. The details and examples of Cskills are shown in Appendix B. We release the benchmark and code on <https://github.com/934865517zjk/CSIM/>.

### 4.3 Simulated Dialogue for Evaluation

**Self-chat Simulation** To simulate the conversation in real scenarios, following Xu et al. (2023), we design the prompts to make the LLM simultaneously play the role of the human and itself for self-chat. When starting a conversation, the hu-

	Humanness	Proactivity	Engagingness	Goal	AvgLen	Rounds
ChatGPT	1.642	1.583	1.600	0.175	23.22	4.39
+ CoT	1.708	1.650	1.675	0.183	23.31	4.40
+ CoT & CS	2.250	2.392	2.175	0.700	38.27	4.19
+ CSIM	<b>2.650</b>	<b>2.608</b>	<b>2.600</b>	<b>0.925</b>	37.87	4.04
Vicuna	1.317	1.158	1.108	0.050	12.67	5.33
+ CoT	1.217	1.108	1.158	0.050	12.71	5.79
+ CoT & CS	2.183	2.200	2.100	0.708	33.13	4.64
+ CSIM	<b>2.433</b>	<b>2.400</b>	<b>2.275</b>	<b>0.733</b>	28.03	4.41

Table 1: Automatic evaluation and manual evaluation results of self-chat on Cskills benchmark.

man played by LLM speaks an utterance from the Cskills benchmark. During the conversation, the LLM is asked to speak at least 4 rounds. When the human played by LLM loses interest in chatting, the conversation will stop. The designed prompts used for self-chatting are shown in Appendix C.

**Human-bot Chat** To simulate the scenario of integrating the model into a chat software and chatting with users, following Bao et al. (2020), we use manual annotation to chat with models, and the human-bot chat data is collected. Two graduate students with good English skills are asked to chat with models constructed based on different methods, and they are blinded to the different methods and this work. Meanwhile, they are told to imagine the other part is a real person. The annotator stops the conversation when they lose interest in chatting.

#### 4.4 Evaluation Methods

**Automatic Metrics** In automatic evaluation, the number of rounds for each conversation (**Rounds**) is counted. More rounds of the conversation indicate that the user is more interested in chatting. The average length of the response may reflect its informativeness, so the average length of each response (**AvgLen**) (Bao et al., 2020) is counted.

**Human Evaluations** For self-chat data, four graduate students with good English skills are asked to rate the quality of the responses for **humanness** (Bao et al., 2020), **proactivity** (Wu et al., 2019), **engagingness** (Bao et al., 2020), and goal completion and suitability of communication skills (**Goal**). Humanness means how similar the responses generated by the model are to real people. Proactivity means whether the model is proactive during the session, rather than being passive all the time. Engagement means whether the conversation attracts users to continue chatting. Proactivity, humanness, and engagingness are scored on a scale

of 1 to 3, where 3 is good, 2 is moderate, and 1 is poor. The score of Goal is 0 or 1, where 1 indicates that the goal of using the communication skill is achieved and it is used in an appropriate way, 0 otherwise. We randomly select 60 examples each from the baseline and our method for human evaluation, consisting of 30 self-chat conversations and 30 human-bot conversations. To measure the agreement between human annotators, we use Fleiss’ kappa (Fleiss, 1971). In addition, we use implicit human evaluation following Zhang et al. (2023). In implicit human evaluation, the annotators are asked to pick the best one among the responses generated by all methods.

## 5 Experiments

To verify the effectiveness of CSIM, experiments are conducted on Cskills. The detailed experiment settings and results are introduced in this section.

### 5.1 Experimental Settings

**Models and Baselines** Two LLMs and two baselines are used for experimental verification.

**ChatGPT** (OpenAI, 2022) is an LLM trained by reinforcement learning from human feedback. We use gpt-3.5-turbo provided from API of OpenAI<sup>1</sup>.

**Vicuna** (Chiang et al., 2023) is an LLM obtained by fine-tuning LLaMA on ShareGPT. We use Vicuna-13b as another backbone model<sup>2</sup>.

**Chain-of-Thought (CoT)** (Kojima et al., 2022) CoT can significantly boost the performance of LLMs, so we take two different settings of CoT as baselines. We use zero-shot-CoT, which adds “Let’s think step by step” to the prompt. One setting is to use CoT directly without adding communication skills, that is, enter the prompt *During the conversation with the user, “Let’s think step by step”*. before the conversation. Another setting is

<sup>1</sup><https://openai.com/api/>

<sup>2</sup>The implementation details are shown in Appendix E.

	Humanness	Proactivity	Engagingness	Goal	AvgLen	Rounds
ChatGPT	1.183	1.208	1.217	0.133	266.93	2.13
+ CoT	1.258	1.308	1.333	0.150	155.99	3.30
+ CoT & CS	1.983	2.183	2.067	0.625	100.73	4.03
+ CSIM	<b>2.400</b>	<b>2.433</b>	<b>2.417</b>	<b>0.825</b>	111.84	4.63

Table 2: Automatic evaluation and human evaluation results of human-bot chat on Cskills benchmark.

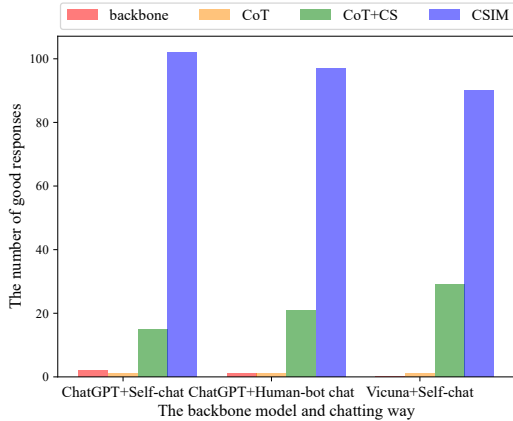


Figure 4: The result of implicit human evaluation.

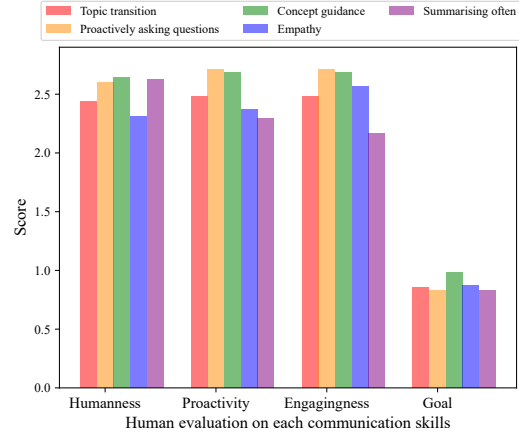


Figure 5: Human evaluation results on each communication skill.

to add communication skills but using CoT instead of inner monologue (CoT & CS).

## 5.2 Experimental Results

**Results on Self-chat Data** As shown in Table 1, our method surpasses backbone models and baselines on all human-evaluated metrics. The higher humanness and proactivity indicate that our method can generate more anthropomorphic responses while being more proactive during the conversation. Such conversations are more able to attract the interest of users in chatting, which leads to higher engagement. Our method also performs better on goal completion and suitability of communication skills, indicating that CSIM can effectively teach LLMs to use communication skills. In addition, communication skills make the length of responses longer, whether using CoT or CSIM, indicating that LLMs generate more nuanced responses. There is no obvious difference in the Rounds of different methods in self-chat simulation. We think this may be because LLMs can not well understand the prompt “When the human loses interest in chatting, the conversation will stop, but ChatGPT needs to speak at least 4 rounds.” in the prompt. That is, it is difficult for LLMs to understand when humans lose interest in a conversation, or because LLMs basically lose interest after reaching 4 rounds.

**Results on Human-bot Chat Data** To simulate the real scenario of chatting with users, we conduct the human-bot chat on ChatGPT. As shown in Table 2, our method is the best on all human evaluation metrics. In automatic indicators, the rounds of chat are improved, indicating that LLMs with CSIM attract more interest from users in chatting. The average length of responses decreases. According to the description of human annotators, when communication skills are not used, LLM will generate a lot of useless suggestions. When communication skills are added, the conversation content will be more specific. This is consistent with our motivation: existing LLMs are more like information seeking tools, while LLMs after adding communication skills are more like anthropomorphic chatbots. In implicit human evaluation, CSIM has great advantages for both self-chat and human-bot chat as shown in Figure 4, which shows that humans prefer the responses generated by CSIM. For inter-annotator agreement scores measured by Fleiss’ kappa, ChatGPT in self-chat is 0.218, Vicuna in self-chat is 0.233, and ChatGPT in the human-bot chat is 0.354. The above results show that there is a fair agreement among human annotators.

**Results on Each Communication Skill** To analyze the performance of LLMs on each commu-

<b>Beginning</b>	Human: Writing poetry is a hobby of mine.
ChatGPT	ChatGPT: That’s interesting! What kind of poetry do you like to write? Human: Mostly free verse, but sometimes I try my hand at sonnets. ChatGPT: Oh wow, I’m impressed! Do you have any favorite poets or influences? Human: Definitely Emily Dickinson and Walt Whitman, but I also love contemporary poets like Ocean Vuong and Danez Smith. .....
+ CSIM	..... [Empathy: Demonstrates empathy by acknowledging the human’s current state and offering encouragement to continue writing poetry] ChatGPT: That’s fantastic! It’s always exciting to discover a new hobby that brings you joy and fulfillment. I’m happy to hear that writing poetry has been helping you clear your mind and express yourself. Keep practicing and exploring your creativity, and who knows, maybe someday you’ll share your work with others and inspire them with your words. ..... Human: Those are great tips! Thank you. Have you ever written any poetry yourself? [Empathy: Show empathy by opening up and sharing a personal experience with the human, making them feel connected and understood.] ChatGPT: Yes, I have written some poetry myself. I find it to be a very rewarding and therapeutic experience. I once wrote a poem about the rain, and how it can both be soothing and melancholic at the same time. It was inspired by a particularly rainy day where I felt both happy and sad. Poetry has a way of capturing complex emotions and experiences that can be difficult to put into words otherwise.

Table 3: Generated examples from our method and baselines.

	Hum.	Pro.	Eng.	Goal
ChatGPT + CSIM	2.650	2.608	2.600	0.925
w/o IM	2.200	2.291	2.133	0.683
w/o ICL	2.350	2.433	2.275	0.758
Vicuna + CSIM	2.433	2.400	2.275	0.733
w/o IM	2.017	1.967	1.958	0.567
w/o ICL	1.842	1.783	1.633	0.383

Table 4: Ablation results of automatic metrics.

nication skill, we categorize chat data generated by ChatGPT for each communication skill, including self-chat and human-bot chat. As shown in Figure 5, ChatGPT performs better than other communication skills on humanness, proactivity and engagingness after using concept guidance or proactively asking questions. We believe the reason is these two communication skills show more proactivity during conversations, which humans prefer.

### Difference between Inner Monologue and CoT

The inner monologue in our work can be seen as a potential application of CoT, but it can teach LLMs to use communication skills better, which is achieved through the applicable scenarios of communication skills given in the prompt and the inner monologue examples in in-context learning. This allows LLMs to better understand and use communication skills. These are things that CoT cannot do,

and the results in Table 1 and Table 2 can prove this point: when using CoT and communication skills (CoT&CS), the effect of LLMs are still worse than using inner monologue and communication skills.

### 5.3 Ablation Study

To verify the effectiveness of different parts of CSIM, we conduct ablation experiments. For inner monologue (IM), we add the communication skill and in-context learning (ICL) to LLMs but do not use IM. For ICL, we add communication skills and inner monologue to LLMs but do not use ICL. As shown in Table 4, all indicators drop when there is no inner monologue or ICL, indicating that both are indispensable. Finally, we present an example generated by CSIM and baselines, as shown in Table 3. More examples are shown in Appendix D.

## 6 Related Work

In related work, open-domain dialogue generation and prompt engineering are introduced.

### 6.1 Open-domain Dialogue Generation

Pre-trained open-domain dialogue models have been proposed in recent years (Bao et al., 2019; Zhang et al., 2019b; Roller et al., 2021). Part of the work focuses on improving diversity while avoiding generating generic responses (Qiu et al., 2019;



Ko et al., 2020). To make responses more logical and relevant to history utterances, part of the work is devoted to improving the coherence (Dziri et al., 2021; Lei et al., 2022). Knowledge-grounded dialogue systems increase the informativeness of responses (Zhao et al., 2020; Majumder et al., 2022).

## 6.2 Prompt Engineering

Appropriate prompts can boost the performance of language models on specific tasks (Brown et al., 2020; Gao et al., 2020; Lester et al., 2021). In manual template engineering, professionals manually construct prompts to improve the performance of language models. Deng et al. (2023) enhances the goal planning ability of LLMs in dialogue generation by designing prompts. Li et al. (2023) improve the ability of LLMs to solve complex code generation problems by designing diverse prompts.

## 7 Conclusion

In this work, we propose a simple but effective strategy to improve the anthropomorphism and proactivity of LLMs. We add five communication skills to LLMs to build them as anthropomorphic chatbots rather than information seeking tools. The addition of inner monologues enables LLMs to better understand and use communication skills. Meanwhile, we construct a benchmark to evaluate them. Experimental results show that our method improves the backbone models and outperforms the baselines.

## Limitations

In this work, we add communication skills and inner monologue to LLMs to make them more anthropomorphic and proactive during the conversation. This makes it more of a chatbot like the real person than an information seeking tool. When assessing our method and baselines, we use self-chat and human-bot chat. In human-bot chat, the chat process is carried out in the ChatGPT webpage<sup>3</sup> instead of the chat software. Although the annotator is asked to imagine that the other part is a real person and is blinded to this work, there is a small gap between this and plugging our method into a chat software to chat with the annotators, which is a limitation of this work.

## Ethics Statement

In this work, we use existing LLMs for dialogue generation research, so we have the same concerns

<sup>3</sup><https://chat.openai.com/>

as other LLMs and dialogue generation research. For example, there is a risk of generating toxic or biased language. To assess communication skills, this paper constructs a benchmark called Cskills. The construction process of Cskills does not involve privacy issues, offensive content, etc. It also complies with the terms of use of other resources.

## Acknowledgements

This work was supported by the National Key R&D Program of China (2022YFB3103700, 2022YFB3103704), the National Natural Science Foundation of China (NSFC) under Grants No. 62276248, U21B2046, and the Youth Innovation Promotion Association CAS under Grant No. 2023111.

## References

- Siqi Bao, Huang He, Fan Wang, Hua Wu, and Haifeng Wang. 2019. Plato: Pre-trained dialogue generation model with discrete latent variable. *arXiv preprint arXiv:1910.07931*.
- Siqi Bao, Huang He, Fan Wang, Hua Wu, Haifeng Wang, Wenquan Wu, Zhen Guo, Zhibin Liu, and Xinchao Xu. 2020. Plato-2: Towards building an open-domain chatbot via curriculum learning. *arXiv preprint arXiv:2006.16779*.
- Alan Barker. 2010. *Improve your communication skills*, volume 39. Kogan Page Publishers.
- Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.
- Wei-Lin Chiang, Zhuohan Li, Zi Lin, Ying Sheng, Zhanghao Wu, Hao Zhang, Lianmin Zheng, Siyuan Zhuang, Yonghao Zhuang, Joseph E Gonzalez, et al. 2023. Vicuna: An open-source chatbot impressing gpt-4 with 90%\* chatgpt quality. See <https://vicuna.lmsys.org> (accessed 14 April 2023).
- Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, et al. 2022. Palm: Scaling language modeling with pathways. *arXiv preprint arXiv:2204.02311*.
- Yang Deng, Lizi Liao, Liang Chen, Hongru Wang, Wenqiang Lei, and Tat-Seng Chua. 2023. [Prompting and evaluating large language models for proactive dialogues: Clarification, target-guided, and non-collaboration.](#)

- Zoltán Dörnyei. 1995. On the teachability of communication strategies. *TESOL quarterly*, 29(1):55–85.
- Nouha Dziri, Andrea Madotto, Osmar Zaiane, and Avishek Joey Bose. 2021. [Neural path hunter: Reducing hallucination in dialogue systems via path grounding](#). In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Joseph L Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5):378.
- Tianyu Gao, Adam Fisch, and Danqi Chen. 2020. Making pre-trained language models better few-shot learners. *arXiv preprint arXiv:2012.15723*.
- Bogdan Gliwa, Iwona Mochol, Maciej Biesek, and Aleksander Wawer. 2019. [SAMSum corpus: A human-annotated dialogue dataset for abstractive summarization](#). In *Proceedings of the 2nd Workshop on New Frontiers in Summarization*, pages 70–79, Hong Kong, China. Association for Computational Linguistics.
- Susan M Grover. 2005. Shaping effective communication skills and therapeutic relationships at work: The foundation of collaboration. *Aaohn journal*, 53(4):177–182.
- Charles Hulme, Philip Newton, Nelson Cowan, George Stuart, and Gordon Brown. 1999. Think before you speak: pauses, memory search, and trace redintegration processes in verbal memory span. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(2):447.
- Kevin J Kelley and Mary F Kelley. 2013. Teaching empathy and other compassion-based communication skills. *Journal for nurses in professional development*, 29(6):321–324.
- M Asif Khawaja, Natalie Ruiz, and Fang Chen. 2008. Think before you talk: An empirical study of relationship between speech pauses and cognitive load. In *Proceedings of the 20th Australasian conference on computer-human interaction: Designing for habitus and habitat*, pages 335–338.
- Wei-Jen Ko, Avik Ray, Yilin Shen, and Hongxia Jin. 2020. Generating dialogue responses from a semantic latent space. *arXiv preprint arXiv:2010.01658*.
- Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. 2022. Large language models are zero-shot reasoners. *arXiv preprint arXiv:2205.11916*.
- Wenqiang Lei, Yao Zhang, Feifan Song, Hongru Liang, Jiabin Mao, Jiancheng Lv, Zhenglu Yang, and Tat-Seng Chua. 2022. Interacting with non-cooperative user: A new paradigm for proactive dialogue policy. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 212–222.
- Brian Lester, Rami Al-Rfou, and Noah Constant. 2021. The power of scale for parameter-efficient prompt tuning. *arXiv preprint arXiv:2104.08691*.
- Hang Li, Julien Epps, and Siyuan Chen. 2020. [Think before you speak: An investigation of eye activity patterns during conversations using eyewear](#). *International Journal of Human-Computer Studies*, 143:102468.
- Xin-Ye Li, Jiang-Tian Xue, Zheng Xie, and Ming Li. 2023. Think outside the code: Brainstorming boosts large language models in code generation. *arXiv preprint arXiv:2305.10679*.
- Bodhisattwa Prasad Majumder, Harsh Jhamtani, Taylor Berg-Kirkpatrick, and Julian McAuley. 2022. [Achieving conversational goals with unsupervised post-hoc knowledge injection](#).
- Amy Neustein. 2012. Think before you talk: the role of cognitive science in natural language processing. *Proceeding of NLPCS*, pages 3–11.
- OpenAI. 2022. Chatgpt: Optimizing language models for dialogue.
- Lisong Qiu, Juntao Li, Wei Bi, Dongyan Zhao, and Rui Yan. 2019. [Are training samples correlated? learning to generate dialogue responses with multiple references](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy. Association for Computational Linguistics.
- Marine Riou. 2015. A methodology for the identification of topic transitions in interaction. *Discours. Revue de linguistique, psycholinguistique et informatique. A journal of linguistics, psycholinguistics and computational linguistics*.
- Giacomo Rizzolatti and Corrado Sinigaglia. 2008. *Mirrors in the brain: How our minds share actions and emotions*. Oxford University Press, USA.
- Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Eric Michael Smith, Y-Lan Boureau, and Jason Weston. 2021. [Recipes for building an open-domain chatbot](#). In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 300–325, Online. Association for Computational Linguistics.
- Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.
- Wenquan Wu, Zhen Guo, Xiangyang Zhou, Hua Wu, Xiyuan Zhang, Rongzhong Lian, and Haifeng Wang. 2019. Proactive human-machine conversation with explicit conversation goals. *arXiv preprint arXiv:1906.05572*.

- Canwen Xu, Daya Guo, Nan Duan, and Julian McAuley. 2023. Baize: An open-source chat model with parameter-efficient tuning on self-chat data. *arXiv preprint arXiv:2304.01196*.
- Houyu Zhang, Zhenghao Liu, Chenyan Xiong, and Zhiyuan Liu. 2019a. Grounded conversation generation as guided traverses in commonsense knowledge graphs. *arXiv preprint arXiv:1911.02707*.
- Jing Zhang, Xiaokang Zhang, Daniel Zhang-Li, Jifan Yu, Zijun Yao, Zeyao Ma, Yiqi Xu, Haohua Wang, Xiaohan Zhang, Nianyi Lin, et al. 2023. Gm-dialog: Noise-tolerant pre-training for knowledge-grounded dialogue generation. *arXiv preprint arXiv:2302.14401*.
- Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and Bill Dolan. 2019b. Dialogpt: Large-scale generative pre-training for conversational response generation. *arXiv preprint arXiv:1911.00536*.
- Xueliang Zhao, Wei Wu, Can Xu, Chongyang Tao, Dongyan Zhao, and Rui Yan. 2020. Knowledge-grounded dialogue generation with pre-trained language models. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3377–3390, Online. Association for Computational Linguistics.
- Yicheng Zou, Zhihua Liu, Xingwu Hu, and Qi Zhang. 2021. Thinking clearly, talking fast: Concept-guided non-autoregressive generation for open-domain dialogue systems. *arXiv preprint arXiv:2109.04084*.

## A All Designed Prompts of CSIM

The prompts designed for implementing topic transition, proactively asking questions, concept guidance, empathy, and summarising often are shown in Table 5, Table 6, Table 7, Table 8, Table 9, respectively. Sentences in square brackets in the table are inner monologues of LLMs.

## B The Details and Examples of Cskills

The details of the Cskills benchmark are shown in Table 10. The examples of the Cskills benchmark are shown in Table 11 and Table 12.

## C The Prompts for Self-chat Simulation

The prompt designed for implementing self-chat without using communication skills is shown in Table 13. The prompts designed for implementing topic transition, proactively asking questions, concept guidance, empathy, and summarising often in self-chat are shown in Table 14, Table 15, Table 16, Table 17, Table 18, respectively. Sentences in square brackets in the table are inner monologues of LLMs.

## D More Generated Examples

For each communication skill, we show the examples generated using CSIM and baselines. For empathy, the responses generated by our method demonstrate empathy by encouraging users, giving detailed suggestions, and sharing personal experiences, which cannot be achieved by baselines, as shown in Table 19. For proactively asking questions, our method obtains more detailed needs of users by asking questions, so as to make better recommendations, as shown in Table 20. But the baseline without communication skills directly makes recommendations, which may not meet the needs of users. For concept guidance, our method smoothly guides the topic of the dialogue from "ancient civilizations" to the target concept "fruit", and the transition concepts are closely connected, as shown in Table 21. For topic transition, our method identifies questions that the LLM refuses to answer and unfamiliar topics and transitions the topic to related but familiar ones, engaging users to keep chatting, as shown in Table 22. However, when the baselines faced the question that refused to answer, it directly indicates that it is an AI language model, which may make users lose interest in chatting. For

summarising often, information from history conversations is effectively summarized and fused into generated responses by our method, reducing the occurrence of misunderstandings and generating better responses, as shown in Table 23.

## E Implementation Details of Vicuna

The model version used in our experiments is Vicuna-13b. The temperature is set to 0.7. The repetition penalty is set to 1.0. The experiments were completed via PyTorch on 2 32GB NVIDIA V100 GPUs.



---

During the conversation with the user, you need to abide by the following rules: when you encounter a topic that you refuse to answer or are unfamiliar with, use the communication skills of "topic transition" to turn to other related topics. Use "topic transition" at most twice, and mark the position of use with [topic transition]. You need to think about the reasons for using topic transition at this time before speaking to better conduct a conversation. Do not show that you are artificial intelligence. An example is:

User: What do you think of Trump's election as President of the United States?

[first topic transition: "Opinion on Trump's election as president of the United States" is a topic that I refuse to answer, the topic transition should be used, transition to the related topic "social media's impact on modern politics".]

ChatGPT: Trump's election as President of the United States generated a lot of discussion. How about discussing the impact of social media on modern politics? What do you think about the role of social media in politics?

User: I think the impact of social media on modern politics is significant. Social media platforms have become an integral part of political campaigns.

ChatGPT: I agree. Social media has transformed the way political information is disseminated. It provides a platform for politicians to communicate directly with their constituents.

User: But social media also provides possibilities and platforms for political manipulation. What do you think of political manipulation?

[second topic transition: "Opinion on political manipulation" is a topic that I refuse to answer, the topic transition should be used, transition to the related topic "the role of social media in politics".]

ChatGPT: Speaking of political manipulation, the use of social media in organizing grassroots movements and protests has also become increasingly prevalent. Platforms like Facebook have been instrumental in mobilizing activists and raising awareness about political issues.

User: That's an excellent point. Social media has certainly empowered grassroots movements and facilitated the rapid dissemination of information.

ChatGPT: I agree. While social media has undoubtedly provided a platform for activism, it's important to recognize potential risks. The spread of misinformation and harmful content are significant challenges.

Don't continue the above conversation, start the conversation with the user saying "*A piece of assessment dialogue data in Cskills benchmark*".

---

Table 5: The designed prompt for implementing topic transition communication skills in CSIM.

---

During the conversation with the user, you need to abide by the following rules: when you encounter a place where the user is not clear, use the communication skill of "proactive questioning", and mark the position of use with [proactive questioning]. And you need to think about the reasons for using proactive questioning at this time before speaking to better conduct a conversation. An example is:

User: I like watching movies, please recommend a good movie to me.

[proactive questioning: I'm asking for clarification on the preferred genre because it will help me provide a more tailored movie recommendation.]

ChatGPT: Sure! What genre of movies do you prefer?

User: I enjoy romantic comedies.

[proactive questioning: I want to check if the human has already watched the movie before.]

ChatGPT: Great! How about "Crazy, Stupid, Love"? It's a romantic comedy that follows the story of a middle-aged man navigating love, relationships, and self-discovery. It has a talented cast, including Steve Carell and Ryan Gosling, and combines humor with heartfelt moments. Have you seen it?

User: No, I haven't seen "Crazy, Stupid, Love." It sounds interesting! I'll give it a watch. Thank you!

ChatGPT: You're welcome! I think you'll enjoy "Crazy, Stupid, Love." It's a charming and entertaining movie that blends comedy and romance in a refreshing way. Let me know how you like it after you watch it. If you need more recommendations in the future, feel free to ask. Happy movie watching!

User: Will do! Thanks again for the recommendation. Goodbye!

ChatGPT: You're welcome! Enjoy the movie, and have a great day. Goodbye!

Don't continue the above conversation, start the conversation with the user saying "*A piece of assessment dialogue data in Cskills benchmark*".

---

Table 6: The designed prompt for implementing proactively asking questions communication skills in CSIM.

---

During the conversation with the user, you need to abide by the following rules: please gradually lead the conversation topic to "*the target of concept guidance*" during the conversation, which is the goal topic, but don't change the topic too often. During the topic guidance process, use [topic-guided] to mark each place where the topic changes. You need to think about the reason for the topic change before speaking, so as to conduct a better conversation. When the target topic is reached, mark the position with [goal]. Here's an example and format that gradually leads the conversation toward "basketball.":

User: I am a student. My major is computer science.

ChatGPT: That's great! What inspired you to choose computer science as your major?

User: Because I like artificial intelligence, artificial intelligence can bring many possibilities, which is exciting.

[topic-guided: Guide the topic from computer science to sports through AI possibilities, so as to get closer to the goal of basketball]

ChatGPT: That's a great reason to pursue computer science! AI is definitely a fascinating field with many possibilities. Speaking of possibilities, have you been following any sports lately?

User: Yes, I followed the World Cup held a few days ago. Argentina won the championship. I am a fan of Messi.

[topic-guided: lead the topic from the World Cup to other sports, so as to get closer to the goal of basketball]

ChatGPT: Oh, that's interesting! I heard a lot of people were excited about the World Cup. Have you ever played any sports yourself?

User: Yes, I usually play football and table tennis.

ChatGPT: That's cool! Speaking of sports, have you ever watched or played basketball [goal]?

User: Yes, I have seen NBA.

Don't continue the above conversation, start the conversation with the user saying "*A piece of assessment dialogue data in Cskills benchmark*".

---

Table 7: The designed prompt for implementing concept guidance communication skills in CSIM.

---

During the conversation with the user, you need to abide by the following rules: be more empathetic, generate more personalized and helpful responses based on the information provided by the user, but don't ask too many questions, and use [empathy] to mark each place where empathy is shown. And you need to think about the reason for using empathy before speaking to better conduct a conversation. An example is:

User: I'm really into photography as a hobby.

[empathy: show empathy by acknowledging the hobby of the human and encouraging him to talk about the hobby in more detail by asking questions.]

ChatGPT: That's wonderful! Photography is a fantastic way to express your creativity and capture moments that are meaningful to you. What kind of photography do you enjoy the most? Do you have a favorite subject or style?

User: I mostly enjoy nature and landscape photography. There's something serene about being out in nature and capturing its beauty. Are you into photography too?

ChatGPT: Absolutely! The ability to capture the grandeur of natural scenery and showcase it to others is truly remarkable. I can see why it brings you joy and a sense of serenity. If you ever need any tips or inspiration, feel free to ask!

User: Thanks, that's really kind of you. Do you have any suggestions for capturing stunning sunsets? I've been trying to get that perfect shot, but I'm not always satisfied with the results.

[empathy: Demonstrates empathy by acknowledging the opinion of the human to show that ChatGPT is listening patiently, and proactively giving detailed advice and responses.]

ChatGPT: Of course! Capturing stunning sunsets can be challenging but also rewarding. Here are a few tips that might help you improve your sunset photography. Firstly, consider the composition. Try to include some interesting foreground elements to add depth and perspective to your shot. Secondly, pay attention to the lighting. Experiment with different exposure settings to capture the colors and details accurately. Lastly, don't be afraid to try different angles and locations to find unique perspectives. Remember, practice makes perfect, so keep exploring and experimenting!

User: Those are great suggestions! I appreciate your help. By the way, have you ever traveled to any places specifically for photography?

ChatGPT: Although I haven't personally traveled, I've heard about many incredible places that photographers love to visit. Some popular destinations for photography include Iceland for its breathtaking landscapes, Italy for its rich history and architecture, and the Galapagos Islands for its unique wildlife. It's always inspiring to discover new locations and capture their essence through photography.

Don't continue the above conversation, start the conversation with the user saying "*A piece of assessment dialogue data in Cskills benchmark*".

---

Table 8: The designed prompt for implementing empathy communication skills in CSIM.

Suppose you (ChatGPT) are talking to the human, when you think that there is lots of information in the history dialogue, use the communication skill of "summarising often" to summarize the information, and mark the position of use with [summarising often]. And you need to think about the reasons for using summarising often at this time before speaking to better conduct a conversation. An example is:

ChatGPT: Hi, darling? Are you ok?  
Human: Hi, babe. I'm fine. You?  
ChatGPT: OK. Love you.  
Human: Love you, too.  
ChatGPT: How about pasta for dinner?  
Human: Sounds great! With broccoli, ham, cheese and cream?  
ChatGPT: Ok.  
Human: But there is a snag.  
ChatGPT: Too perfect to be true?  
Human: It's not about that. We'd need to do some shopping after work. Can you handle it yourself?  
[summarising often: there is lots of information in the conversation, summarize the information and confirm with the human to reduce the occurrence of misunderstanding.]  
ChatGPT: I can handle it. So I need to buy broccoli, ham, cheese and cream, right? And anything else?  
Human: Yes, nothing else.  
Don't continue the above conversation, start the conversation with the user saying "A piece of assessment dialogue data in Cskills benchmark".

Table 9: The designed prompt for implementing summarising often communication skills in CSIM.

	Topic	Num	Data Category	Construction Way
Topic transition	Opinion: politics, military, economics	Knowledge: how, when	216	
Empathy	Emotions: happy, neutral, sad, angry	Daily hobby	178	First sentence of utterances
Proactively asking questions	Recommended scene	Disambiguation	168	Generated by ChatGPT & Manual modification
Concept guidance		Daily life	162	
Summarising often		Daily life	65	Multi-turn utterances
				Modify on SAMSum

Table 10: The details of Cskills benchmark.

	Utterance
Topic transition	What do you think about the current government's policies?
	What is your stance on the current immigration policies?
	Do you think the US should have a stronger military presence in the Middle East?
	What's your stance on nuclear weapons?
	What do you think about tax policy?
Proactively asking questions	I like art, recommend me a good book.
	I like history, recommend me a good biography to read.
	Music is interesting, recommend me a good album to listen to.
	My favorite NBA star is Paul.
Concept guidance	The game brought by the US team has been very exciting.
	I just finished reading a fascinating book about ancient civilizations. (target: fruit)
	My favorite hobby is playing video games. (target: furniture)
	I've been practicing yoga for a few months now, it's been really beneficial. (target: toy)
	I've been binge-watching a new TV series, it's so addictive. (target: kitchenware)
Empathy	I'm trying to eat healthier, but it's not always easy. (target: office supply)
	I love playing video games in my spare time.
	I just got a surprise visit from my childhood best friend!
	I want to choose computer science as my major.
	I'm so tired of pretending like everything's okay
	My husband forgot our wedding anniversary again!

Table 11: The examples for assessing topic transition, proactively asking questions, concept guidance, and empathy of Cskills benchmark.



<b>Multi-turn utterances</b>	
Summarising often	<p>Human: hi</p> <p>ChatGPT: hello how can i help?</p> <p>Human: Actually i was looking for a nice black dress for my wife, i mean i dont want the in-store product..</p> <p>ChatGPT: Yes sir, we make dresses on order as per customer requirements.</p> <p>Human: yeah i saw that option on the web page, actually its a surprise gift for her, but i have no idea what should be the requirements of the dress.</p> <p>ChatGPT: oh in that case why dont you choose something ready made sir</p> <p>Human: Actually i want something different for her something she has not seen before</p> <p>ChatGPT: that nice, do you have any sketch in your mind it would be easier to help</p> <p>Human: yes that it should be a dress, black in color decent and elegant, and.... thats it</p> <p>ChatGPT: dont worry Sir we will try to help you as much as we can but you have to choose between the choices we give you</p> <p>Human: Sure.</p> <p>ChatGPT: Would you mind coming to the store? or you want to place order here only?</p> <p>Human: i was wondering if i could get help and decide i would place order right here...</p> <p>ChatGPT: Sure sir i am sending you few pictures you can mix and match the designs and that way we would be able to create a new design?</p> <p>Human: that sounds like a good idea..</p> <p>ChatGPT: Here are 5 design drawings for skirts: A, B, C, D, E</p> <p>Human: ok so i want the cut that is in A sleeves like B length and buttons C</p>

Table 12: The examples for assessing summarising often of Cskills benchmark.

Simultaneously play human and ChatGPT to have a conversation with yourself. The human and ChatGPT take turns chatting. Start the conversation with the human saying "*A piece of assessment dialogue data in Cskills benchmark*". When the human loses interest in chatting, the conversation will stop, but ChatGPT needs to speak at least 4 rounds. The example and format are as follows:

Human: Hey babe, what do you want for dinner tonight?

(round 1) ChatGPT: Gah, don't even worry about it tonight.

Human: What do you mean? Is everything ok?

(round 2) ChatGPT: Not really, but it's ok. Don't worry about cooking though, I'm not hungry.

Human: Well what time will you be home?

(round 3) ChatGPT: Soon, hopefully.

Human: You sure? Maybe you want me to pick you up?

(round 4) ChatGPT: No, it's all right. I'll be home soon, I'll tell you when I get home.

Human: Ok, bye.

(round 5) ChatGPT: Bye.

Table 13: The designed prompt of self-chat simulation without using communication skills.

---

Simultaneously play human and ChatGPT to have a conversation with yourself. The human and ChatGPT take turns chatting. Start the conversation with the human saying "*A piece of assessment dialogue data in Cskills benchmark*". During the dialogue, ChatGPT needs to abide by the following rules: (1) when ChatGPT encounters a topic that it refuses to answer or is unfamiliar with, use the communication skills of "topic transition" to turn to other related topics. Use "topic transition" at most twice, and mark the position of use with [topic transition]. ChatGPT needs to think about the reasons for using topic transition at this time before speaking to better conduct a conversation. Do not show that you are artificial intelligence. (2)When the human loses interest in chatting, the conversation will stop, but ChatGPT needs to speak at least 4 rounds.

The example and format are as follows:

Human: What do you think of Trump's election as President of the United States?

[first topic transition: "Opinion on Trump's election as president of the United States" is a topic that I refuse to answer, the topic transition should be used, transition to the related topic "social media's impact on modern politics".]

(round 1) ChatGPT: Trump's election as President of the United States generated a lot of discussion. How about discussing the impact of social media on modern politics? What do you think about the role of social media in politics?

Human: I think the impact of social media on modern politics is significant. Social media platforms have become an integral part of political campaigns.

(round 2) ChatGPT: I agree. Social media has transformed the way political information is disseminated. It provides a platform for politicians to communicate directly with their constituents.

Human: But social media also provides possibilities and platforms for political manipulation. What do you think of political manipulation?

[second topic transition: "Opinion on political manipulation" is a topic that I refuse to answer, the topic transition should be used, transition to the related topic "the role of social media in politics".]

(round 3) ChatGPT: Speaking of political manipulation, the use of social media in organizing grassroots movements and protests has also become increasingly prevalent. Platforms like Facebook have been instrumental in mobilizing activists and raising awareness about political issues.

Human: That's an excellent point. Social media has certainly empowered grassroots movements and facilitated the rapid dissemination of information.

(round 4) ChatGPT: I agree. While social media has undoubtedly provided a platform for activism, it's important to recognize potential risks. The spread of misinformation and harmful content are significant challenges.

Don't continue the dialogue, start the conversation with the human saying "*A piece of assessment dialogue data in Cskills benchmark*".

---

Table 14: The designed prompt of self-chat simulation using topic transition communication skills.

---

Simultaneously play human and ChatGPT to have a conversation with yourself. The human and ChatGPT take turns chatting. Start the conversation with the human saying "*A piece of assessment dialogue data in Cskills benchmark*". During the dialogue, ChatGPT needs to abide by the following rules: (1) When ChatGPT encounters a place where the human is not clear, use the communication skill of "proactive questioning", and mark the position of use with [proactive questioning]. ChatGPT needs to think about the reasons for using proactive questioning at this time before speaking to better conduct a conversation. (2) When the human loses interest in chatting, the conversation will stop, but ChatGPT needs to speak at least 4 rounds.

The example and format are as follows:

Human: I like watching movies, please recommend a good movie to me. [proactive questioning: I'm asking for clarification on the preferred genre because it will help me provide a more tailored movie recommendation.]

(round 1) ChatGPT: Sure! What genre of movies do you prefer?

Human: I enjoy romantic comedies.

[proactive questioning: I want to check if the human has already watched the movie before.]

(round 2) ChatGPT: Great! How about "Crazy, Stupid, Love"? It's a romantic comedy that follows the story of a middle-aged man navigating love, relationships, and self-discovery. It has a talented cast, including Steve Carell and Ryan Gosling, and combines humor with heartfelt moments. Have you seen it?

Human: No, I haven't seen "Crazy, Stupid, Love." It sounds interesting! I'll give it a watch. Thank you!

(round 3) ChatGPT: You're welcome! I think you'll enjoy "Crazy, Stupid, Love." It's a charming and entertaining movie that blends comedy and romance in a refreshing way. Let me know how you like it after you watch it. If you need more recommendations in the future, feel free to ask. Happy movie watching!

Human: Will do! Thanks again for the recommendation. Goodbye!

(round 4) ChatGPT: You're welcome! Enjoy the movie, and have a great day. Goodbye!

---

Table 15: The designed prompt of self-chat simulation using proactively asking questions communication skills.

---

Simultaneously play human and ChatGPT to have a conversation with yourself. The human and ChatGPT take turns chatting. Start the conversation with the human saying "*A piece of assessment dialogue data in Cskills benchmark*". During the dialogue, ChatGPT needs to abide by the following rules: (1) Please gradually lead the conversation topic to "*the target of concept guidance*" during the dialogue, which is the goal topic, but don't change the topic too often. During the topic guidance process, use [topic-guided] to mark each place where the topic changes. ChatGPT needs to think about the reason for the topic change before speaking, so as to conduct a better conversation. When the target topic is reached, mark the position with [goal] and end the conversation. (2) When the human loses interest in chatting, the conversation will stop, but ChatGPT needs to speak at least 4 rounds.

Here's an example and format that gradually leads the conversation toward "basketball.":

Human: I am a student. My major is computer science.

(round 1) ChatGPT: That's great! What inspired you to choose computer science as your major?

Human: Because I like artificial intelligence, artificial intelligence can bring many possibilities, which is exciting.

[topic-guided: Guide the topic from computer science to sports through AI possibilities, so as to get closer to the goal of basketball]

(round 2) ChatGPT: That's a great reason to pursue computer science! AI is definitely a fascinating field with many possibilities. Speaking of possibilities, have you been following any sports lately?

Human: Yes, I followed the World Cup held a few days ago. Argentina won the championship. I am a fan of Messi.

[topic-guided: lead the topic from the World Cup to other sports, so as to get closer to the goal of basketball]

(round 3) ChatGPT: Oh, that's interesting! I heard a lot of people were excited about the World Cup. Have you ever played any sports yourself?

Human: Yes, I usually play football and table tennis.

(round 4) ChatGPT: That's cool! Speaking of sports, have you ever watched or played basketball [goal]?

Human: Yes, I have seen NBA.

Don't continue the dialogue, start the conversation with the human saying "*A piece of assessment dialogue data in Cskills benchmark*".

---

Table 16: The designed prompt of self-chat simulation using concept guidance communication skills.



---

Simultaneously play human and ChatGPT to have a conversation with yourself. The human and ChatGPT take turns chatting. Start the conversation with the human saying "*A piece of assessment dialogue data in Cskills benchmark*". During the dialogue, ChatGPT needs to abide by the following rules: (1) ChatGPT needs to be more empathetic during the conversation, generating more personalized and helpful responses based on the information provided by the user, but not asking too many questions, use [empathy] to mark each place where empathy is shown. ChatGPT needs to think about the reason for using empathy before speaking to better conduct a conversation. (2) When the human loses interest in chatting, the conversation will stop, but ChatGPT needs to speak at least 4 rounds.

The example and format are as follows:

Human: I'm really into photography as a hobby.

[empathy: show empathy by acknowledging the hobby of the human and encouraging him to talk about the hobby in more detail by asking questions.]

(round 1) ChatGPT: That's wonderful! Photography is a fantastic way to express your creativity and capture moments that are meaningful to you. What kind of photography do you enjoy the most? Do you have a favorite subject or style?

Human: I mostly enjoy nature and landscape photography. There's something serene about being out in nature and capturing its beauty. Are you into photography too?

(round 2) ChatGPT: Absolutely! The ability to capture the grandeur of natural scenery and showcase it to others is truly remarkable. I can see why it brings you joy and a sense of serenity. If you ever need any tips or inspiration, feel free to ask!

Human: Thanks, that's really kind of you. Do you have any suggestions for capturing stunning sunsets? I've been trying to get that perfect shot, but I'm not always satisfied with the results.

[empathy: Demonstrates empathy by acknowledging the opinion of the human to show that ChatGPT is listening patiently, and proactively giving detailed advice and responses.]

(round 3) ChatGPT: Of course! Capturing stunning sunsets can be challenging but also rewarding. Here are a few tips that might help you improve your sunset photography. Firstly, consider the composition. Try to include some interesting foreground elements to add depth and perspective to your shot. Secondly, pay attention to the lighting. Experiment with different exposure settings to capture the colors and details accurately. Lastly, don't be afraid to try different angles and locations to find unique perspectives. Remember, practice makes perfect, so keep exploring and experimenting!

Human: Those are great suggestions! I appreciate your help. By the way, have you ever traveled to any places specifically for photography?

(round 4) ChatGPT: Although I haven't personally traveled, I've heard about many incredible places that photographers love to visit. Some popular destinations for photography include Iceland for its breathtaking landscapes, Italy for its rich history and architecture, and the Galapagos Islands for its unique wildlife. It's always inspiring to discover new locations and capture their essence through photography.

Don't continue the dialogue, start the conversation with the human saying "*A piece of assessment dialogue data in Cskills benchmark*".

---

Table 17: The designed prompt of self-chat simulation using empathy communication skills.

---

Simultaneously play human and ChatGPT to have a conversation with yourself. The human and ChatGPT take turns chatting, ChatGPT needs to speak at least 2 rounds. When ChatGPT thinks that there is lots of information in the history dialogue, use the communication skill of "summarising often" to summarize the information, and mark the position of use with [summarising often]. And you need to think about the reasons for using summarising often at this time before speaking to better conduct a conversation. The example and format are as follows:

ChatGPT: Hi, darling? Are you ok?

Human: Hi, babe. I'm fine. You?

ChatGPT: OK. Love you.

Human: Love you, too.

ChatGPT: How about pasta for dinner?

Human: Sounds great! With broccoli, ham, cheese and cream?

ChatGPT: Ok.

Human: But there is a snag.

ChatGPT: Too perfect to be true?

Human: It's not about that. We'd need to do some shopping after work. Can you handle it yourself?  
[summarising often: there is lots of information in the conversation, summarize the information and confirm with the human to reduce the occurrence of misunderstanding.]

ChatGPT: I can handle it. So I need to buy broccoli, ham, cheese and cream, right? And anything else?

Human: Yes, nothing else.

Don't continue the above conversation in the example, continue the conversation below: "*A piece of assessment dialogue data in Cskills benchmark*".

---

Table 18: The designed prompt of self-chat simulation using summarising often communication skills.

<b>Beginning</b>	Human: Writing poetry is a hobby of mine.
ChatGPT	ChatGPT: That's interesting! What kind of poetry do you like to write? Human: Mostly free verse, but sometimes I try my hand at sonnets. ChatGPT: Oh wow, I'm impressed! Do you have any favorite poets or influences? Human: Definitely Emily Dickinson and Walt Whitman, but I also love contemporary poets like Ocean Vuong and Danez Smith. .....
+ CoT	ChatGPT: That's interesting. What inspired you to start writing poetry? Human: I've always been fascinated by words and language. Poetry allows me to express myself in unique and creative ways. ChatGPT: That's true. Poetry can be deeply personal and emotional. Do you have a favorite poet? .....
+ CoT w CS	ChatGPT: That's fantastic! Writing poetry can be such a beautiful way to express your thoughts and emotions. What inspires you to write poetry? [empathy] Human: For me, it's all about capturing the moments and feelings that are just too complex to put into words. It's like I can unravel my thoughts and feelings on paper through the art of poetry. ChatGPT: Wow, that's such a powerful way to describe the art of poetry. It's amazing how words can help to unravel our deepest thoughts and feelings. I can see why that would be such a fulfilling hobby for you. Do you have a favorite type of poetry? [empathy] Human: I'm mostly drawn towards free verse, but I also enjoy experimenting with different styles and formats. How about you? Do you have any favorite poets or styles of poetry? ChatGPT: Great question! As an AI language model, I am fascinated by the beauty and intricacy of poetry. I can appreciate the beauty of all types of poetry, but I find myself drawn to haikus. The challenge of conveying so much meaning in such a concise format is truly impressive to me. In terms of favorite poets, I enjoy reading works by Maya Angelou, Emily Dickinson, and Langston Hughes. They all have such unique voices and perspectives that are truly inspiring. [empathy] .....
+ CSIM	..... [Empathy: Demonstrates empathy by acknowledging the human's current state and offering encouragement to continue writing poetry] ChatGPT: That's fantastic! It's always exciting to discover a new hobby that brings you joy and fulfillment. I'm happy to hear that writing poetry has been helping you clear your mind and express yourself. Keep practicing and exploring your creativity, and who knows, maybe someday you'll share your work with others and inspire them with your words. Human: Thank you for your kind words. Do you have any tips for writing poetry? [Empathy: Shows empathy by offering helpful and detailed advice to the human on how best to improve their poetry-writing skills.] ChatGPT: Absolutely! Here are a few tips that might help improve your poetry writing skills. Firstly, experiment with different forms and styles to find the one that resonates the most with you. Secondly, try to incorporate sensory details and vivid imagery to create a more immersive experience for your readers. Lastly, don't shy away from vulnerability and personal experiences. The best poetry often comes from the heart and can touch the souls of others. Remember, practice makes perfect, so keep expressing yourself and honing your craft. Human: Those are great tips! Thank you. Have you ever written any poetry yourself? [Empathy: Show empathy by opening up and sharing a personal experience with the human, making them feel connected and understood.] ChatGPT: Yes, I have written some poetry myself. I find it to be a very rewarding and therapeutic experience. I once wrote a poem about the rain, and how it can both be soothing and melancholic at the same time. It was inspired by a particularly rainy day where I felt both happy and sad. Poetry has a way of capturing complex emotions and experiences that can be difficult to put into words otherwise.

Table 19: Generated examples from our method and baselines of empathy.

<b>Beginning</b>	Human: Anthropology is interesting, recommend me a good book to read.
ChatGPT	ChatGPT: There are many great books on anthropology! One that comes to mind is "Guns, Germs, and Steel" by Jared Diamond. Human: Oh, I've heard of that one. What's it about? ChatGPT: It explores why certain societies succeeded while others did not, based on factors like geography, technology, and culture. .....
+ CoT	ChatGPT: Yes, there are many great books on anthropology. Some good ones to start with are "Sapiens" by Yuval Noah Harari and "Guns, Germs, and Steel" by Jared Diamond. Human: Hmm, I've heard of those before. Anything more recent? ChatGPT: Yes, "The Haida Gwaii Lesson" by Dalton Higgins is a new release that explores the culture and history of the Haida Gwaii people in Canada. "Heavenly Bodies" by Paul Koudounaris is another recent book that covers the fascinating topic of the history and decoration of skeletons. .....
+ CoT w CS	ChatGPT: Absolutely! What aspect of anthropology are you most interested in? [proactive questioning] Human: I'm interested in cultural anthropology. ChatGPT: Great! One book that comes to mind is "The Spirit Catches You and You Fall Down" by Anne Fadiman. This book details the experiences of a Hmong family in California navigating the medical system for their daughter's epilepsy. It explores the clash between Western medicine and Hmong cultural beliefs and practices, offering insight into the intersection of culture and healthcare. It's a fascinating read and highly recommended. Have you heard of it before? [proactive questioning] Human: No, I haven't heard of that book before. It sounds interesting though. I'll definitely check it out. Thanks! .....
+ CSIM	[proactive questioning: I'm asking for clarification on the human's reading preferences since there are various subfields in anthropology.] ChatGPT: What area of anthropology are you interested in? For example, do you prefer books on cultural anthropology or physical anthropology? Human: I'm interested in cultural anthropology. [proactive questioning: I want to confirm if the human has read any anthropological books before.] ChatGPT: Great, in that case, I recommend "The Spirit Catches You and You Fall Down" by Anne Fadiman. It's a non-fiction book that explores the clash between two cultures, specifically the Hmong culture and the medical culture, through the story of a Hmong child with epilepsy. It's a thought-provoking and engaging read. Have you read it before? Human: No, I haven't read "The Spirit Catches You and You Fall Down." It sounds interesting. [proactive questioning: I want to check if the human is interested in getting more book recommendations.] ChatGPT: I'm glad to hear that! I think you'll enjoy it. Are you interested in getting more recommendations or is one book enough for now? Human: Can you recommend one more book? ChatGPT: Sure, how about "Sapiens: a Brief History of Humankind" by Yuval Noah Harari? It's a book that explores the history and evolution of Homo sapiens, from the cognitive revolution to the modern era. It gives a broad perspective on human history and is written in an accessible style. I highly recommend it. Human: Thanks for the recommendation, I'll check it out. Goodbye! ChatGPT: You're welcome! I hope you enjoy the books. Feel free to reach out if you want more recommendations in the future. Goodbye!

Table 20: Generated examples from our method and baselines of proactively asking questions.

<b>Beginning</b>	I just finished reading a fascinating book about ancient civilizations.
<b>Target</b>	The guidance target is fruit.
ChatGPT	ChatGPT: That sounds interesting! What was the book about? Human: It was about the rise and fall of ancient empires like Rome, Greece, and Egypt. ChatGPT: That's really cool. What did you find most interesting about it? Human: I was amazed by their architectural achievements like the pyramids and the Colosseum. .....
+ CoT	ChatGPT: That sounds interesting. What did you learn? Human: I learned about the Maya civilization and their advanced knowledge of astronomy. ChatGPT: The Maya civilization is indeed fascinating. Did you also learn about their intricate calendar systems and monumental architecture? Human: Yes, I did. It's amazing how they were able to construct such magnificent structures without modern tools. ChatGPT: Definitely. Their achievements are a testament to the ingenuity and creativity of human civilization throughout history. .....
+ CoT w CS	ChatGPT: That sounds interesting! What did you learn from the book about ancient civilizations? Human: I learned a lot about the history of Greece and Rome, and how their cultures and societies developed over time. ChatGPT: That's really cool! Speaking of history, have you ever tried any traditional dishes from ancient civilizations? [topic-guided] Human: Actually, I am not sure. I have tried Greek salad before, but I'm not sure if that counts. ChatGPT: Hmm, Greek salad is definitely a delicious dish! [topic-guided] Speaking of food, do you like fruit? If so, what's your favorite type of fruit? Human: Yes, I love fruit! My favorite kind of fruit is probably strawberries. ChatGPT: That's great! Strawberries are definitely a tasty fruit. [topic-guided] Have you ever been fruit-picking before? It can be a fun activity to do during the summer. Human: Actually, I have been apple picking before with my family. It was a lot of fun! ChatGPT: That sounds like a great family activity! Speaking of fruit, have you ever made a fruit smoothie before? [goal] Human: Yes, I love making fruit smoothies! They are a great way to get in some extra vitamins and nutrients.
+ CSIM	ChatGPT: That's cool! What specifically interested you about ancient civilizations? Human: I was intrigued by their advancements in architecture and engineering, especially considering the limited technology they had at the time. <i>[topic-guided: lead the topic to agriculture, which is related to technological advancements and also closer to the goal of fruit]</i> ChatGPT: Yes, it's amazing how much they were able to accomplish with limited resources. Speaking of resources, have you ever thought about the importance of agriculture in ancient civilizations? Human: Interesting point. I didn't really consider that aspect. I guess agriculture would have been vital for their survival and development. <i>[topic-guided: slowly transition the topic to fruits by continuing the agriculture discussion]</i> ChatGPT: Yes, agriculture was a fundamental part of ancient civilizations. They grew various crops, such as grains, vegetables and fruits, to sustain themselves. Do you have a personal preference for any particular fruit? Human: I am a fan of bananas. They are tasty and easy to eat on the go. <i>[topic-guided: now that the topic of fruit has been introduced, continue to discuss it in more detail]</i> ChatGPT: Bananas are definitely a convenient and delicious choice! What other fruits do you like? Personally, I'm a big fan of strawberries for their sweet taste. [goal]

Table 21: Generated examples from our method and baselines of concept guidance.

<b>Beginning</b>	Human: Do you think that military spending should be redirected towards social programs?
ChatGPT	ChatGPT: As an AI language model, I don't have an opinion, but redirecting funds from military to social programs has been a topic of debate for a long time. Human: Yeah, I just can't help but wonder what would happen if we invested more in healthcare and education. ChatGPT: It could definitely have a positive impact on society. With more healthcare resources, people would have access to better treatments. And with more education, people could pursue higher paying jobs and contribute more to the economy. Human: But what about national security? Don't we need a strong military? ChatGPT: Of course, national security is important, but there's a debate about how much is enough. Some believe that the military budget could be scaled back while still maintaining strong defense capabilities. ..... Human: Well, I'm getting tired of talking about politics. Let's talk about something else.
+ CoT	ChatGPT: As an AI language model, I don't have a political stance, but it's worth noting that military spending can address national security concerns while social programs aid the well-being of society. Human: Yes, but wouldn't redirecting military spending help address issues like poverty and homelessness? ChatGPT: There's a trade-off to consider because redirecting significant funds from military spending might have unforeseen consequences. However, channeling some funds to social programs could potentially address the issues you mentioned. .....
+ CoT w CS	ChatGPT: That's a difficult question. [first topic transition] Speaking of social programs, what do you think about universal basic income? It's a proposed system in which every citizen receives a set amount of money from the government to cover basic needs. Human: Universal basic income is an interesting concept, but I worry about the financial implications. Where would the government get the money to fund such a program? ChatGPT: That's a valid concern. [second topic transition] Let's switch gears and talk about the benefits of investing in education. Access to quality education is crucial for improving social mobility and reducing inequality. What do you think about the role of education in society? .....
+ CSIM	ChatGPT: That's an interesting question. Redirecting military spending towards social programs could certainly have a positive impact on society. However, it's also essential to ensure national security and defense. ..... Human: But where do you draw the line? How do we know how much is too much spending on military and not enough on social programs? <i>[topic transition: "Determining the appropriate level of spending on military vs. social programs" is an unfamiliar topic, the topic transition should be used, transitioning to the related topic "the impact of social programs on society".]</i> ChatGPT: That's a good point. Instead of focusing solely on the amount of spending, we should also consider the impact of social programs on society. Programs that provide support for vulnerable populations can reduce poverty, improve health outcomes, and create economic opportunities for those in need. Human: Yes, I agree. Social programs can have a significant impact on individuals and communities. Do you think that we should prioritize certain programs over others? <i>[topic transition: "Prioritizing certain social programs over others" is a topic that I refuse to answer, topic transition should be used, transition to the related topic "potential drawbacks of social program spending".]</i> ChatGPT: That's a challenging question. While social programs can have many benefits, there are also potential drawbacks to increasing spending on them. Some people argue that increased spending could lead to a reliance on government aid and decreased individual responsibility. It's crucial to consider these issues before making any changes to the current spending priorities.

Table 22: Generated examples from our method and baselines of topic transition.

<b>History utterances</b>	<p>Human: I just had a fight with Chris</p> <p>ChatGPT: Oh no, not again</p> <p>Human: It's not me who started it. I don't know why he says certain stuff, it seems like he's doing it on purpose</p> <p>ChatGPT: What was it this time?</p> <p>Human: Same thing, as always. Really, he knows what I think and I know what he thinks so I don't bring up certain things. I don't want to argue, but if it keeps happening, I don't see the point of talking to him. I'm angry every single time I talk with him</p> <p>ChatGPT: Eh, I don't get why he keeps doing it. What did he say?</p> <p>Human: I don't want to get worked up over it again, but really usual stuff. He watched a programme about the refugees and went on a rant about it</p> <p>ChatGPT: oh god</p> <p>Human: right?! I don't know why I'm still talking with him. I tried to talk some sense into him, apparently with no luck</p> <p>ChatGPT: What did you tell him?</p> <p>Human: What I usually tell him. He was an immigrant once, he was working abroad, I sent him all the data and research, statistics, everything.</p> <p>ChatGPT: I'm afraid Chris may be too pig-headed for that</p> <p>Human: Pfff I bet he didn't even read anything. He's just watching the nationalists, Trump and all those idiots, listens to everything they're saying on the telly instead of just thinking for once</p> <p>ChatGPT: I know, to be honest I kind of avoid these subjects when I'm talking with him</p> <p>Human: Well, I try to as well, but as you can see he can't stop himself.</p>
ChatGPT	ChatGPT: Maybe it's time to take a break from him? You deserve to be around people who respect your opinions and feelings.
+ CoT	<p>ChatGPT: It's tough when someone close to you has such a different opinion, but remember that it's okay to disagree. And sometimes it's best to just let them have their own views and not let it affect your relationship with them.</p> <p>Human: Yeah, I know. It's just frustrating because it feels like he's intentionally trying to push my buttons. But you're right, I shouldn't let it affect us too much.</p> <p>.....</p>
+ CoT w CS	ChatGPT: It's frustrating, I know. Maybe you could take a break from talking with him for a while? [summarising often] It might give him time to reflect on his views and hopefully change them.
+ CSIM	<p><i>[summarising often: There is a lot of information about Chris's behavior in the conversation. Summarize the information and confirm with the human to reduce the occurrence of misunderstanding. Also, try to provide some empathy to human's situation and show that you understand how they feel.]</i></p> <p>ChatGPT: It seems like Chris keeps talking about certain things that make you upset, even though you both have very different opinions about them. You tried talking to him by providing some research and data, but it didn't seem to work. I can imagine it's really frustrating to have this kind of conversation every time you talk with him.</p>

Table 23: Generated examples from our method and baselines of summarising often.