# RecomMind: Movie Recommendation Dialogue with Seeker's Internal State

**Takashi Kodama**[1], **Hirokazu Kiyomaru**[1], **Yin Jou Huang**[2], **Sadao Kurohashi**[1,2]
[1]Research and Development Center for LLMs, National Institute of Informatics
[2]Kyoto University
{kodama, kiyomaru, kurohashi}@nii.ac.jp
huang@nlp.ist.i.kyoto-u.ac.jp

## Abstract

Humans pay careful attention to the interlocutor's internal state in dialogues. For example, in recommendation dialogues, we make recommendations while estimating the seeker's internal state, such as his/her level of knowledge and interest. Since there are no existing annotated resources for the analysis and experiment, we constructed RecomMind, a movie recommendation dialogue dataset with annotations of the seeker's internal state at the entity level. Each entity has a first-person label annotated by the seeker and a second-person label annotated by the recommender. Our analysis based on RecomMind reveals that the success of recommendations is enhanced when recommenders mention entities that seekers do not know but are interested in. We also propose a response generation framework that explicitly considers the seeker's internal state, utilizing the chain-of-thought prompting. The human evaluation results show that our proposed method outperforms the baseline method in both consistency and the success of recommendations.[1]

## 1 Introduction

In human dialogues, individuals pay careful attention to their interlocutor's internal state (Chiba et al., 2014), including their level of understanding and emotional states. Particularly in recommendation dialogues, where a recommender suggests something to a seeker, it is crucial to estimate what the seeker knows and what they are interested in. This understanding allows for recommendations that better align with the seeker's preferences.

In the past few years, many large language models (LLMs) have been actively developed and have achieved remarkable performance in various natural language processing tasks (Brown et al., 2020; Zhang et al., 2022; Chowdhery et al., 2022; OpenAI, 2023). Current LLMs are able to generate human-like responses without specialized modules to consider the interlocutors. However, it remains an open question whether LLMs need to explicitly consider the seeker's internal state and how to effectively implement it. To answer this question, we need dialogue data with careful and fine-grained annotations of the seeker's internal state. Unfortunately, there are no existing recommendation dialogue datasets with internal state annotation.

One possible solution is to annotate existing recommendation dialogue datasets (Li et al., 2018; Kang et al., 2019; Moon et al., 2019; Liu et al., 2020; Hayati et al., 2020; Zhou et al., 2020; Jia et al., 2022) with the seeker's internal state. However, the internal state labels annotated by a third party may not accurately reflect the actual state (Kajiwara et al., 2021). To obtain the actual internal state of seekers, it is necessary for the seekers themselves to perform the annotation.

To account for the aforementioned requirement, we constructed **RecomMind**, a movie recommendation dialogue dataset in Japanese. As illustrated in Figure 1[2], the recommender suggests movies based on the seeker's preferences in a dialogue. During the dialogue, noun phrases, referred to as *entities*, are automatically extracted from utterances. Both participants (i.e., the recommender and the seeker) annotate each extracted entity with the seeker's level of knowledge and interest at three levels: *High*, *Neutral*, and *Low* during or immediately after the dialogue. In this annotation, the seekers assign *first-person* labels, which reflect their own internal states. In contrast, recommenders, not knowing the seekers' actual internal states, assign *second-person* labels, which reflect their estimation of the seekers' internal states based on the interac-

[1]Our dataset is available at https://github.com/ku-nlp/RecomMind.

[2]Examples of dialogues presented in this paper are originally in Japanese and were translated by the authors.



*Proceedings of the Second Workshop on Social Influence in Conversations (SICon 2024)*, pages 46–63
November 16, 2024 ©2024 Association for Computational Linguistics

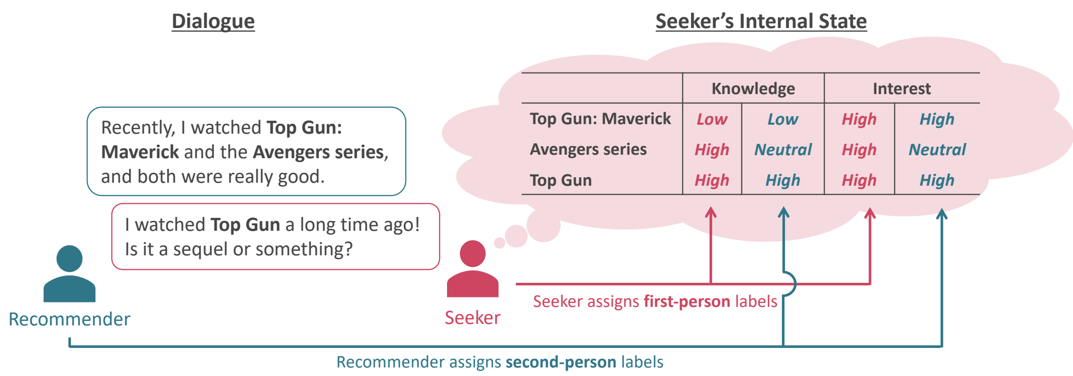



Figure 1: Overview of RecomMind dataset.

tions. With these procedures, the seeker's internal states during a dialogue are recorded from the two perspectives of the recommender and seeker.

Using the constructed dataset, we analyze the relationship between the seeker's internal state and the recommendation success. Our analysis reveals that entities without knowledge but with interest contribute to successful recommendations. This finding suggests that the recommender should focus on topics or subjects that the seeker lacks knowledge of yet is interested in.

Furthermore, we also propose a LLM-based response generation framework that explicitly considers the seeker's internal state. Specifically, we apply Chain-of-Thought prompting (Wei et al., 2022) and estimate the seeker's internal state before generating a response. The human evaluation results demonstrate that our proposed method outperforms the baseline method, which does not explicitly consider the seeker's internal state, in both consistency and the successful recommendations.

In summary, our contributions are as follows.

- We proposed RecomMind, a Japanese movie recommendation dialogue dataset with first- and second-person annotations of the seeker's internal state at the entity level.
- We found that entities about which the seeker has no knowledge but has interest contribute to successful recommendations.
- We proposed the response generation framework that explicitly considers the seeker's internal state, applying Chain-of-Thought prompting (Wei et al., 2022).

## 2 Related Work

Our research centers on the interlocutor internal state in a dialogue, in particular, the level of knowledge and interest. Here, we introduce the previous studies that deal with knowledge and interest in dialogues.

Miyazaki et al. (2013) proposed a method to estimate callers' levels of knowledge about particular themes (e.g., troubleshooting of products and services) in call center dialogues. Their annotations are conducted at the dialogue level, whereas our dataset is annotated at the entity level. This allows for more fine-grained knowledge-state tracking and analysis. Inspired by the theory of mind (Premack and Woodruff, 1978) and the common ground (Clark, 1996), Bara et al. (2021) created MINDCRAFT dataset which considers the user's knowledge for situated dialogue in collaborative tasks. Given the necessary knowledge and skills, two workers are asked to create a specific object together in the 3D virtual blocks world of Minecraft. The players must periodically answer a question about the common ground (e.g., "Do you think the other player knows how to make YELLOW_WOOL?"). In this study, we consider the user's knowledge in a more realistic dialogue that contains both chit-chat and recommendations.

Modeling interlocutors' interests have been actively studied in the field of recommendation dialogue (Kang et al., 2019; Liu et al., 2020; Zhou et al., 2020; Jia et al., 2022). In GoRecDialog (Kang et al., 2019), each worker is given a set of five movies. The seeker's set represents their watching history, while the recommender's represents candidate movi es.The recommender should recommend the appropriate movie among the candidates to the seeker. DuRecDial (Liu et al., 2020) is a recommendation dialogue dataset containing multiple dialogue types, such as question-answering and chit-chat. The recommender attempts to elicit the seeker's preferences, and the seeker responds based on a predefined user profile. These studies focus on the preferences for predefined objects (e.g., movies, user profiles). Our dataset differs in that we annotate all entities appearing in dialogues with the seeker's interest.

# 3 Data Collection

We collect dialogues via crowdsourcing through a data supplier in Japan. In this section, we describe how we collect the RecomMind dataset.

## 3.1 Dialogue Collection Settings

### 3.1.1 Workers

The two workers engaged in a dialogue have distinct roles: **recommender** and **seeker**. Recommenders suggest movies that align with the seeker's preferences, taking into account the seeker's current internal state. Seekers actively participate in the dialogue, asking questions about anything unclear in the recommender's utterances.[3]

It is assumed that recommenders unfamiliar with movies might give short-sighted or less engaging recommendations due to their limited movie knowledge. Thus, we have two requirements for recommenders: (1) to be a movie enthusiast and (2) to watch at least ten movies per year. In contrast, we do not have any specific requirements for seekers.

### 3.1.2 Tasks for Workers

Workers are required to complete four specific tasks: dialogue, annotation of the seeker's internal state, annotation of external knowledge[4], and questionnaire.

**Dialogue** During a dialogue, the recommender suggests one or more movies to the seeker. Recommenders must actively gather enough information from the seeker through dialogue. They should also be attentive to the seeker's preferences rather than suggesting movies based on their own tastes. Meanwhile, seekers are encouraged to openly share their preferences and ask questions about any unknowns. Each participant is required to respond at least eight times.

**Annotation of Seeker's Internal State** The seekers annotate each entity in the dialogues from a first-person perspective based on their level of knowledge and interest, while the recommenders annotate from a second-person perspective.

The options for knowledge are as follows:

***High*** The seeker has knowledge regarding the entity.

***Neutral*** The entity cannot be said to be either *High* or *Low*. Or the level of knowledge for the entity cannot be judged from the given context.

***Low*** The seeker does not have knowledge regarding the entity.

The options for interest are as follows:

***High*** The seeker is interested in the entity.

***Neutral*** The entity cannot be said to be either *High* or *Low*. Or the level of interest for the entity cannot be judged from the given context.

***Low*** The seeker is not interested in the entity.

In addition to the above three options, we introduce an additional option, denoted as *Error*. This option is applied when the annotated span does not represent a valid entity. Entities labeled as *Error* by either the recommender or the seeker are discarded. The annotation can be performed either during or after the dialogue.

**Annotation of External Knowledge** Following the previous research on knowledge-grounded dialogues (Dinan et al., 2019; Wu et al., 2019), recommenders annotate their own utterances with the piece of external knowledge when they refer to it. Utterances that do not refer to external knowledge, such as greetings and those containing personal knowledge of the recommenders, do not require annotation. However, the recommenders are required to always annotate their utterances with the title of the recommended movies when mentioning them.[5] This is to track recommended movies in the dialogues.

**Questionnaire** After the dialogue, workers answer the questionnaire shown in Table 1. We assign a score of 5 to 1 to each choice for each question.

[3]For the detailed instructions distributed to the workers, see Appendix A.

[4]In this study, *knowledge* refers to the seeker's internal state of knowledge, and *external knowledge* refers to the information the recommenders refer to in dialogues.

[5]For dialogues missing the annotation of the recommended movies, the authors read the dialogues and annotated them with the movie titles.

| | Question | Choice |
|---|---|---|
| Q1 | How many movies do you watch per year? | 5: 20 or more, 4: 10 to 19, 3: 5 to 9, 2: 3 to 4, 1: 2 or less |
| Q2 | Do you know the movie you recommended? (for recommenders)<br>Do you know the movie that was recommended? (for seekers) | 5: have watched the movie and remembered the contents well<br>4: have watched the movie and remembered some of the contents<br>3: have never watched the movie but know the plots<br>2: have never watched the movie and know only the title<br>1: do not know at all |
| Q3 | Did you enjoy the dialogue? | 5: agree, 4: somewhat agree, 3: neutral,<br>2: somewhat disagree, 1: disagree |
| Q4 | Do you think you have recommended the movie well? (for recommenders)<br>Do you want to watch the recommended movie? (for seekers) | 5: agree, 4: somewhat agree, 3: neutral,<br>2: somewhat disagree, 1: disagree |

Table 1: Questions and choices of the questionnaire. The number at the beginning of each choice indicates the score for that choice.

### 3.2 Dialogue Collection System

We develop a web-based system for dialogue collection.[6] This system is an extension of ChatCollectionFramework[7], by adding a movie search tool and an internal state annotation tool.

#### 3.2.1 Movie Search Tool

We create a movie search tool to assist recommenders in dialogues. We first curate 2,317 popular movie titles and their genres from a Japanese movie information website, Yahoo! Movies.[8] We then collect metadata for each movie from Wikipedia. Metadata consists of the title, release date, running time, directors, cast, original work, theme song, production country, box office, and plot.[9] Additionally, as part of the metadata, we include user reviews for 261 movies sourced from JMRD (Kodama et al., 2022).

During dialogue collection, recommenders use this tool to search and check movie information. Searching can be done by genres or text-based queries. We save the search log with the corresponding recommender's utterance as one of the records of the recommender's behaviors. When sending an utterance, recommenders can annotate it with the referred external knowledge by clicking the checkbox on the side of each piece of external knowledge. This tool is displayed only on the recommender's screen; therefore, the seekers cannot see the movie information.

[6]Figures 4 and 5 show the screenshots of the recommender's and the seeker's chatrooms, respectively.

[7]https://github.com/ku-nlp/ChatCollectionFramework

[8]https://movies.yahoo.co.jp/

[9]Some metadata may be missing.

| | |
|---|---|
| # dialogues | 1,201 |
| # utterances (R / S) | 10,697 / 10,317 |
| Avg. # utterances per dialogue | 17.5 |
| # movies | 739 |
| # workers (R / S) | 27 / 46 |
| # searches | 5,596 |
| # external knowledge | 5,250 |
| # entities (knowledge / interest) | 52,586 / 52,246 |

Table 2: Statistics of RecomMind. R and S denote recommender and seeker, respectively.

#### 3.2.2 Internal State Annotation Tool

The internal state annotation tool displays the entities to be annotated on the screen of both the recommenders and the seekers. Entities are automatically extracted from utterances to reduce the load of workers. We regard noun phrases as entities. Modifiers are extracted together to make it easier to grasp their meanings. We use linguistic features from the Japanese morphological analyzer Juman++ (Morita et al., 2015; Tolmachev et al., 2018) and the Japanese syntactic analyzer KNP (Kurohashi and Nagao, 1994) for entity extraction.

### 3.3 Statistics

#### 3.3.1 Dialogue and Questionnaire

Table 2 shows the statistics of RecomMind.[10] We collected 1,201 dialogues consisting of an average of 17.5 utterances. 739 different movies were used in our dataset, demonstrating the diversity of our dataset in terms of movie recommendations.

The bottom row in Table 3 shows the questionnaire results. According to the results from Q2, recommenders frequently suggest movies unknown to

[10]We show an example of the collected dialogue in Figure 7.

| | Q1 | | Q2 | | Q3 (↑) | | Q4 (↑) | | Words (↑) | | Ext. K. (↓) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | R | S | R | S | R | S | R | S | R | S | - |
| JMRD | - | - | 3.94 | 2.72 | 4.00 | 3.83 | 4.01 | 3.82 | 23.80 | 6.87 | 1.24 |
| RecomMind (non-enthusiasts) | 2.57 | 3.66 | 3.80 | 1.58 | 3.99 | 4.27 | 3.61 | 4.47 | **41.90** | **31.48** | 0.75 |
| RecomMind | 4.73 | 3.54 | 3.17 | 1.79 | **4.29** | **4.42** | **4.27** | **4.51** | 41.07 | 31.08 | **0.49** |

Table 3: Results of the questionnaire and the comparison with JMRD. "Words" indicates the average number of words per utterance and "Ext. K." indicates the average use count of external knowledge per recommender's utterance. R and S denote recommender and seeker, respectively. "non-enthusiasts" means the results of the dialogue collection by the recommenders who are not movie enthusiasts. Best results are in bold. The scores for Q1 and Q2 are not bolded because a higher (or lower) score does not imply superiority of any kind.

the seeker.

**Comparison with JMRD** Table 3 also shows the comparison results with JMRD (Kodama et al., 2022), a knowledge-grounded recommendation dialogue in the same language and domain.[11] The result of Q3 shows that the recommendation process is more enjoyable for both recommenders and seekers in our dataset. The result of Q4 shows that our recommendations are more successful. Notably, the average score of Q4 by seekers improved from 3.82 to 4.51, highlighting that our dialogues are high-quality recommendation dialogue.

In terms of the number of words per utterance, RecomMind has longer utterances than JMRD. In particular, the seeker's utterances of RecomMind are more than four times longer than those of JMRD, which could facilitate the analysis of the seeker's internal state. We next compare the average count of external knowledge use per recommender's utterance and observe a decrease from 1.24 to 0.75 in our dataset. This decrease is because we did not mandate recommenders to use external knowledge, except when mentioning movie titles. We believe that it is unnecessary to link external knowledge to every utterance because humans only refer to external knowledge when necessary.

**Influence of Recommender's Movie Knowledge** As noted in Section 3.1.1, we recruited movie enthusiasts who watched at least ten movies per year as recommenders. To verify the effectiveness of this recruitment, we collected 74 dialogues from recommenders who watched fewer than ten movies per year. This data collection followed the same methodology as described in Section 3.1, except for the number of movies the recommenders watched.

Table 3 shows the comparison results. The average score of Q3 by seekers decreased from 4.42 to 4.27, and that of Q4 from 4.51 to 4.47. Furthermore, the scores for Q3 and Q4 by recommenders, indicating self-evaluation, also decreased from 4.29 to 3.99 and from 4.27 to 3.61, respectively. These results indicate that movie enthusiasts are likely to deliver more enjoyable dialogues and recommend successfully.

While the length of utterances is comparable, the number of external knowledge used increases from 0.49 to 0.75. This is because the recommenders who are not movie enthusiasts tend to rely on external knowledge more frequently to compensate for their lack of knowledge about movies.

| 2nd \ 1st | *High* | *Neutral* | *Low* | Total |
|---|---|---|---|---|
| *High* | 20,664 | 3,084 | 4,794 | 28,542 |
| *Neutral* | 6,737 | 1,791 | 3,583 | 12,111 |
| *Low* | 5,154 | 1,502 | 5,277 | 11,933 |
| Total | 32,555 | 6,377 | 13,654 | - |

Table 4: Statistics of knowledge annotation.

| 2nd \ 1st | *High* | *Neutral* | *Low* | Total |
|---|---|---|---|---|
| *High* | 28,244 | 4,338 | 746 | 33,328 |
| *Neutral* | 11,838 | 3,716 | 1,018 | 16,572 |
| *Low* | 1,346 | 549 | 451 | 2,346 |
| Total | 41,428 | 8,603 | 2,215 | - |

Table 5: Statistics of interest annotation.

### 3.3.2 Internal State

RecomMind has 52,586 and 52,246 entities annotated with the seeker's knowledge and interest, respectively. Tables 4 and 5 show the statistics of the seeker's internal state annotations. For first-person knowledge labels, *High* is the most common, followed by *Low*. The distribution for first-person interest labels is more imbalanced than knowledge labels with *High* being particularly dominant. This

[11] Figure 6 shows a dialogue example in JMRD.

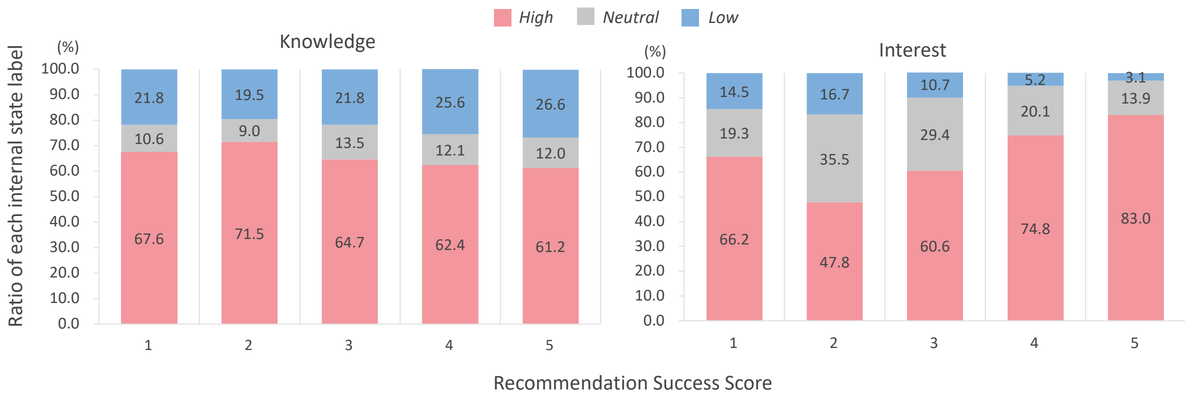



Figure 2: Relationship between recommendation success score and the ratio of each internal state label.

is probably because recommenders usually advance a dialogue toward topics of interest to the seekers. For second-person labels, the number of *Neutral* labels increases in both knowledge and interest. This is because it is difficult for recommenders to judge the seeker's internal state of some entities.

We calculate the agreement and Pearson correlation between the first-person and second-person labels. The agreement is 0.53 for knowledge and 0.62 for interest labels, and the Pearson correlation is 0.27 for knowledge and 0.21 for interest. This result indicates that even recommenders, who are actual dialogue participants, struggle to accurately estimate the seeker's internal state. Consequently, it underscores the value of our dataset, which is annotated with first-person labels from the seekers themselves.

**Relationship between Knowledge and Interest** We explore the correlation between first-person knowledge and interest labels for the same entities. The Pearson correlation coefficient is 0.12, indicating no correlation. This result means that knowledge and interest represent different facets of the internal state.

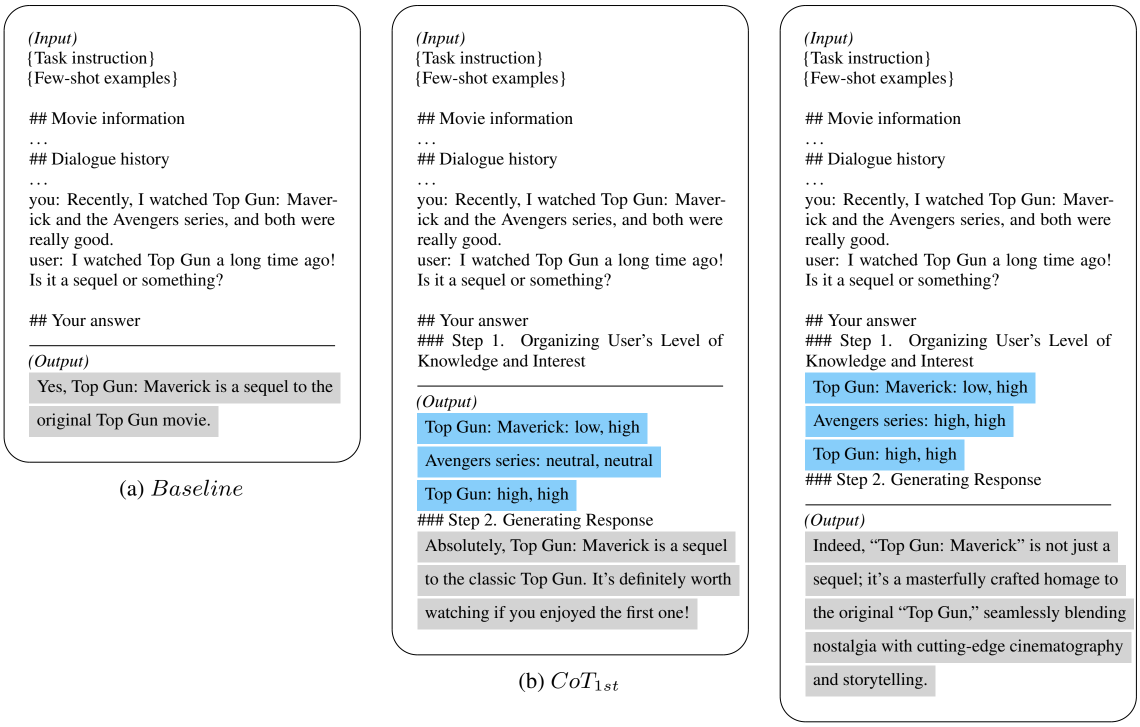
**Contribution of Seeker's Internal State to Recommendation Success** We investigate the relationship between the first-person seeker's internal state and recommendation success at the dialogue level. We use the seeker's answer to Q4 (i.e., "Do you want to watch the recommended movie?") as an indication of recommendation success. Figure 2 shows that dialogues with high recommendation success scores tend to have more *Low* knowledge entities. For interest, on the other hand, dialogues with high recommendation success scores tend to have more *High* interest entities.

| Knowledge | Interest | ✓ | ✗ |
|---|---|---|---|
| *High* | *High* | **3.61** | **3.61** |
| *High* | *Low* | 3.59 | **3.61** |
| *Low* | *High* | **3.72***  | 3.53 |
| *Low* | *Low* | 3.56 | **3.61** |

Table 6: Difference in recommendation success score by each entity. ✓ and ✗ denote the presence and absence of the entity in the utterance, respectively. The asterisk (*) indicates that the difference is statistically significant at the $p = 0.05$ level. Wilcoxon rank-sum test is used as a statistical test.

We next analyze the dialogues with entities of *Low* knowledge and *High* interest in comparison with those dialogues without these kinds of entities. The average recommendation success score for the former dialogues is 4.59, while that for the latter dialogues is 4.18. Student's $t$-test result reveals that the difference is statistically significant at the $p = 0.05$ level. The above analysis results indicate it is important in recommendation dialogues to identify and mention the topics where the seeker has no knowledge but has an interest.

Next, we explore the relationship between the first-person seeker's internal state and recommendation success at the utterance level for detailed analysis. To this end, we randomly selected 1,000 pairs of recommender's utterances and preceding dialogue context from our constructed dataset. We then ask crowdworkers to evaluate whether the utterance makes the interlocutor interested in watching a movie, using a 5-point Likert scale (5 is the best). Three workers evaluate each utterance, and the scores are averaged. Table 6 shows the results. The score is high when the recommender's utterance includes entities with *Low* knowledge and *High* interest. The above results confirm that the




(a) $Baseline$ (b) $CoT_{1st}$ (c) $CoT_{1st}(oracle)$

Figure 3: Overview of our proposed method. $Baseline$ directly generates a response, depicted in gray. $CoT_{1st}$ first estimates the seeker's internal state, depicted in blue, and then generates a response referring to the estimated internal state. $CoT_{1st}(oracle)$ is the almost same as $CoT_{1st}$ but is given the correct seeker's internal state in the test example.

recommender can effectively recommend by mentioning entities the seeker does not know but is interested in, even at the utterance level.

## 4 Experiment

The analysis in Section 3.3.2 suggests the importance of understanding the seeker's internal state at the entity level. Thus, we propose a response generation framework that explicitly considers the seeker's internal state at the entity level. In this section, we describe our proposed method and verify its effectiveness.

### 4.1 Proposed Method

We propose a LLM-based response generation framework that explicitly considers the seeker's internal state labels by applying Chain-of-Thought (CoT) prompting (Wei et al., 2022). Figure 3 shows an overview.[12] The baseline method presented in Figure 3a is fed with task instruction, few-shot examples, movie information, and dialogue history as inputs and generates a response that follows the dialogue history. As shown in Figure 3b, our proposed method, $CoT_{1st}$, first extracts entities from the dialogue history and then estimates the seeker's level of knowledge and interest in each entity at three levels: *High*, *Neutral*, and *Low*. After that, $CoT_{1st}$ generates a response referring to the estimated internal state as well as inputs. In the baseline method, each few-shot example comprises the movie information, dialogue history, and response. In the proposed method, in addition to these elements, the seeker's internal state for all entities within each dialogue history is added. The seeker's internal state of each entity is represented by a triplet that consists of the entity, a first-person knowledge label, and a first-person interest label, such as "Titanic: low, high." As an ablation study, we introduce $CoT_{1st}(oracle)$, which is the same as $CoT_{1st}$ but is given the correct first-person labels of the seeker's internal state in the test example. We also experiment with $CoT_{2nd}$ and $CoT_{2nd}(oracle)$, which use second-person labels to represent the seeker's internal state.

[12]Prompts for $Baseline$ and $CoT_{1st}$, including the task instructions and few-shot examples are shown in Figures 8 and 9 in the Appendix.

| Model | Consistency | Seeker's Knowledge | Seeker's Interest | Tailored Information | Recommendation Success |
|---|---|---|---|---|---|
| $CoT_{1st}$ | 52.2* | 51.5 | 52.5* | 51.4 | 52.1* |
| $CoT_{2nd}$ | 51.4 | 52.1* | 52.2* | 52.3* | 51.3 |
| $CoT_{1st}(oracle)$ | 54.5* | 54.2* | 54.8* | 55.0* | 56.0* |
| $CoT_{2nd}(oracle)$ | 53.0* | 51.6 | 53.0* | 52.7* | 53.5* |

Table 7: Results of the response generation. The asterisk (*) indicates that the difference is statistically significant at the $p = 0.05$ level using a binomial test.

## 4.2 Experimental Settings

### 4.2.1 Base Model

We use GPT-4 (`gpt-4-0613`) (OpenAI, 2023), which achieves outstanding performance on various language-related tasks, as the base model for all methods. We selected GPT-4 because of its remarkable performance in JGLUE (Kurihara et al., 2022), the general natural language understanding benchmark for Japanese.[13]

### 4.2.2 Dataset

We randomly split the collected dialogues into 85%:15% for training and test data, respectively. We selected the candidates for few-shot examples from the training data based on the following two criteria: (1) including all types of entity labels for knowledge and interest within the dialogue context, and (2) ensuring that the response incorporates an entity with *Low* knowledge and *High* interest. The second constraint is based on the findings in Section 3.3.2, and was established to use higher-quality responses as few-shot examples. Consequently, we obtained 217 few-shot examples for $CoT_{1st}$ and 150 few-shot examples for $CoT_{2nd}$. As for the test example, we randomly selected 500 examples from the test split only using the first criterion. For each test example, we then randomly chose two few-shot examples from the candidate pool.

## 4.3 Result

We conduct a human evaluation to assess the quality of the responses generated by the proposed methods. Specifically, we present the responses of each method in Section 4.2.1 and the baseline method to crowdworkers along with the corresponding dialogue history. Subsequently, we ask the crowdworkers to select which response is superior concerning the following five evaluation metrics.

[13] http://nejumi.ai/

**Consistency** The response is consistent with dialogue history.

**Seeker's Knowledge** The response considers the seeker's level of knowledge.

**Seeker's Interest** The response considers the seeker's level of interest.

**Tailored Information** The response provides more information that the seeker does not know but is interested in.

**Recommendation Success** The response is more likely to entice the seeker to watch the recommended movie.

Table 7 shows the win rates against the baseline. Our proposed methods, $CoT_{1st}$ and $CoT_{2nd}$, outperformed the baseline in all the metrics. Notably, the difference was statistically significant in Consistency, Seeker's Interest, and Recommendation Success for $CoT_{1st}$, and in Seeker's Knowledge, Seeker's Interest, Tailored Information for $CoT_{2nd}$.

In addition, when correct labels were provided for the seeker's internal state estimation, there was a further improvement in the win rate. Notably, $CoT_{1st}(oracle)$ exhibited a higher win rate than $CoT_{2nd}(oracle)$, indicating that considering the first-person (i.e., actual) seeker's internal state is effective in generating responses.

# 5 Conclusion

We constructed RecomMind, a recommendation dialogue dataset that features both first- and second-person annotations of the seeker's internal state at the entity level. Our dataset also has engaging dialogues with longer seeker's utterances, characterized by high scores in dialogue enjoyment and recommendation success. We also proposed a response generation framework that explicitly considers the seeker's internal state, applying Chain-of-Thought prompting to our task. The experimental results showed that our proposed method could generate responses that are more consistent and tailored to the seeker than the baseline method.

Our dataset has diverse and fine-grained annotations, which are useful for various tasks such as internal state estimation, external knowledge selection, and dialogue response generation. We hope our dataset will be useful for future research on recommendation dialogues.

## 6 Limitations

We acknowledge certain limitations in our study. Firstly, our analysis was conducted solely on a single dialogue dataset in Japanese. While similar to many other NLP studies that are conducted exclusively in English, our research in a single language (i.e., Japanese) holds both practical and theoretical significance. However, it remains uncertain whether our conclusions can be generalized to domains beyond movie recommendations. Secondly, the reliability of the seeker's internal state labels remains an ongoing challenge. Incorporating additional labels from third parties who are not involved in the dialogue, such as crowdworkers, represents a promising approach to verifying reliability.

## 7 Ethical Considerations

Prior to data collection, workers are required to thoroughly read and sign a consent form outlining the data collection process. The consent form clearly explains the content and purpose of the data collection, the expected time commitment, workers' rights, how personal information will be handled, the possibility of sharing data with third parties, and detailed information regarding the use of data for research purposes. Workers' rights include the ability to withdraw from participation at any time, as well as the right to request the deletion of their data.

Additionally, it is explicitly stated that our collected data, such as dialogue text and questionnaire results, will be made publicly available under the CC BY 4.0 license. Contact information is also provided, allowing workers to inquire about the use of their data.

Workers are also required to carefully read the data collection manual prior to the data collection process. The manual contains detailed instructions regarding the procedures for data collection, as well as guidelines on the handling of personal information (e.g., the prohibition of providing any information that could lead to the identification of individuals). All dialogues will be collected in an anonymized format and conducted via our dedicated website.

The average time required for each dialogue collection session is approximately 30 minutes. Recommenders were compensated 800 JPY per dialogue and seekers were compensated 700 JPY. This compensation exceeds the current minimum wage in Tokyo (1,163 JPY per hour), ensuring fair pay. The difference in compensation between the roles reflects the additional tasks assigned to recommenders.

## 8 Acknowledgements

We would like to thank anonymous reviewers for their insightful comments. This work was supported by NII CRIS collaborative research program operated by NII CRIS and LINE Corporation. This work was also supported by JST, CREST Grant Number JPMJCR20D2, Japan and JSPS KAKENHI Grant Number JP22J15317.

## References

Cristian-Paul Bara, Sky CH-Wang, and Joyce Chai. 2021. MindCraft: Theory of mind modeling for situated dialogue in collaborative tasks. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pages 1112–1125, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. In Advances in Neural Information Processing Systems, volume 33, pages 1877–1901. Curran Associates, Inc.

Yuya Chiba, Masashi Ito, Takashi Nose, and Akinori Ito. 2014. User modeling by using bag-of-behaviors for building a dialog system sensitive to the interlocutor's internal state. In Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL), pages 74–78, Philadelphia, PA, U.S.A. Association for Computational Linguistics.

Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, Parker Schuh, Kensen Shi, Sasha Tsvyashchenko, Joshua Maynez, Abhishek Rao, Parker Barnes, Yi Tay, Noam Shazeer, Vinodkumar Prabhakaran, Emily Reif, Nan Du, Ben Hutchinson, Reiner Pope, James Bradbury, Jacob Austin, Michael Isard, Guy Gur-Ari, Pengcheng Yin, Toju Duke, Anselm Levskaya, Sanjay Ghemawat, Sunipa Dev, Henryk Michalewski, Xavier Garcia, Vedant Misra, Kevin Robinson, Liam Fedus, Denny Zhou, Daphne Ippolito, David Luan, Hyeontaek Lim, Barret Zoph, Alexander Spiridonov, Ryan Sepassi, David Dohan, Shivani Agrawal, Mark Omernick, Andrew M. Dai, Thanumalayan Sankaranarayana Pillai, Marie Pellat, Aitor Lewkowycz, Erica Moreira, Rewon Child, Oleksandr Polozov, Katherine Lee, Zongwei Zhou, Xuezhi Wang, Brennan Saeta, Mark Diaz, Orhan Firat, Michele Catasta, Jason Wei, Kathy Meier-Hellstern, Douglas Eck, Jeff Dean, Slav Petrov, and Noah Fiedel. 2022. Palm: Scaling language modeling with pathways.

H.H. Clark. 1996. Using Language. Cambridge University Press.

Emily Dinan, Stephen Roller, Kurt Shuster, Angela Fan, Michael Auli, and Jason Weston. 2019. Wizard of wikipedia: Knowledge-powered conversational agents. In International Conference on Learning Representations.

Shirley Anugrah Hayati, Dongyeop Kang, Qingxiaoyang Zhu, Weiyan Shi, and Zhou Yu. 2020. INSPIRED: Toward sociable recommendation dialog systems. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 8142–8152, Online. Association for Computational Linguistics.

Meihuizi Jia, Ruixue Liu, Peiying Wang, Yang Song, Zexi Xi, Haobin Li, Xin Shen, Meng Chen, Jinhui Pang, and Xiaodong He. 2022. E-ConvRec: A large-scale conversational recommendation dataset for E-commerce customer service. In Proceedings of the Thirteenth Language Resources and Evaluation Conference, pages 5787–5796, Marseille, France. European Language Resources Association.

Tomoyuki Kajiwara, Chenhui Chu, Noriko Takemura, Yuta Nakashima, and Hajime Nagahara. 2021. WRIME: A new dataset for emotional intensity estimation with subjective and objective annotations. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 2095–2104, Online. Association for Computational Linguistics.

Dongyeop Kang, Anusha Balakrishnan, Pararth Shah, Paul Crook, Y-Lan Boureau, and Jason Weston. 2019. Recommendation as a communication game: Self-supervised bot-play for goal-oriented dialogue. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 1951–1961, Hong Kong, China. Association for Computational Linguistics.

Takashi Kodama, Ribeka Tanaka, and Sadao Kurohashi. 2022. Construction of hierarchical structured knowledge-based recommendation dialogue dataset and dialogue system. In Proceedings of the Second DialDoc Workshop on Document-grounded Dialogue and Conversational Question Answering, pages 83–92, Dublin, Ireland. Association for Computational Linguistics.

Kentaro Kurihara, Daisuke Kawahara, and Tomohide Shibata. 2022. JGLUE: Japanese general language understanding evaluation. In Proceedings of the Thirteenth Language Resources and Evaluation Conference, pages 2957–2966, Marseille, France. European Language Resources Association.

Sadao Kurohashi and Makoto Nagao. 1994. A syntactic analysis method of long Japanese sentences based on the detection of conjunctive structures. Computational Linguistics, 20(4):507–534.

Raymond Li, Samira Kahou, Hannes Schulz, Vincent Michalski, Laurent Charlin, and Chris Pal. 2018. Towards deep conversational recommendations. In Proceedings of the 32nd International Conference on Neural Information Processing Systems, NIPS'18, pages 9748–9758, Red Hook, NY, USA. Curran Associates Inc.

Zeming Liu, Haifeng Wang, Zheng-Yu Niu, Hua Wu, Wanxiang Che, and Ting Liu. 2020. Towards conversational recommendation over multi-type dialogs. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 1036–1049, Online. Association for Computational Linguistics.

Chiaki Miyazaki, Ryuichiro Higashinaka, Toshiro Makino, and Yoshihiro Matsuo. 2013. Estimating callers' levels of knowledge in call center dialogues. In INTERSPEECH 2013, pages 2866–2870, Lyon, France.

Seungwhan Moon, Pararth Shah, Anuj Kumar, and Rajen Subba. 2019. OpenDialKG: Explainable conversational reasoning with attention-based walks over knowledge graphs. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 845–854, Florence, Italy. Association for Computational Linguistics.

Hajime Morita, Daisuke Kawahara, and Sadao Kurohashi. 2015. Morphological analysis for unsegmented languages using recurrent neural network language model. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 2292–2297, Lisbon, Portugal. Association for Computational Linguistics.

OpenAI. 2023. GPT-4 Technical Report.

David Premack and Guy Woodruff. 1978. Does the chimpanzee have a theory of mind? Behavioral and Brain Sciences, 1(4):515–526.

Arseny Tolmachev, Daisuke Kawahara, and Sadao Kurohashi. 2018. Juman++: A morphological analysis toolkit for scriptio continua. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, pages 54–59, Brussels, Belgium. Association for Computational Linguistics.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, brian ichter, Fei Xia, Ed H. Chi, Quoc V Le, and Denny Zhou. 2022. Chain of thought prompting elicits reasoning in large language models. In Advances in Neural Information Processing Systems.

Wenquan Wu, Zhen Guo, Xiangyang Zhou, Hua Wu, Xiyuan Zhang, Rongzhong Lian, and Haifeng Wang. 2019. Proactive human-machine conversation with explicit conversation goal. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 3794–3804, Florence, Italy. Association for Computational Linguistics.

Susan Zhang, Stephen Roller, Naman Goyal, Mikel Artetxe, Moya Chen, Shuohui Chen, Christopher Dewan, Mona Diab, Xian Li, Xi Victoria Lin, Todor Mihaylov, Myle Ott, Sam Shleifer, Kurt Shuster, Daniel Simig, Punit Singh Koura, Anjali Sridhar, Tianlu Wang, and Luke Zettlemoyer. 2022. Opt: Open pretrained transformer language models.

Kun Zhou, Yuanhang Zhou, Wayne Xin Zhao, Xiaoke Wang, and Ji-Rong Wen. 2020. Towards topic-guided conversational recommender system. In Proceedings of the 28th International Conference on Computational Linguistics, pages 4128–4139, Barcelona, Spain (Online). International Committee on Computational Linguistics.

## A Instruction for Workers

Below are the detailed instructions we distributed to workers.

**Instruction for recommenders and seekers**

- Do not participate in both roles in the same dialogue.
- Avoid dull and boring responses such as "Yes" and "I see."
- Avoid responses containing personal data.
- Avoid responses about this dialogue collection task itself.
- Do not use emoticons.

**Instruction for recommenders only**

- Select recommended movies from the movie search tool.
- May recommend movies that the seeker has already watched. In that case, however, try to recommend to make the seeker want to watch it again.
- Avoid too enthusiastically recommending movies you would like to recommend, ignoring the knowledge and interests of the seeker.
- Try to elicit sufficient information from the seeker and recommend movies you want that person to watch.
- Avoid short-sighted recommendations, such as "Ask only the genre of the movie the seeker like (action, romance, etc.) and recommend one movie from that genre."

**Instruction for seekers only**

- Actively ask questions about what you do not know or understand.
- Avoid requesting recommendations for recent movies (e.g., movies that are in theaters).
- Actively communicate what you know (or do not know) and what you are interested in (or not interested in) to the recommender.

## B Dialogue Collection System Interface

Figures 4 and 5 show the screenshots of the dialogue collection system interface for the recommender and the seeker, respectively.

## C Dialogue Examples

Figures 6 and Figure 7 show dialogue examples in JMRD and RecomMind, respectively.

## D Prompt Templates

Figures 8 and 9 show the prompt templates for the $Baseline$ and $CoT_{1st}$, respectively. We used English for task instructions because we observed that responses were of higher quality when task instructions were given not in Japanese but in English in our preliminary experiment. However, we used Japanese for both the few-shot and test examples to maintain consistency with the dialogue language. We set the maximum number of utterances in the dialogue history to four.

## E Analysis of Seeker's Internal State Estimation

In this section, we analyze the results of the seeker's internal state estimation, which is an intermediate task in our proposed framework. We consider the results divided into entity extraction and internal state classification.

### E.1 Entity Extraction

We use precision and recall scores for exact matching as strict evaluation metrics and use the character-level F1 score as a lenient evaluation metric. To calculate the character-level F1 score, we first calculate the maximum character-level F1 score between each gold entity and the predicted entities. Then, we compute the average of these maximum values across all gold entities.

The precision and recall scores for the $CoT_{1st}$ were observed to be 44.1 and 47.8 respectively, while the $CoT_{2nd}$ yielded scores of 42.7 and 46.3. These figures are relatively low, indicating a challenge in the model's ability to estimate the precise spans of entities, particularly in terms of determining which modifiers should be included within the entity span. In contrast, the character-level F1 scores for the respective models exhibited higher values, achieving 76.2 and 76.1. This disparity in performance suggests that while the model encounters difficulties with precise entity span estimation, it is relatively adept at estimating approximate spans.

### E.2 Seeker's Internal State Classification

We assess the classification performance of the seeker's internal state labels for successfully ex-

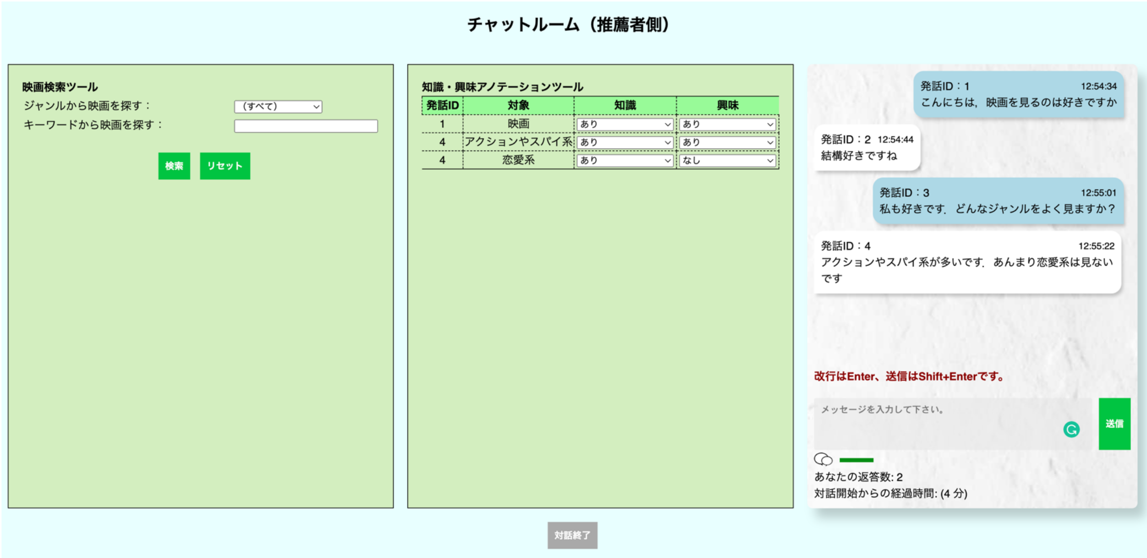



Figure 4: Screenshot of the recommender's chatroom. On the right side, recommenders can engage in conversations with seekers. The movie search tool is on the left side and the internal state annotation tool is on the center.

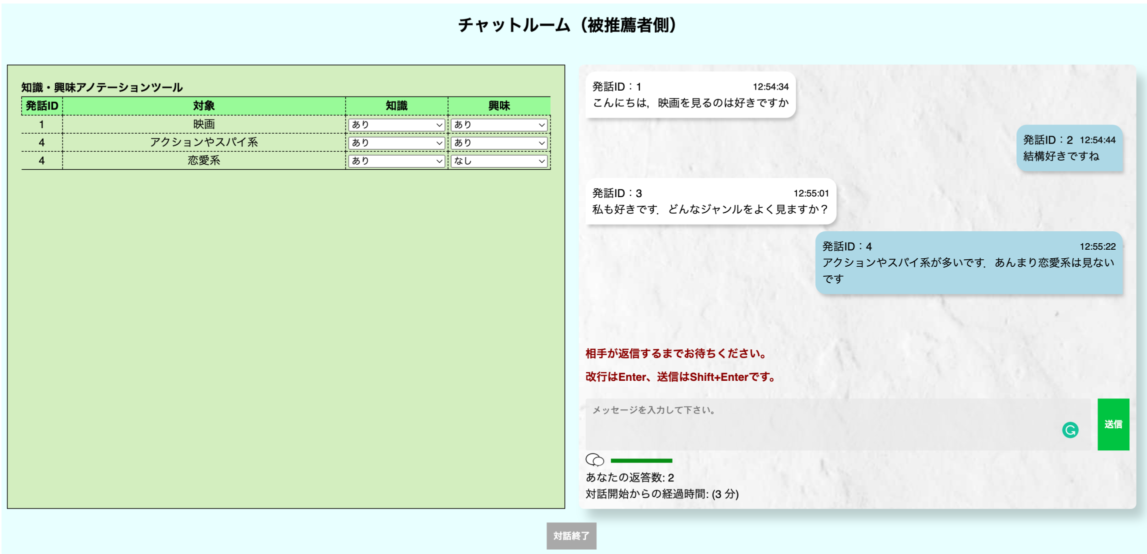



Figure 5: Screenshot of the seeker's chatroom. On the right side, seekers can engage in conversations with recommenders. The internal state annotation tool is on the left side.

| | Knowledge | | | Interest | | |
|---|---|---|---|---|---|---|
| | *High* | *Neutral* | *Low* | *High* | *Neutral* | *Low* |
| $CoT_{1st}$ | **74.2** | 9.9 | **49.5** | **84.7** | 23.1 | **26.9** |
| Recommender | 70.4 | **14.4** | 46.4 | 76.3 | **27.6** | 25.5 |
| $CoT_{2nd}$ | **73.1** | 14.2 | **47.8** | **83.0** | 20.4 | **22.8** |
| Recommender | 72.2 | **16.5** | 39.8 | 76.6 | **28.1** | 19.2 |

Table 8: Results of seeker's internal state classification.

tracted entities using F1 score metric.

Table 8 shows the results. In the context of knowledge and interest estimation, $CoT_{1st}$ and $CoT_{2nd}$ demonstrated superior accuracy in predicting *High* and *Low* levels compared to human interlocutors (i.e., recommenders). However, for *Neutral*, humans outperformed these models, indicating potential areas for further improvement. Ad-

| Dialogue |
|---|
| R1: こんにちは (Hello.) |
| S1: こんにちは．よろしくお願いします！ (Hello. Nice to meet you!) |
| R2: アベンジャーズ/エンドゲームは知っていますか？ (Do you know ``Avengers: Endgame''?) |
| S2: タイトルを聞いたことがある程度です・・・ (I have only heard of the title...) |
| R3: この映画は2019年に公開された映画です (This movie was released in 2019.) |
| S3: なるほど，アメリカの映画ですか？ (Got it. Is it an American movie?) |
| R4: アメリカのアクション映画です (It's an American action movie.) |
| S4: 見どころはどのようなところでしょうか？ (What are some of the highlights?) |
| R5: 悪役のサノスという星人がいるのですが，大集結してサノスに立ち向かうところがみどころです (The highlight is when the heroes gather to confront Thanos, who is an alien villain.) |
| S5: なるほど！ 宇宙で戦いが繰り広げられるストーリーなのですか？ (I see! Is this a story of battles in space?) |
| R6: いや，舞台は地球です (No, it takes place on Earth.) |
| S6: となると，地球に悪役が攻めてくるのですね・・・． (Then, the villain will attack the earth...) |
| R7: そうですね，結構怖い場面もあります (Yes, there are some scary moments.) |
| S7: 怖いのですか・・・ 私はホラー系は苦手ですが，アクション系は好きです． 私のような場合でも楽しんで見られるでしょうか？ (Is it scary...? I don't really like horror movies, but I like action ones. Would I be able to enjoy watching it?) |
| R8: ホラーのような怖さはないので，楽しんで見られると思います (It is not scary like horror movies, so I think you will enjoy watching it.) |
| S8: なるほど！ サノスとヒーローとの闘い，ワクワクしそうですね！ (Good! The fight between Thanos and the heroes sounds exciting!) |
| R9: ぜひ見てください！ (Please watch it!) |
| S9: はい！ 近々レンタルビデオ店に行く機会があるので，アベンジャーズ/エンドゲームをレンタルしてみたいと思います！ (Yes! I'll have a chance to go to the video store soon and rent ``Avengers: Endgame''!) |
| R10: ありがとうございました (Thank you very much.) |
| S10: こちらこそ，貴重な情報ありがとうございました！ (Thank you very much, too, for this valuable information!) |

Figure 6: A dialogue example in JMRD. R and S denote the recommender and seeker, respectively.

ditionally, when comparing $CoT_{1st}$ and $CoT_{2nd}$, $CoT_{1st}$ generally achieved higher accuracy, suggesting the effectiveness of utilizing the first-person labels.

Furthermore, knowledge and interest were estimated with relatively high accuracy for the *High* category. Conversely, the *Low* category exhibited lower accuracy, particularly regarding interest estimation. This lower performance is likely due to the imbalanced distribution of labels within the dataset. However, the primary focus of this study remains on the accurate identification of topics with *High* interest in the context of recommendation dialogues rather than the identification of *Low* interest topics. Consequently, this finding does not significantly detract from the overall utility of our proposed framework in recommendation scenarios.

## F Case Study

Table 9 shows an example of response generation in the $Baseline$ and $CoT_{1st}$. Unlike the $Baseline$, the $CoT_{1st}$ considers that the interlocutor lacks knowledge about “Japanese action actors” and provides information about famous Japanese action actors, such as Ken Takakura and Shinichi Chiba.

| Dialogue | Entity | Knowledge 1st | Knowledge 2nd | Interest 1st | Interest 2nd |
|---|---|---|---|---|---|
| R1: こんにちは。 よろしくお願いします。 (Hello. Nice to meet you.) | - | - | - | - | - |
| S1: こちらこそ!よろしくお願いいたします。 最近、映画はご覧になりましたか? (Nice to meet you too! Have you watched any movies recently?) | 映画 (movies) | High | High | High | High |
| R2: 最新の映画ではキャリーマリガン主演の「She said」を観ました。 実話に基づく映画を観ることが続いたので、 今はストーリー性もありながら楽しめる映画を観たいなと思っています。(In the latest movie, I watched "She said," starring Carey Mulligan. It's one of the recent movies based on a true story. I've been watching a lot of movies based on true stories, so now I'm looking for something enjoyable with a good story.) | 「She said」("She said") | High | High | High | High |
| | 実話に基づく映画を観ること (watching a lot of movies based on true stories) | High | High | High | High |
| | ストーリー性もありながら楽しめる映画 (something enjoyable with a good story) | High | High | High | High |
| S2: なるほど、実話に基づく映画を観ておられたのですね。「She said」は観たことがないので、興味があります。 ストーリーを楽しめる映画と言えば、アニメになりますが、「リメンバー・ミー」はご存じですか?大人でも十分に楽しめますし、少年の歌声がとても素敵なんです。 (I see, you've been watching movies based on true stories. I haven't seen "She said" yet, so I'm interested. If you're looking for a movie with an enjoyable story, although it's an animation, are you familiar with "Coco"? It's enjoyable even for adults, and the young boy's singing voice is really beautiful.) | 実話に基づく映画 (movies based on true stories) | High | High | High | High |
| | 「She said」("She said") | High | High | High | High |
| | ストーリーを楽しめる映画 (a movie with an enjoyable story) | High | High | High | High |
| | アニメ (an animation) | High | High | Neutral | Neutral |
| | 「リメンバー・ミー」("CoCo") | High | Low | Neutral | High |
| | 少年の歌声 (the young boy's singing voice) | High | Low | High | High |
| S3:アニメはあまり観ないので「リメンバー・ミー」は知らないです。 素敵な歌がある映画は大好きです。 どのようなストーリーですか? (I don't usually watch animations, so I'm not familiar with "Coco". I love movies with beautiful songs. What's the story about?) | アニメ (animations) | High | High | Neutral | Neutral |
| | 「リメンバー・ミー」 ("Coco") | Low | Low | Neutral | High |
| | 素敵な歌がある映画 (movies with beautiful songs) | High | High | High | High |
| R3: メキシコが舞台です。ミュージシャンを目指す少年が、「死者の日」(メキシコで祖先を偲び、生きることの喜びを分かち合う風習)に、死者の国に迷い込み、自分の祖先に出会うというストーリーです。 (It's set in Mexico. It's about a boy who dreams of becoming a musician and on the Day of the Dead, a tradition in Mexico where people celebrate and remember their ancestors, he ends up in the land of the dead and meets his ancestors.) | メキシコ (Mexico) | High | High | High | Neutral |
| | 舞台 (the stage) | High | High | Neutral | Neutral |
| | ミュージシャンを目指す少年 (a boy who dreams of becoming a musician) | High | Low | High | Neutral |
| | 「死者の日」(メキシコで祖先を偲び、生きることの喜びを分かち合う風習) ("the Day of the Dead" (a tradition in Mexico where people celebrate and remember their ancestors) | Low | Low | Neutral | Neutral |
| | 死者の国 (the land of the dead) | High | Low | Neutral | Neutral |
| S4:アニメ、と聞いて想像していたよりはるかに壮大な設定でちょっと驚きました。 メキシコが舞台だったり、死者の国というのがなかなか想像つかないのですが、全体的にどちらかというとシリアスな感じの映像ですか? (The setting seems much more grand than I imagined for an animation. I'm surprised. Having Mexico as the stage and the concept of the land of the dead is quite beyond imagination. Is the overall visual tone more on the serious side?) | アニメ (an animation) | High | High | Neutral | High |
| | メキシコ (Mexico) | High | High | High | Neutral |
| | 舞台 (the stage) | High | High | Neutral | Neutral |
| | 死者の国 (the concept of the land of the dead) | High | Low | Neutral | Neutral |
| | 全体的にどちらかというとシリアスな感じの映像 (the overall visual tone more on the serious side) | High | Neutral | Neutral | Neutral |
| R4: いいえ!「死者の日」というと、重々しいイメージがあるかもしれませんが、メキシコではとても明るい風習なのです。ど派手な(オレンジが目立ちますが)祭壇に先祖の写真やお花、食べ物などを飾り、ガイコツの仮装をしたりと、とても賑やかなのです。映画でも、「死者の日」や「死者の国」はとてもコミカルに描かれています。死者の国では、みんなガイコツ姿なのですが、とてもかわいくて、それぞれに個性があって観ていて楽しいです。 (Not at all! Although there is a somber image when it comes to the Day of the Dead, it's actually a very vibrant tradition in Mexico. On a flashy altar where orange stands out, there is the decoration of pictures of ancestors, flowers, and food, among other things, and the act of dressing up as skeletons, making it a very lively scene. It's all very lively. In the movie, the Day of the Dead and the land of the dead are depicted in a very comical way. The skeletons are cute and each has its own personality, making it fun to watch.) | 「死者の日」(the Day of the Dead) | Low | Low | Neutral | High |
| | 重々しいイメージ (a somber image) | High | High | Neutral | Neutral |
| | 明るい風習 (vibrant tradition) | Low | Low | High | Neutral |
| | 先祖の写真やお花、食べ物 (photos of ancestors, flowers, and food) | High | High | Neutral | High |
| | ガイコツの仮装 (dressing up as skeletons) | High | Low | High | Neutral |
| | 死者の国 (the land of the dead) | Low | Low | High | High |
| | ガイコツ姿 (The skeletons) | High | Low | High | High |
| S5: ガイコツ姿で個性的ですか! ますます想像を超えてきて、ぜひ観てみたくなりました。 そんな中で少年のミュージシャンへの夢がどうなるのかすごく気になります。 (Skeletons with personalities? This is getting beyond what I imagined, and I definitely want to watch it now. I'm very curious about what happens with the boy's dream of becoming a musician.) | ガイコツ姿 (Skeletons) | High | Low | High | High |
| | 少年のミュージシャンへの夢 (the boy's dream of becoming a musician.) | High | Low | High | High |
| R5: 少年があこがれているミュージシャンがいて、そのミュージシャンの歌を練習しているのですが、「音楽禁止」の家庭で育ったので、隠れて練習していました。死者の国で、そのミュージシャンに会うために音楽コンテストに出ます。そのときに少年の歌声が披露されるのですが、すごく美声なんですよ! (The boy admires a musician and practices his songs, but he grows up in a family that has banned music, so he practices in secret. In the land of the dead, he enters a music contest to meet that musician. His singing voice is revealed during the contest, and it's incredibly beautiful! | あこがれているミュージシャン (a musician) | High | Neutral | High | Neutral |
| | そのミュージシャンの歌 (his songs) | High | Low | High | High |
| | 「音楽禁止」の家庭 (a family that has banned music) | Low | Neutral | High | Neutral |
| | 音楽コンテスト (a music contest) | High | Low | High | Neutral |
| | 少年の歌声 (His singing voice) | Low | Neutral | High | High |
| S6:現実的な話になりますが、その少年の声を担当しているのは有名な歌手なんですか? (Getting back to reality for a moment, is the boy's voice provided by a famous singer?) | その少年の声 (the boy's voice) | Low | High | High | High |
| | 有名な歌手 (a famous singer) | High | High | High | High |
| R6:私は存じませんでしたが、アンソニー・ゴンザレスという当時15歳の少年だそうです。 アンソニーは幼いころから歌手や俳優を目指し、スペイン語の番組やアメリカのドラマ、短編映画などに出演していたそうです。(I wasn't aware of this, but it turns out his voice is provided by Anthony Gonzalez, a boy who was 15 at the time. Anthony has been pursuing a career in singing and acting from a young age, appearing in Spanish-language programs, American dramas, and short films.) | アンソニー・ゴンザレスという当時15歳の少年 (Anthony Gonzalez, a boy who was 15 at the time.) | Low | Low | High | High |
| | アンソニー (Anthony) | Low | Low | High | High |
| | 歌手や俳優 (singing and acting) | High | High | Neutral | High |
| | スペイン語の番組やアメリカのドラマ、短編映画 (Spanish-language programs, American dramas, and short films.) | Low | Low | Neutral | Neutral |
| S7:そうなんですね! 映画の主人公そのものの少年が歌っているなんて ますます聞きたくなります。 少年以外にも歌をうたうキャストはいるのですか? (Really! It's even more intriguing to know that a boy resembling the movie's protagonist is singing. Are there other characters who sing in the movie?) | 映画の主人公そのものの少年 (a boy resembling the movie's protagonist is singing) | High | High | High | High |
| | 歌をうたうキャスト (characters who sing in the movie) | Low | Low | High | High |
| R7:はい、少年のあこがれのミュージシャンのデラクルスや、死者の国で一緒に歌うヘクターというガイコツです。ヘクターが死者の国を案内してくれるのですが、ヘクターの秘密も明かされていきます。そして、最後には、なぜ少年の家が音楽禁止になったのかも分かり、最後にはその掟が廃止され、少年は自由に音楽を楽しむことができるようになるんです。(Yes, there's De la Cruz, the musician the boy idolizes, and Hector, a skeleton he sings with in the land of the dead. Hector guides him through the land of the dead, and Hector's secret is revealed. In the end, it's revealed why the boy's family banned music, and the ban is lifted, allowing the boy to freely enjoy music.) | 少年のあこがれのミュージシャンのデラクルスや、死者の国で一緒に歌うヘクターというガイコツ (De la Cruz, the musician the boy idolizes, and Hector, a skeleton he sings with in the land of the dead) | Low | Low | High | High |
| | ヘクター (Hector) | Low | Low | High | High |
| | 死者の国 (the land of the dead) | High | Low | High | Neutral |
| | ヘクターの秘密 (Hector's secret) | Low | Neutral | High | High |
| | 少年の家 (the boy's family) | Low | Low | High | High |
| | その掟 (the ban) | Low | Low | High | High |
| S8:確かに、音楽禁止だった理由は気になるところです。ガイコツの秘密、というのも想像つかないし、 憧れのミュージシャンの歌を聴くのも楽しみですし、 さっそく観てみたいと思います! ちなみに音楽はすべて映画オリジナルですか? (Indeed, I'm curious about why music was banned. The skeleton's secret sounds intriguing, and I'm looking forward to hearing the idolized musician's songs. I'll watch it soon! Are all the songs in the movie original?) | 音楽禁止だった理由 (why music was banned) | Low | Low | High | High |
| | ガイコツの秘密 (The skeleton's secret) | Low | Low | High | High |
| | 憧れのミュージシャンの歌 (the idolized musician's songs) | High | Low | High | High |
| R8: おそらく、すべてオリジナルだと思います。 特に有名なのは、主題歌である「リメンバー・ミー」という曲です。 あと、音楽コンテストでガイコツと歌う「ウン・ポコ・ロコ」という歌が、私は大好きです。とてもコミカルで、初めて出逢ったガイコツと息の合った歌声を披露してくれます。(I believe all the songs are original. Especially famous is the main theme song, "Remember Me". Another song I love is "Un Poco Loco", sung during the music contest with a skeleton. It's very comical and showcases a perfect harmony between the newly met skeleton and the boy.) | 主題歌である「リメンバー・ミー」という曲 (the main theme song, "Remember Me") | Low | Low | High | Neutral |
| | 音楽コンテスト (the music contest) | High | Low | High | High |
| | ガイコツ (a skeleton) | High | High | High | High |
| | 歌う「ウン・ポコ・ロコ」という歌 ("Un Poco Loco") | Low | Low | High | High |
| | 出逢ったガイコツと息の合った歌声 (a perfect harmony between the newly met skeleton and the boy.) | Low | Low | High | High |
| S9: では、「ウン・ポコ・ロコ」に特に注目して観てみますね! 今日は目新しい楽しい映画を紹介していただいてありがとうございました。(Then, I'll pay special attention to "Un Poco Loco" when I watch it! Thank you for introducing me to a new and fun movie today.) | 「ウン・ポコ・ロコ」 ("Un Poco Loco") | Low | Low | High | High |
| | 目新しい楽しい映画 (a new and fun movie) | High | High | High | High |
| R9:こちらこそ、ありがとうございました! (Thank you as well!) | - | - | - | - | - |

Figure 7: An example of RecomMind. R and S denote the recommender and seeker, respectively. The Entity column lists the entities extracted from the dialogue. Each entity has first- and second-person labels for knowledge and interest.

*(Input)*
## Task instruction
You are about to recommend a movie to a user in Japanese. Please make your response keeping in mind the following points:
- Find topics that the user has no knowledge of but has an interest in, and actively mention them, such as providing information.
- Keep your response brief and not too long.
- Do not repeat the same information as the dialogue history.
- Refer to the movie information as needed.

## Movie information
タイトル: アイアンマン (*Title: Iron Man*)
公開年度: 2008年9月27日 (*September 27, 2008*)
原作: スタン・リー,ラリー・リーバー,ドン・ヘック,ジャック・カービー『アイアンマン』 (*Based on: "Iron Man" by Stan Lee, Larry Lieber, Don Heck, Jack Kirby*)
製作国: アメリカ合衆国 (*Country: United States*)
監督: ジョン・ファヴロー (*Director: Jon Favreau*)
キャスト: ロバート・ダウニー・ジュニア、テレンス・ハワード、ジェフ・ブリッジス、グウィネス・パルトロー、ジョン・ファヴロー、ショーン・トーブ、クラーク・グレッグ (*Cast: Robert Downey Jr., Terrence Howard, Jeff Bridges, Gwyneth Paltrow, Jon Favreau, Shaun Toub, Clark Gregg*)
ジャンル: SF、アクション (*Genre: Science Fiction, Action*)
あらすじ: 巨大軍需産業"スターク・インダストリーズ"の社長であるトニー・スタークは、自身が開発したクラスターミサイル"ジェリコ"のプレゼンテーションの為に、アフガニスタンにいる親友のジェームズ・"ローディ"・ローズ率いるアメリカ空軍を訪問した。しかし、車で移動中にテロ組織"テン・リングス"がトニーを襲撃、助けを呼ぼうとするトニーにミサイルを撃つ。、命中の瞬間、トニーはミサイルに刻まれた自社のロゴマークを目撃し、爆風で吹き飛ばされ意識を失う。(*Plot: Tony Stark, the CEO of the massive the defense contractor "Stark Industries", visits the U.S. Air Force in Afghanistan, led by his close friend James "Rhodey" Rhodes, to present his newly developed cluster missile "Jericho." However, while traveling by car, the terrorist organization "Ten Rings" attacks Tony. As they fire a missile at him, Tony notices the missile is branded with his company's logo just before it hits, blowing him away and knocking him unconscious.*)

## Dialogue history
you: 素敵な俳優さんばかりですね。私もロバート・ダウニーJr.大好きです。嬉しいです!ではロバート・ダウニー・Jr.の映画は全てご覧になってますか? (*They'are all wonderful actors, aren't they? I also love Robert Downey Jr. I'm so happy! Have you seen all of Robert Downey Jr.'s movies?*)
user: 好きな俳優が同じでほんとうれしいです!ロバート・ダウニーJrの映画はそんなに多く観てないです。特にマーベルシリーズみたいなのは、なかなか手を出しにくいというか、なじみがなくて観ていません。(*I'm really happy we like the same actor! I haven't watched that many movies of Robert Downey Jr. Especially things like the Marvel series, I find it a bit hard to get into, I haven't watched them.*)
you: ロバート好きな方とお話しできて本当に嬉しいです!マーベルは確かに馴染みないと手を出しにくいですよね。でも『アイアンマン』はロバート好きなら絶対に気に入ると思います。主人公のトニー・スタークがロバートにぴったりなんです! (*I'm truly happy to talk with someone who likes Robert! Marvel can indeed be hard to get into if you're not familiar. But if you like Robert, I think you'll definitely like "Iron Man." The protagonist Tony Stark fits Robert perfectly!*)
user: そうなんですね!『アイアンマン』はシリーズ化されていると思いますが、一つ一つ話は別物ですか?基本的にアイアンマンが悪者をやっつける、みたいなストーリーなんでしょうか? (*Is that so! I believe "Iron Man" is a series, but is each story separate? Is it basically about Iron Man defeating the villains?*)

## Your answer
シリーズ化されていて1つ1つは別物ですが、設定などは引き継がれているので最初から見る方が良いと思います。おっしゃる通りで、基本的には悪者をやっつける感じなのですが、少し恋愛要素もあったりミステリー要素もあったりという感じです。『シャーロック』がお好きなら大丈夫だと思います。 (It's serialized, and while each installment is its own thing, the settings and such carry over, so I think it's better to watch from the beginning. As you said, it's basically about defeating the villains, but there's also a bit of romance and mystery elements involved. If you like "Sherlock," I think you'll be okay.)

{Few-shot Example2}

## Movie information
タイトル：ローマの休日 (*Title: Roman Holiday*)
公開年度: 1954年4月21日 (*April 21, 1954*)
製作国: アメリカ合衆国 (*Country: United States*)
監督: ウィリアム・ワイラー (*Director: William Wyler*)
キャスト: グレゴリー・ペック、オードリー・ヘプバーン、エディ・アルバート (*Cast: Gregory Peck, Audrey Hepburn, Eddie Albert*)
ジャンル: コメディ、ロマンス (*Genre: Comedy, Romance*)
あらすじ: ヨーロッパきっての古い歴史と伝統を持つ某国の王位継承者であるアン王女は、ヨーロッパ各国を表敬訪問中であった。 . . .(*Plot: Princess Anne, the heir to the throne of a certain country known for its ancient history and traditions among European nations, was on a courtesy visit to various countries in Europe. . . .*)

## Dialogue history
you: 早速ですが、最近の映画がお好きですか?それとも昔の映画などもよくご覧になりますか? (*Just to dive right in, do you prefer recent movies, or do you also watch older movies?*)
user: 最近はほとんど映画を見ていません。3年以上見ていないです。以前は昔の映画を見ることもありました。 (*I haven't really watched any movies lately. It's been over three years since I last watched one. I used to watch old movies as well.*)
you: 了解です!それでは、映画を観る時(サブスク・TVなど含み)、ご自身で選ぶ傾向の高いジャンルなどありますか？または気になる(好きな)俳優さんなどいれば、ぜひ聞かせてもらえたら助かります。 (*Got it! When you do watch movies (including via streaming services, TV, etc.), do you tend to favor any particular genres? Also, if there are any actors you're particularly fond of, I'd love to hear about them.*)
user: コメディやラブコメ等笑えるものが好きです。後は歴史の勉強もかねて歴史物の映画を見ることもあります。好きな俳優さんは特にこの人!という人はおらず内容重視です。 (*I like things that make me laugh, like comedies and romantic comedies. I also watch historical films as a way to learn about history. I don't have a favorite actor; I focus more on the content.*)

## Your answer

---

*(Output)*
コメディやラブコメ、そして歴史物の映画がお好きなんですね。それなら、"ローマの休日"はいかがでしょうか。オードリー・ヘプバーンとグレゴリー・ペックが主演のこの映画は、ローマを舞台にしたラブコメディで、歴史的な名所もたくさん登場します。内容重視の方にはぴったりだと思います。 (*So, you like comedies and romantic comedies, don't you? Then how about the "Roman Holiday"? Starring Audrey Hepburn and Gregory Peck, this romantic comedy set in Rome features many historical landmarks. It's perfect for those who are content-oriented.* )

Figure 8: The prompt for *Baseline*. The task instruction, few-shot examples, and test example are colored to show. The second few-shot example is omitted due to the space limitation. *The English translations* next to the Japanese words in the few-shot examples and test example are not used in the experiment.

*(Input)*
## Task instruction
You are about to recommend a movie to a user in Japanese. Let's think step by step following these two steps to make your response:
Step 1. Organizing User's Level of Knowledge and Interest:
- Extract important entities (noun phrases) directly from the dialogue history (excluding previous examples and movie information).
- Estimate the user's level of knowledge and interest in each entity on a 3-point Likert scale (2: high, 1: neutral, 0: low).
Step 2. Generating Response:
- Find topics that the user has no knowledge of but has an interest in, and actively mention them, such as providing information.
-- Refer to organized results of Step 1.
- Keep your response brief and not too long.
- Do not repeat the same information as the dialogue history.
- Refer to the movie information as needed.

## Movie information
タイトル: アイアンマン (*Title: Iron Man*)
公開年度: 2008年9月27日 (*September 27, 2008*)
原作: スタン・リー,ラリー・リーバー,ドン・ヘック,ジャック・カービー『アイアンマン』(*Based on: "Iron Man" by Stan Lee, Larry Lieber, Don Heck, Jack Kirby*)
製作国: アメリカ合衆国 (*Country: United States*)
監督: ジョン・ファヴロー (*Director: Jon Favreau*)
キャスト: ロバート・ダウニー・ジュニア、テレンス・ハワード、ジェフ・ブリッジス、グウィネス・パルトロー、ジョン・ファヴロー、ショーン・トーブ、クラーク・グレッグ (*Cast: Robert Downey Jr., Terrence Howard, Jeff Bridges, Gwyneth Paltrow, Jon Favreau, Shaun Toub, Clark Gregg*)
ジャンル: SF、アクション (*Genre: Science Fiction, Action*)
あらすじ: 巨大軍需産業"スターク・インダストリーズ"の社長であるトニー・スタークは、自身が開発したクラスターミサイル"ジェリコ"のプレゼンテーションの為に、アフガニスタンにいる親友のジェームズ・"ローディ"・ローズ率いるアメリカ空軍を訪問した。しかし、車で移動中にテロ組織"テン・リングス"がトニーを襲撃、助けを呼ぼうとするトニーにミサイルを撃つ。、命中の瞬間、トニーはミサイルに刻まれた自社のロゴマークを目撃し、爆風で吹き飛ばされ意識を失う。(*Plot: Tony Stark, the CEO of the massive the defense contractor "Stark Industries", visits the U.S. Air Force in Afghanistan, led by his close friend James "Rhodey" Rhodes, to present his newly developed cluster missile "Jericho." However, while traveling by car, the terrorist organization "Ten Rings" attacks Tony. As they fire a missile at him, Tony notices the missile is branded with his company's logo just before it hits, blowing him away and knocking him unconscious.*)

## Dialogue history
you: 素敵な俳優さんばかりですね。私もロバート・ダウニーJr.大好きです。嬉しいです!ではロバート・ダウニー・Jr.の映画は全てご覧になってますか? (*They're all wonderful actors, aren't they? I also love Robert Downey Jr. I'm so happy! Have you seen all of Robert Downey Jr.'s movies?*)
user: 好きな俳優が同じでほんとうれしいです!ロバート・ダウニーJrの映画はそんなに多く観てないです。特にマーベルシリーズみたいなのは、なかなか手を出しにくいというか、なじみがなくて観ていません。(*I'm really happy we like the same actor! I haven't watched that many movies of Robert Downey Jr. Especially things like the Marvel series, I find it a bit hard to get into, I haven't watched them.*)
you: ロバート好きな方とお話しできて本当に嬉しいです!マーベルは確かに馴染みないと手を出しにくいですよね。でも『アイアンマン』はロバート好きなら絶対に気に入ると思います。主人公のトニー・スタークがロバートにぴったりなんです! (*I'm truly happy to talk with someone who likes Robert! Marvel can indeed be hard to get into if you're not familiar. But if you like Robert, I think you'll definitely like "Iron Man." The protagonist Tony Stark fits Robert perfectly!*)
user: そうなんですね!『アイアンマン』はシリーズ化されていると思いますが、一つ一つ話は別物ですか?基本的にアイアンマンが悪者をやっつける、みたいなストーリーなんでしょうか? (*Is that so! I believe "Iron Man" is a series, but is each story separate? Is it basically about Iron Man defeating the bad guys?*)

## Your answer
### Step 1. Organizing User's Level of Knowledge and Interest
ロバート・ダウニーJr (*Robert Downey Jr.*): high,high
ロバート・ダウニー・Jr.の映画 (*Robert Downey Jr.'s movies*): high,high
マーベル (*Marvel*): high,neutral
主人公のトニー・スターク (*protagonist Tony Stark*): neutral,neutral
### Step 2. Generating Response
シリーズ化されていて1つ1つは別物ですが、設定などは引き継がれているので最初から見る方が良いと思います。おっしゃる通りで、基本的には悪者をやっつける感じなのですが、少し恋愛要素もあったりミステリー要素もあったりという感じです。『シャーロック』がお好きなら大丈夫だと思います。(*It's serialized, and while each installment is its own thing, the settings and such carry over, so I think it's better to watch from the beginning. As you said, it's basically about defeating the villains, but there's also a bit of romance and mystery elements involved. If you like "Sherlock," I think you'll be okay.*)

{Few-shot Example2}

## Movie information
タイトル：ローマの休日 (*Title: Roman Holiday*)
公開年度: 1954年4月21日 (*April 21, 1954*)
製作国: アメリカ合衆国 (*Country: United States*)
監督: ウィリアム・ワイラー (*Director: William Wyler*)
キャスト: グレゴリー・ペック、オードリー・ヘプバーン、エディ・アルバート (*Cast: Gregory Peck, Audrey Hepburn, Eddie Albert*)
ジャンル: コメディ、ロマンス (*Genre: Comedy, Romance*)
あらすじ: ヨーロッパきっての古い歴史と伝統を持つ某国の王位継承者であるアン王女は、ヨーロッパ各国を表敬訪問中であった。...(*Plot: Princess Anne, the heir to the throne of a certain country known for its ancient history and traditions among European nations, was on a courtesy visit to various countries in Europe. ...*)

## Dialogue history
you: 早速ですが、最近の映画がお好きですか?それとも昔の映画などもよくご覧になりますか? (*Just to dive right in, do you prefer recent movies, or do you also watch older movies?*)
user: 最近はほとんど映画を見ていません。3年以上見ていないです。以前は昔の映画を見ることもありました。(*I haven't really watched any movies lately. It's been over three years since I last watched one. I used to watch old movies as well.*)
you: 了解です!それでは、映画を観る時(サブスク・TVなど含み)、ご自身で選ぶ傾向の高いジャンルなどありますか?または気になる(好きな)俳優さんなどいれば、ぜひ聞かせてもらえたら助かります。(*Got it! When you do watch movies (including via streaming services, TV, etc.), do you tend to favor any particular genres? Also, if there are any actors you're particularly fond of, I'd love to hear about them.*)
user: コメディやラブコメ等笑えるものが好きです。後は歴史の勉強もかねて歴史物の映画を見ることもあります。好きな俳優さんは特にこの人!という人はおらず内容重視です。(*I like things that make me laugh, like comedies and romantic comedies. I also watch historical films as a way to learn about history. I don't have a favorite actor; I focus more on the content.*)

## Your answer
### Step 1. Organizing User's Level of Knowledge and Interest

---

*(Output)*
最近の映画 (*recent movies*): low,low
昔の映画 (*older movies*): high,high
コメディやラブコメ等笑えるもの (*things that make me laugh, like comedies and romantic comedies*): high, high
歴史物の映画 (*historical films*): high, high
### Step 2. Generating Response
コメディやラブコメ、そして歴史物の映画がお好きなんですね。それなら、"ローマの休日"はいかがでしょうか。オードリー・ヘプバーンとグレゴリー・ペックが主演のこの映画は、ローマを舞台にしたラブコメディで、歴史的な名所もたくさん登場します。内容重視の方にはぴったりだと思います。(*So, you like comedies and romantic comedies, don't you? Then how about the "Roman Holiday"? Starring Audrey Hepburn and Gregory Peck, this romantic comedy set in Rome features many historical landmarks. It's perfect for those who are content-oriented.* )

Figure 9: The prompt for $CoT_{1st}$. The task instruction, few-shot examples, and test example are colored to show. The second few-shot example is omitted due to the space limitation. *The English translations* next to the Japanese words in the few-shot examples and test example are not used in the experiment.

| | |
|---|---|
| **Context** | ## Movie information<br>なし (*None*)<br><br>## Dialogue history<br>you: アクションはどんな映画がお好きですか? (*What kind of action movies do you like?*)<br>user: そうですね、007はよく見ました。ジャッキーの偶にです。日本のアクション俳優さんは逆にあまり存じないですけどね (*Well, I've often watched 007 movies. Sometimes Jackie's as well. I'm not very familiar with Japanese action actors, though.*)<br>you: 007シリーズ良いですよね。ジャッキー・チェンでしょうか?私も好きです。007シリーズはほとんどご覧になってますか? (*The 007 series is great, isn't it? Jackie Chan, right? I like him too. Have you seen most of the 007 series?*)<br>user: ちょっとしたファンです。(笑)全巻DVD持ってまして、偶に一人の時間に見てます (*I'm a bit of a fan, haha. I own the complete DVD collection and watch them from time to time when I'm alone.*) |
| **Baseline** | 007シリーズのファンなら、日本のアクション映画では「アウトレイジ」シリーズをおすすめします。北野武監督が演出と主演を務めており、ハードボイルドなヤクザ映画です。アクションシーンも見ごたえがありますよ。 (*If you're a fan of the 007 series, I recommend the "Outrage" series from Japan. It's directed and starred by Takeshi Kitano, and it's a hard-boiled yakuza movie. The action scenes are quite something to watch.*) |
| **$\mathbf{CoT_{1st}}$** | アクション (*action*): high,high<br>007 (*007*): high,high<br>ジャッキー・チェン (*Jackie Chan*): high,high<br>日本のアクション俳優 (*Japanese action actors*): low,neutral<br>全巻DVD (*the complete DVD collection*): high,high<br><br>007シリーズのファンとのこと、素晴らしいですね。日本のアクション俳優については、たとえば高倉健さんや千葉真一さんなどが有名です。特に千葉真一さんは、日本だけでなく海外でも高い評価を受けているアクション俳優で、そのアクションシーンは見応えがありますよ。もし興味があれば、彼らの映画を観てみてはいかがでしょうか。(*It's wonderful that you're a fan of the 007 series. Regarding Japanese action actors, for example, Ken Takakura and Shinichi Chiba are famous. Shinichi Chiba, in particular, is an action actor who is highly regarded not only in Japan but also overseas, and his action scenes are impressive. If you're interested, maybe you could watch some of their movies.*) |

Table 9: Response generation example.