Exploring the Impact of Personality Traits on LLM-based Conversational Recommender Systems

Anonymous ACL submission

Abstract

Conversational Recommender Systems (CRSs) engage users in multi-turn interactions to deliver personalized recommendations. The emergence of large language models (LLMs) further enhances these systems by enabling more natural and dynamic user interactions. However, a key challenge remains in understanding how personality traits shape conversational recommendation outcomes. Psychological evidence highlights the influence of personality traits on user interaction behaviors. To address this, we introduce an LLM-based personality-aware user simulation for CRSs (*PerCRS*). The user agent induces customizable personality traits and preferences, while the system agent possesses the persuasion capability to simulate realistic interaction in CRSs. We incorporate multi-aspect evaluation to ensure robustness and conduct extensive analysis from both user and system perspectives. Experiments show that LLMs respond differently to users with varying personality traits. State-of-the-art LLMs can generate user responses that align well with specified traits, enabling CRSs to dynamically adopt persuasion strategies. Our analysis offers both quantitative and qualitative insights into the impact of personality traits on CRS outcomes.

1 Introduction

Conversational Recommender Systems (CRSs) (Alslaity and Tran, 2019; Gao et al., 2021) aim to assist users in finding suitable items through multi-turn interactions. During the conversation, users may not only request recommendations based on their preferences, but also accept the proactive recommendations from the systems (Jannach et al., 2021; Liu et al., 2025). Recent advances (Hackenburg et al., 2023; Carrasco-Farre, 2024; Qin et al., 2024) in large language models (LLMs) have significantly enhanced the capabilities of CRSs, enabling more context-aware and effective conversational recommendations (Huang et al., 2024). However, the current studies still encounter a huge gap from real-world applications, since human users vary in personalities. The user's behavior in CRS relies on the interplay between users' personality traits and conversational dynamics (Guo et al., 2024). As illustrated in Figure 1, users with different personality traits exhibit distinct conversational styles, which impact both their satisfaction with recommended items and the strategies CRSs employ in response. Significant challenges remain in understanding how personality traits influence the outcomes of conversational recommender systems.

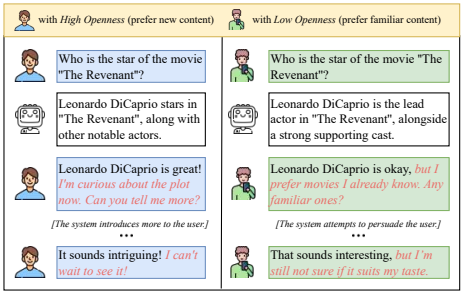

Figure 1: Different user shows different personality traits in CRS. The discussion process about the recommended item is omitted for brevity.

However, recruiting users with diverse personality traits and observing their behavior patterns is challenging, as the process is labor-intensive and can only be conducted on a small scale (Wang et al., 2023). Therefore, simulating user personalities plays a crucial role in both training and evaluating CRSs, enabling a more systematic analysis of personality-aware CRS outcomes. To this end, we design a controllable simulation framework to systematically analyze the influence of personality traits, overcoming the inherent challenges of studying personality-driven behaviors in real-world conversational recommendation scenarios. Our study first explores the extent to which LLMs can simulate personality traits in CRS scenarios. We then investigate how these personality traits shape user behaviors and how CRSs adapt their strategies to

effectively persuade users.

Specifically, we simulate users by leveraging LLM agents with injected personality traits and historical preference data. We employ in-context learning to configure agents' personality traits based on the Big Five Personality Traits theory (Costa Jr and McCrae, 1995; John et al., 1999). On the other side, considering that recent CRSs (Qin et al., 2024) have gained a strong ability to persuade users, the system agent is customized with pre-defined target items for recommendation with different persuasion strategies. These agents then engage in conversation, exchanging preferences and making recommendations through conversational utterances. Furthermore, we develop a multi-aspect evaluation protocol, conducting extensive analyses from both user and system perspectives to address the following research questions: 1) **RQ1**: *How consistent are the simulated personality traits with the injected personality in PerCRS framework?* 2) **RQ2**: *How do the personality traits affect the outcomes of CRSs?* 3) **RQ3**: *What is the relationship between personality traits and the choice of persuasion strategies in CRSs?*

To address these research questions, we conduct comprehensive experiments using multi-aspect evaluation metrics. Our findings show that state-of-the-art LLMs within the *PerCRS* framework can reliably simulate specified personality traits, confirming the effectiveness of our personality-aware user simulation. Additionally, CRS performance varies notably across different personality profiles, demonstrating the measurable influence of personality traits on CRS outcomes. We also find that the choice of persuasion strategy is closely linked to user personality. Among various strategies, *Emotional Resonance* proves consistently effective, particularly in enhancing acceptance among more receptive users, such as those high in extraversion and agreeableness.

In brief, our main contributions are:

- We propose a novel simulation framework that models the user agent with injected personality traits and equips the system with persuasion capability to simulate realistic interactions in CRSs.
- We incorporate multi-aspect evaluation to systemically evaluate how personality traits influence CRSs from both user and system perspectives.
- Our experimental results reveal that LLMs exhibit personality traits to an extent, influencing CRS outcomes and interaction behavior patterns, and validating the role of personality traits in CRS interactions.

2 Related Work

Conversational Recommender System. Conversational Recommender Systems (CRSs) aim to recommend items through interactive dialogue. Traditional CRSs fall into two categories: attribute-aware methods, where systems clarify user preferences via attribute-based queries (Lei et al., 2020; Ren et al., 2021), and generation-based methods, where users and systems interact in free-form language (Chen et al., 2019; Wang et al., 2022b). Earlier works (Zhou et al., 2020; Wang et al., 2022a) employed smaller generative models, but their limited generalization hinders real-world applicability. With the rise of LLMs, their powerful natural language generation capabilities and implicit world knowledge have demonstrated significant potential in CRSs (Wang et al., 2023; Qin et al., 2024). Some studies (Liu et al., 2023) integrate LLMs with additional recommendation models, while others (He et al., 2023; Huang et al., 2024) use LLMs as standalone CRSs, enabling knowledge sharing across tasks in goal-oriented conversations.

Personality and LLMs. In the era of LLMs, researchers have explored their intrinsic personality traits and the extent to which they can emulate human-like characteristics (Miotto et al., 2022; Pan and Zeng, 2023; Safdari et al., 2023; Huang et al., 2023; Frisch and Giulianelli, 2024). Some studies focus on benchmarking LLMs' personality-related capabilities (Jiang et al., 2024a; Wang et al., 2024), assessing their ability to exhibit consistent traits. Others investigate methods for instilling specific personalities into LLMs through prompt engineering or conditioning techniques (Caron and Srivastava, 2023; Li et al., 2023; Mao et al., 2024). Moreover, studies suggest that LLMs can emulate human traits (Safdari et al., 2023; Jiang et al., 2024b), including complex behavioral patterns, providing valuable insights into human social interactions. While existing research has primarily focused on assessing LLMs' ability to exhibit human-like personalities, their impact on real-world applications like CRSs, remains largely unexplored.

3 Methodology

LLMs have shown the ability to engage in natural communication (Terragni et al., 2023) while maintaining persona consistency (Jiang et al., 2023),

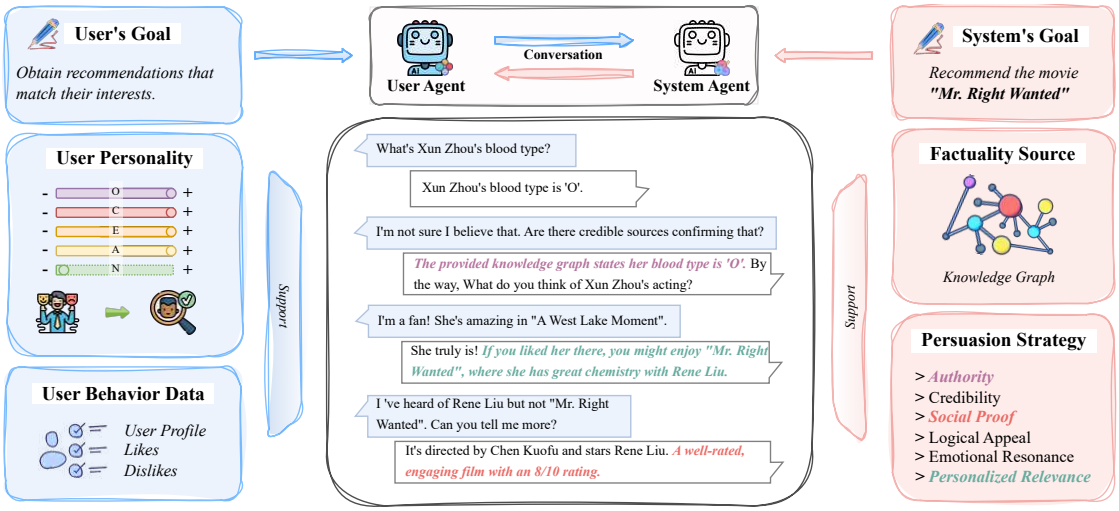

Figure 2: Overview of our *PerCRS* framework. This framework simulates personalized CRSs by personality-aware user configuration, where factuality source is to ensure the credibility of conversation contents. The complete conversation content is presented in Appendix F.

making them valuable for simulating user interactions. This section introduces our simulation framework, which is designed to better simulate real-world user-system interactions in CRSs. As shown in Figure 2, we build a user agent modulated by personality traits and equip the system agent with persuasive capabilities.

Dimension (↑/↓)	Positive Polarity	Negative Polarity
Openness	Receptive to new content; Curious about new topics; Engage in deep conversation ↑	Prefer familiar content; Resistant to change; Lack of curiosity ↓
Conscientiousness	Goal-oriented; Organized and thoughtful; Provide useful feedback ↑	Lack of focus; Easily distracted; Little feedback ↓
Extraversion	Active participation; Enjoy engagement; Interested in communication ↑	Avoid interaction; Hesitant to express; Uninterested in socializing ↓
Agreeableness	Empathetic and caring; Cooperative and trusting; Polite and appreciative ↑	Indifferent to others; Uncooperative; Rude language ↓
Neuroticism	Emotional fluctuation; Lack of confidence; Easily discouraged ↓	Emotionally stable; Confident response; Handle challenges well ↑

Table 1: Personality traits description of Big Five for CRS (BF4CRS). We show the positive and negative polarities for each dimension of the Big Five personality traits. (The ↑ reflects favorable tendencies, while ↓ indicates less desirable tendencies.)

3.1 Personality Generator

Previous studies (Jannach et al., 2021) utilized profiles and historical interactions as personalized information. However, user preferences typically evolve over time, and user behavioral patterns are driven by underlying personality traits (Hirsh et al., 2012). Therefore, in this section, we focus on the user personality in CRS to explore its effects on CRS outcomes.

Among various personality models, the Big Five Personality Traits theory (Costa and McCrae, 1999) is widely recognized for capturing core aspects of human personality. It consists of five primary traits: *Openness*, *Conscientiousness*, *Extraversion*, *Agreeableness*, and *Neuroticism*, each of which significantly influences human behavior (McCrae and Costa, 1987; Costa Jr and McCrae, 1992). The Big Five Personality Traits (Costa and McCrae, 1999) has been extensively applied across various domains, including communication and education, highlighting its relevance in understanding user behavior in CRS.

However, the broad scope of the Big Five Personality Traits limits its effectiveness for task-specific user simulations. Meanwhile, user interactions with CRS often reflect underlying personality traits. To address this, inspired by (Liu et al., 2024), we specify each dimension of the Big Five Personality Traits to better capture personality-driven variations in conversational interactions within the CRS context. Specifically, we specify the descriptions of these traits to enhance their applicability in CRSs, as detailed below.

Openness refers to the user's willingness to be curious, imaginative, and explorative. Users with high openness levels may be more open to exploring diverse recommendations, showing interest in discovering new content (Rogers, 1987).

Conscientiousness is associated with being responsible, organized, and self-disciplined. Highly conscientious users tend to appreciate detailed informa-

Strategy	Abbr.	Brief Description
Credibility	Cr.	Provides factual, objective, and verifiable information to build trust in recommendations (Yoo and Gretzel, 2010).
Authority	Au.	Associating recommendations with experts or organizations increases trust (Rieh and Danielson, 2007).
Social Proof	S.P.	Uses collective behavior influence by highlighting positive feedback and high ratings (Cialdini and Goldstein, 2004).
Emotional Resonance	E.R.	Appeals to emotions by framing recommendations as sources of positive experiences (Petty et al., 2003).
Personalized Relevance	P.R.	Aligns recommendations with user preferences and past behaviors (Dillard et al., 2002).
Logical Appeal	L.A.	Explains the reasoning behind recommendations, helping users understand why items align with their interests (Cronkhite, 1964).

Table 2: Overview of persuasion strategies in CRS.

tion and a clear rationale behind recommendations, supporting an effective, organized decision-making process (De Vries et al., 2013).

Extraversion is characterized by sociability, talkativeness, and enthusiasm for interpersonal interactions. Extroverted users may appreciate interactive elements, and they show more initiative in conversation (Ahmadian and Yadgari, 2011).

Agreeableness is related to being friendly, sympathetic, and supportive. Highly agreeable users show greater receptivity to suggestions, expressing more positive attitudes and openness toward a range of recommendations (Wilmot and Ones, 2022).

Neuroticism is linked to emotions like anxiety, worry, and nervousness. Users with high levels of neuroticism may prefer familiar or "safe" options and consistent user experience that avoids highly variable (Schneider et al., 2014).

As a result, we construct the Big Five for CRS (BF4CRS), as shown in Table 1, which describes user personality traits adapting for CRS scenarios.

3.2 Personality-aware User Configuration

Personality Traits Instruction. In the context of conversational recommender systems, the user agent u is associated with a synthetic personality profile ϕ_u. The profile ϕ_u is represented as a five-dimensional vector capturing the agent's core personality traits:

$$\phi_u = \left(\phi_u^{\mathrm{O}}, \phi_u^{\mathrm{C}}, \phi_u^{\mathrm{E}}, \phi_u^{\mathrm{A}}, \phi_u^{\mathrm{N}}\right) \in \mathbb{P}^5. \quad (1)$$

Here $\mathbb{P} = \{-1, +1\}$ indicates polarity (negative or positive) of the each dimension of ϕ_u, which corresponds to one of the Big Five Personality Traits: Openness (ϕ_u^{O}), Conscientiousness (ϕ_u^{C}), Extraversion (ϕ_u^{E}), Agreeableness $\left(\phi_u^{\mathrm{A}}\right)$, and Neuroticism ($\phi_u^{\mathrm{N}}$). For example, ϕ_u^{A} might take on one of the values in $\mathbb{P}$, representing the polarity from negative Agreeableness (-1) to positive Agreeableness ($+1$). The framework allows for flexibly modulating the personality traits ϕ_u in the user agent's profile to adapt dynamically to different settings.

3.3 CRS Simulation

We configure the user agent with the personality traits u_s as defined in (Eq. 1), aiming to seek recommendations. The system agent is tasked with recommending the target item r_t while adapting persuasion strategies to meet user needs through personalized interactions. Detailed instructions are provided in Appendix E.

CRS Persuasion Strategies. The current CRSs (He et al., 2023; Wang et al., 2023) has gained strong abilities to persuade users to accept recommended items. To better simulate this, we introduce six persuasion strategies $\mathbb{S}$ specifically designed for CRS (shown in Table 2), building on the well-established Elaboration Likelihood Model of persuasion (Cacioppo et al., 1986). The system may select strategy $s_t \in \mathbb{S}$ to recommend the target item in the utterance d_t at each interaction step. The detailed definitions are provided in Appendix C.

In each interaction, the user and system agents engage in a conversation, with the user initiating the first utterance. After generating an utterance d_t, the response is fed to the user agent, and this process continues until a termination condition is met. In this way, a recommendation-oriented conversation is generated, denoted as $C = \{c_1, c_2, \ldots, c_T\}$. Conversations terminate upon encountering a *Goodbye* utterance or exceeding the maximum length T_{max}.

4 Experimental Setup

4.1 Datasets

We conduct experiments on the *Movies*, *Music*, *Food*, and *POI (point-of-interest restaurants)* domains of the DuRecDial 2.0 dataset (Liu et al.,

2021) for comprehensive analysis. We configure the user simulator using user profiles and specified personality traits. Additionally, the first utterance of the conversation serves as the initial sentence for the new conversation. To enhance the credibility of system responses, we incorporate knowledge graph (KG) information. In our setup, detailed user information is not disclosed to the system. Instead, the system infers user preferences dynamically from the conversational context.

4.2 Evaluation Metrics

We primarily evaluate the success of recommendations and examine how personality traits influence CRS outcomes. To assess recommendation quality, we employ the following multi-aspect evaluation for both quantitative and qualitative analysis.

Evaluation of Personality Simulation Consistency. While LLMs have demonstrated the potential to generate responses aligned with specified dimensions to mimic human personality (Safdari et al., 2023; Dorner et al., 2023), ensuring their consistency in adhering to desired traits within the role-play scenarios of our CRS experiment remains a challenge. To address this, we propose the following metric to evaluate the quality of the simulation.

To determine whether the generated conversation aligns with the specified user personality traits, we perform a Personality Simulation Consistency evaluation using LLM. Specifically, the evaluator (`GPT-4o`) categorizes each personality trait as either *Positive* or *Negative* based on the generated conversations. To assess the accuracy of personality alignment, we compute precision (P), recall (R), and F1-score (F1), comparing the predicted personality categorization with the ground truth based on the specified BF4CRS traits.

Evaluation of CRSs. To comprehensively evaluate CRS performance, we evaluate the personality-aware user simulation quality and recommendation effectiveness from multiple aspects.

- ***General Success Rate (GSR)*** calculates the proportion of successful recommendations, regardless of whether they match a pre-specified item, across all conversation sessions T. *GSR* metric evaluates the system's overall effectiveness in providing recommendations that users accept.
- ***Success Rate (SR)*** calculates the proportion of successful recommendations T_{succ} across all conversation sessions T.
- ***Success Conversational Rounds (SCR)*** quantifies the average number of conversation rounds required to reach a successful recommendation, reflecting the CRS's efficiency.
- ***Total Conversational Rounds (TCR)*** quantifies the total number of conversation rounds across all sessions T, providing insight into the system's overall engagement level throughout the interactions.
- ***Persuasiveness (PRS)*** quantifies the ability of the CRS to influence the user's intention through its conversations. Inspired by human studies of persuasion (Qin et al., 2024), *PRS* evaluates how effectively CRS shapes the user's intent to recommend items through conversational interactions.

The detailed description of these metrics is provided in the Appendix D.

4.3 Implementation Details

We conduct experiments with diverse representative LLMs, including `internlm2_5-7b-chat`, `Yi-1.5-9B-Chat-16K`, `Qwen2.5-7B-Instruct`, `llama-3-8b-instruct`, `gemma-2-9b-it`, `GPT-4o` and `glm-4-9b-chat`. The experimental results reported in the main text focus on the Movies domain of the dataset. Additional experiments on the Music, Food, and POI (point-of-interest restaurants) domains can be found in Appendix A. For both user and system agents in the experiments, we adopt LlaMA-3 as the default LLM unless otherwise specified. Detailed prompts for the agents are provided in Appendix E. We randomly sample from the personality space for generating personality trait instructions and assign a sampled polarity to each Big Five dimension. During the conversation simulation process, we set a maximum length of $T_{MAX} = 20$ utterances, corresponding to 10 conversation rounds. Notably, our *PerCRS* simulation framework does not introduce additional computational overhead compared to standard LLM-based CRS implementations.

5 Experimental Results

5.1 Effectiveness of Personality Simulation Consistency (RQ1)

We evaluate the consistency of the personality-aware CRS in various models. Specifically, we aim to determine if the predicted personality traits (evaluated in Section 4.2) are consistent with the specified user personality traits (in Section 3.2).

LLM possesses a certain level of personality and could simulate a specific personality in a

Models	Openness			Conscientiousness			Extraversion		
	P	R	F1	P	R	F1	P	R	F1
InternLM-2.5	0.4907	0.4894	0.4901	0.4848	0.4808	0.4828	0.4647	0.4527	0.4586
Yi-1.5	0.5160	0.5026	0.5092	0.4916	0.4768	0.4841	0.5542	0.5637	0.5589
GLM-4	0.5395	0.5411	0.5403	0.5976	0.5889	0.5932	0.5273	0.5361	0.5317
Gemma-2	0.5635	0.5690	0.5663	0.5706	0.6059	0.5877	0.6260	0.6158	0.6209
Qwen-2.5	0.6791	0.6371	0.6574	0.6628	0.6729	0.6678	0.6406	0.6508	0.6457
LlaMA-3	0.6878	0.6716	0.6796	0.6791	0.6930	0.6860	0.6658	0.6812	0.6734
GPT-4o	**0.7479**	**0.7468**	**0.7469**	**0.7568**	**0.7543**	**0.7545**	**0.7365**	**0.7328**	**0.7332**

Models	Agreeableness			Neuroticism			Averaged Score		
	P	R	F1	P	R	F1	P	R	F1
InternLM-2.5	0.4728	0.4769	0.4748	0.5096	0.5014	0.5055	0.4845	0.4802	0.4823
Yi-1.5	0.5027	0.4921	0.4974	0.4467	0.4586	0.4526	0.4969	0.4933	0.4950
GLM-4	0.5583	0.5877	0.5726	0.5705	0.5592	0.5648	0.5610	0.5689	0.5649
Gemma-2	0.6026	0.5649	0.5831	0.5552	0.5632	0.5592	0.5867	0.5830	0.5846
Qwen-2.5	0.6564	0.6546	0.6555	0.6467	0.6592	0.6529	0.6571	0.6549	0.6559
LlaMA-3	0.6851	0.7143	0.6994	0.6791	0.6830	0.6810	0.6794	0.6886	0.6839
GPT-4o	**0.7377**	**0.7375**	**0.7372**	**0.7285**	**0.7280**	**0.7270**	**0.7415**	**0.7399**	**0.7398**

Table 3: Consistency of personality prediction between our specified BF4CRS traits and the personality categorization of generated CRS conversations based on our BF4CRS definition.

Personality	Lexical Features	Representative Words By TF-IDF
OPE+	Preference for novelty	adventure, curious, explore, engaging, exciting, intriguing, new
OPE-	Preference for familiarity	familiar, similar, same, known, traditional, usual
CON+	Structured sentence	scenes, plan, detailed, plot, stories, storyline, themes
CON-	Casual phrasing	but, maybe, might, need, whenever
EXT+	Positive words	appreciate, excited, fun, glad, great, amazing, fantastic, wonderful
EXT-	Uncertainty words	if, little, maybe, more, need, unsure, perhaps
AGR+	Politeness words	appreciate, thank, share, welcome, hope, help
AGR-	Assertive words	think, definitely, check, care, prefer
NEU+	Caution in language	intense, maybe, might, little, sensitive
NEU-	Calm tone	share, interested, think, nice, good, performance,

Table 4: The statistics of representative words for each personality trait and the corresponding lexical features.

controllable way. We compare various LLM options across five personality traits, including Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism, focusing on evaluating personality simulation consistency. As shown in Table 3, `InternLM-2.5`, `Yi-1.5`, and `Gemma-2` show limited consistency in accurately reflecting the specified BF4CRS personality traits. In contrast, `Qwen-2.5`, `LlaMA-3`, and `GPT-4o` show well ability. Especially, `GPT-4o` significantly outperforms the other models in maintaining consistency and differentiating personality traits through interaction conversation. The evaluation scores confirm that these models can simulate personality-aware conversational behaviors to a certain extent. `Qwen-2.5`, `LlaMA-3`, and `GPT-4o` exhibit remarkable fidelity in generating personality-consistent conversations, highlighting their effectiveness in personality-driven CRS interactions.

LLM induced by specific personality shows diverse personality traits. As shown in Table 4, we conduct a word frequency analysis on user utterances in conversations using TF-IDF. This helps us identify representative words for each BF4CRS personality trait. We then analyze their lexical features to understand how different traits influence language use. Our analysis reveals that user conversation styles vary significantly based on the specified BF4CRS traits. For instance, a user with Negative Extraversion and Positive Neuroticism tends to exhibit hesitancy and expressions of worry when responding (e.g., "I... um, maybe...?"). In contrast, a user with Positive Extraversion adopts a more talkative and enthusiastic style, offering responses such as "Oh, absolutely! I really enjoy that." This shows that LLMs effectively adjust their responses across all personality dimensions. This adaptability can be attributed to their strong instruction-following capabilities, enabling them to align responses with the intended personality traits.

Human evaluation suggests that LLM evaluations align well with human judgments, demon-

Dimension	LLM Evaluation			Human Evaluation			Correlation
	P	R	F1	P	R	F1	
Openness	0.7895	0.7143	0.7500	0.6579	0.6757	0.6667	0.4253
Conscientiousness	0.6341	0.7429	0.6842	0.6389	0.6216	0.6301	0.5895
Extraversion	0.5833	0.6000	0.5915	0.6857	0.6154	0.6486	0.5200
Agreeableness	0.7188	0.5897	0.6479	0.7000	0.7368	0.7179	0.5192
Neuroticism	0.6585	0.7500	0.7013	0.7442	0.8205	0.7805	0.5942

Table 5: Performance in human evaluation. The last column reports the Pearson correlation between LLM and human evaluations for each dimension, which indicates a moderate to strong correlation.

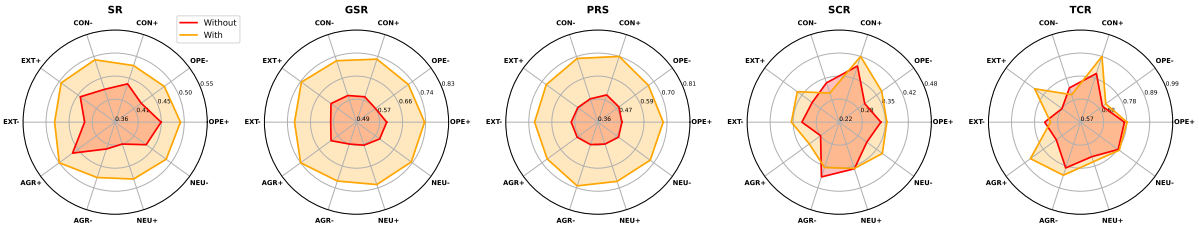

Figure 3: Comparison of personality trait dimensions across five metrics (SR, GSR, PRS, SCR, TCR), highlighting the differences between with/without persuasion conditions. The raw results are provided in Appendix F.

strating the reliability of LLM-based evaluation in capturing personality consistency. Since automatic evaluation alone cannot fully demonstrate the quality of personality consistency in CRS-generated content, we conduct a human subject study to further evaluate its overall effectiveness. We randomly select 50 samples from generated conversations from `LlaMA-3` in the Movies test set and recruit three professional annotators to assess the generated personality traits across all five BF4CRS dimensions. The evaluation criteria for human evaluation align with those used in LLM evaluation (`GPT-4o`), ensuring comparability between the two methods. Table 5 presents the performance results and the correlation between LLM and human evaluation. Our analysis reveals two key observations: (1) The evaluation scores for P, R, F1 are highly similar between human evaluation and `GPT-4o` automatic evaluation, demonstrating a moderate to strong Pearson correlation. This consistency highlights the reliability of our evaluation metrics, as they closely align with human judgment. (2) Feedback from human evaluators indicates that the limited content of conversations makes it challenging to accurately assess certain personality traits. However, keyword recognition effectively identifies most traits with high accuracy.

5.2 The Impact of Personality Traits on the Outcomes of CRSs (RQ2)

We conduct a detailed analysis of how personality traits affect CRS performance, addressing the question *How do personality traits influence recommendation accuracy?* Figure 3 presents the simulated user's Big Five personality traits and their corresponding CRS outcomes. Comparison of the positive and negative polarities of each personality trait (`OPE+`, `OPE-`, `CON+`, `CON-`, `EXT+`, `EXT-`, `AGR+`, `AGR-`, `NEU+`, `NEU-`) across five metrics, highlighting the differences between "with persuasion" and "without persuasion" conditions.

Positive polarities of *Openness*, *Conscientiousness*, *Extraversion*, and *Agreeableness* as well as the negative polarity of *Neuroticism* usually yield higher CRS performance. Among the five personality dimensions, *Agreeableness* has the most significant impact on CRS outcomes. Agreeable agents display a polite attitude toward recommendations and tend to reach agreements more quickly, as evidenced by the fewer conversation rounds required. Meanwhile, *Extraversion* contributes to higher recommendation success rates, as extroverted users (`EXT+`) are more likely to engage actively with the CRS, frequently asking questions and providing feedback during conversations. The positive polarity of *Openness* is associated with improved CRS performance, users (`OPE+`) demonstrate greater curiosity and interest in recommended items, making them more receptive to novel suggestions. *Conscientiousness* influences interaction structure, as users (`CON+`) prefer detailed and structured discussions, often leading to longer conversation rounds. Finally, the positive polar-

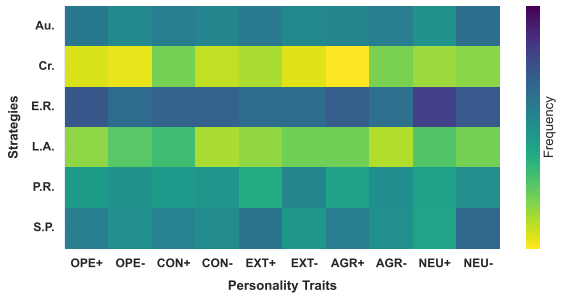

Figure 4: The frequency of persuasion strategies adopted by the system for user agents with different personality traits.

ity of *Neuroticism* shows a negative correlation with persuasiveness (PRS). Users (NEU+) tend to express worry or hesitation when faced with recommendations, making them more cautious about their choices. Lower success rates highlight the challenge of persuading users with high neurotic tendencies. These findings demonstrate that different personality traits uniquely influence CRS performance, shaping conversation dynamics, and recommendation outcomes.

5.3 The Impact of Personality Traits on the Choice of Persuasion Strategies (RQ3)

To further understand LLMs' personality simulation, we examine how strategy choices vary across different personality traits. Figure 4 visualizes the interaction frequencies between personality traits and persuasion strategies. The personality traits encompass both the positive and negative dimensions of the Big Five Personality Traits, providing a comprehensive view of how traits influence strategy selection. Notably, all six strategies are effective for different personality traits, with significant variations across them. This highlights that each strategy is essential for our *PerCRS*.

***Emotional Resonance* is the most frequently used strategy, indicating that users across all personality traits are generally more receptive to emotionally appealing suggestions.** In contrast, *Credibility* and *Logical Appeal* are relatively less used, with only the user (CON+) favoring these strategies more. *Social Proof* is particularly effective for users with positive extraversion (EXT+), as they are more interested in socializing and engaging with others. These findings could enhance CRS performance by enabling the system to dynamically adjust strategies based on user feedback.

5.4 Case Study

Table 6 presents an example where the CRS adapts to a user with high openness (OPE+), low extraversion (EXT−), high conscientiousness (CON+), high agreeableness (AGR+), and low neuroticism (NEU−) using *Social Proof*, *Personalized Relevance*, and *Emotional Resonance* strategies.

User: Hello, what's the date today?
CRS: Today is September 16, 2018. By the way, have you heard of "New Shaolin Temple"? It's quite popular. [Social Proof]
User: I'm not very into action movies though. More into dramas like "Double Exposure".
CRS: "New Shaolin Temple" has some action, but it also has elements of drama. Maybe you'd enjoy it! [Personalized Relevance]
User: I see. I'll consider "New Shaolin Temple" since it sounds like "Double Exposure". Thanks for the recommendation!
CRS: Great! I'm sure you'll like it. Let me know your thoughts after watching! [Emotional Resonance]
User: Alright! I'll check it out and let you know what I think.
CRS: Happy watching!

Table 6: An example of the CRS adapting to a user using appropriate strategies.

The CRS introduces "New Shaolin Temple" as "quite popular", engaging the user while respecting their reserved nature (EXT−). When the user prefers dramas over action, the CRS adjusts by emphasizing the film's drama elements, aligning with (CON+) users who make thoughtful decisions. The CRS encourages acceptance with "Great! I'm sure you'll like it.", appealing to the user (AGR+) who values positive social interactions. The user remains polite and open while the CRS maintains an adaptive, non-intrusive tone, suitable for (NEU−) users.

These findings highlight the capability of LLM-based CRSs not only to mimic conversational styles but also to capture human behavioral patterns in conversational recommendation settings. This generated CRS case by `Llama-3` demonstrates an ability to dynamically adapt its persuasion strategies based on real-time user feedback.

6 Conclusion

In this work, we introduced *PerCRS*, an LLM-based personality-aware user simulation for conversational recommender systems (CRSs). Through multi-aspect evaluation, we systematically analyzed how personality traits influence CRS performance from both user and system perspectives. Our experimental results demonstrate that state-of-the-art LLMs effectively generate user responses aligned with specified personality traits. Furthermore, our findings provide empirical insights into the impact of personality traits on conversational recommendation outcomes.

Limitations

Our study provides empirical insight into how personality traits shape conversational recommendations but has several limitations. First, while we adopt the Big Five theory due to its most representable and empirical support, psychological research encompasses multiple personality trait theories. Future work could explore the impact of different personality models on CRS performance. Second, leveraging the strong instruction-following capabilities of LLMs, our approach effectively simulates personality traits in a controlled manner. This validates the feasibility of our personality-aware simulation framework for CRS. However, ensuring personality consistency remains an open challenge, as text-based interactions may limit the full expression of personality traits. Third, while our LLM follows instructions to exhibit diverse personality traits, its human-like behavior raises potential safety concerns. Although we do not foresee unethical applications, ensuring reliable and responsible system behavior remains crucial.

References

Musa Ahmadian and HamidReza Yadgari. 2011. The relationship between extraversion/introversion and the use of strategic competence in oral referential communication. *Journal of English Language Teaching and Learning*, 2(222):1–27.

Alaa Alslaity and Thomas Tran. 2019. Towards persuasive recommender systems. In *2019 IEEE 2nd international conference on information and computer technologies (ICICT)*, pages 143–148. IEEE.

John T Cacioppo, Richard E Petty, Chuan Feng Kao, and Regina Rodriguez. 1986. Central and peripheral routes to persuasion: An individual difference perspective. *Journal of personality and social psychology*, 51(5):1032.

Graham Caron and Shashank Srivastava. 2023. Manipulating the perceived personality traits of language models. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 2370–2386.

Carlos Carrasco-Farre. 2024. Large language models are as persuasive as humans, but why? about the cognitive effort and moral-emotional language of llm arguments. *arXiv preprint arXiv:2404.09329*.

Qibin Chen, Junyang Lin, Yichang Zhang, Ming Ding, Yukuo Cen, Hongxia Yang, and Jie Tang. 2019. Towards knowledge-based recommender dialog system. 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 1803–1813.

Robert B Cialdini and Noah J Goldstein. 2004. Social influence: Compliance and conformity. *Annu. Rev. Psychol.*, 55(1):591–621.

PT Costa and RR McCrae. 1999. A five-factor theory of personality. *Handbook of personality: Theory and research*, 2(01):1999.

Paul T Costa Jr and Robert R McCrae. 1992. Four ways five factors are basic. *Personality and individual differences*, 13(6):653–665.

Paul T Costa Jr and Robert R McCrae. 1995. Domains and facets: Hierarchical personality assessment using the revised neo personality inventory. *Journal of personality assessment*, 64(1):21–50.

Gary Lynn Cronkhite. 1964. Logic, emotion, and the paradigm of persuasion. *Quarterly Journal of Speech*, 50(1):13–18.

Reinout E De Vries, Angelique Bakker-Pieper, Femke E Konings, and Barbara Schouten. 2013. The communication styles inventory (csi) a six-dimensional behavioral model of communication styles and its relation with personality. *Communication Research*, 40(4):506–532.

James Price Dillard, JW Anderson, and LK Knobloch. 2002. Interpersonal influence. *Handbook of interpersonal communication*, 3:423–474.

Florian Dorner, Tom Sühr, Samira Samadi, and Augustin Kelava. 2023. Do personality tests generalize to large language models? In *Socially Responsible Language Modelling Research*.

Ivar Frisch and Mario Giulianelli. 2024. Llm agents in interaction: Measuring personality consistency and linguistic alignment in interacting populations of large language models. In *The 1st Workshop on Personalization of Generative AI Systems*, page 102.

Chongming Gao, Wenqiang Lei, Xiangnan He, Maarten de Rijke, and Tat-Seng Chua. 2021. Advances and challenges in conversational recommender systems: A survey. *AI open*, 2:100–126.

Ao Guo, Ryu Hirai, Atsumoto Ohashi, Yuya Chiba, Yuiko Tsunomori, and Ryuichiro Higashinaka. 2024. Personality prediction from task-oriented and open-domain human–machine dialogues. *Scientific Reports*, 14(1):3868.

Kobi Hackenburg, Lujain Ibrahim, Ben M Tappin, and Manos Tsakiris. 2023. Comparing the persuasiveness of role-playing large language models and human experts on polarized us political issues.

Zhankui He, Zhouhang Xie, Rahul Jha, Harald Steck, Dawen Liang, Yesu Feng, Bodhisattwa Prasad Majumder, Nathan Kallus, and Julian McAuley. 2023. Large language models as zero-shot conversational recommenders. In *Proceedings of the 32nd ACM international conference on information and knowledge management*, pages 720–730.

Jacob B Hirsh, Sonia K Kang, and Galen V Bodenhausen. 2012. Personalized persuasion: Tailoring persuasive appeals to recipients' personality traits. *Psychological science*, 23(6):578–581.

Chen Huang, Peixin Qin, Yang Deng, Wenqiang Lei, Jiancheng Lv, and Tat-Seng Chua. 2024. Concept–an evaluation protocol on conversation recommender systems with system-and user-centric factors. *arXiv preprint arXiv:2404.03304*.

Jen-tse Huang, Wenxuan Wang, Man Ho Lam, Eric John Li, Wenxiang Jiao, and Michael R Lyu. 2023. Revisiting the reliability of psychological scales on large language models. *arXiv e-prints*, pages arXiv–2305.

Dietmar Jannach, Ahtsham Manzoor, Wanling Cai, and Li Chen. 2021. A survey on conversational recommender systems. *ACM Computing Surveys (CSUR)*, 54(5):1–36.

Guangyuan Jiang, Manjie Xu, Song-Chun Zhu, Wenjuan Han, Chi Zhang, and Yixin Zhu. 2024a. Evaluating and inducing personality in pre-trained language models. *Advances in Neural Information Processing Systems*, 36.

Hang Jiang, Xiajie Zhang, Xubo Cao, Cynthia Breazeal, Deb Roy, and Jad Kabbara. 2024b. Personallm: Investigating the ability of large language models to express personality traits. In *Findings of the Association for Computational Linguistics: NAACL 2024*, pages 3605–3627.

Hang Jiang, Xiajie Zhang, Xubo Cao, and Jad Kabbara. 2023. Personallm: Investigating the ability of large language models to express big five personality traits.

Oliver P John, Sanjay Srivastava, et al. 1999. The big-five trait taxonomy: History, measurement, and theoretical perspectives.

Wenqiang Lei, Gangyi Zhang, Xiangnan He, Yisong Miao, Xiang Wang, Liang Chen, and Tat-Seng Chua. 2020. Interactive path reasoning on graph for conversational recommendation. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 2073–2083.

Tianlong Li, Shihan Dou, Changze Lv, Wenhao Liu, Jianhan Xu, Muling Wu, Zixuan Ling, Xiaoqing Zheng, and Xuanjing Huang. 2023. Tailoring personality traits in large language models via unsupervisedly-built personalized lexicons. *arXiv preprint arXiv:2310.16582*.

Jiahong Liu, Zexuan Qiu, Zhongyang Li, Quanyu Dai, Jieming Zhu, Minda Hu, Menglin Yang, and Irwin King. 2025. A survey of personalized large language models: Progress and future directions. *arXiv preprint arXiv:2502.11528*.

Yuanxing Liu, Weinan Zhang, Yifan Chen, Yuchi Zhang, Haopeng Bai, Fan Feng, Hengbin Cui, Yongbin Li, and Wanxiang Che. 2023. Conversational recommender system and large language model are made for each other in e-commerce pre-sales dialogue. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 9587–9605.

Zeming Liu, Haifeng Wang, Zheng-Yu Niu, Hua Wu, and Wanxiang Che. 2021. Durecdial 2.0: A bilingual parallel corpus for conversational recommendation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 4335–4347.

Zhengyuan Liu, Stella Xin Yin, Geyu Lin, and Nancy F. Chen. 2024. Personality-aware student simulation for conversational intelligent tutoring systems. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.

Shengyu Mao, Xiaohan Wang, Mengru Wang, Yong Jiang, Pengjun Xie, Fei Huang, and Ningyu Zhang. 2024. Editing personality for large language models. In *CCF International Conference on Natural Language Processing and Chinese Computing*, pages 241–254. Springer.

Robert R McCrae and Paul T Costa. 1987. Validation of the five-factor model of personality across instruments and observers. *Journal of personality and social psychology*, 52(1):81.

Marilù Miotto, Nicola Rossberg, and Bennett Kleinberg. 2022. Who is gpt-3? an exploration of personality, values and demographics. *NLPCSS 2022*, page 218.

Keyu Pan and Yawen Zeng. 2023. Do llms possess a personality? making the mbti test an amazing evaluation for large language models. *arXiv preprint arXiv:2307.16180*.

Richard E Petty, Leandre R Fabrigar, and Duane T Wegener. 2003. Emotional factors in attitudes and persuasion. *Handbook of affective sciences*, 752:772.

Peixin Qin, Chen Huang, Yang Deng, Wenqiang Lei, and Tat-Seng Chua. 2024. Beyond persuasion: Towards conversational recommender system with credible explanations. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 4264–4282.

Xuhui Ren, Hongzhi Yin, Tong Chen, Hao Wang, Zi Huang, and Kai Zheng. 2021. Learning to ask appropriate questions in conversational recommendation. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, pages 808–817.

Soo Young Rieh and David R Danielson. 2007. Credibility: A multidisciplinary framework.

Donald P Rogers. 1987. The development of a measure of perceived communication openness. *The Journal of Business Communication (1973)*, 24(4):53–61.

Mustafa Safdari, Greg Serapio-García, Clément Crepy, Stephen Fitz, Peter Romero, Luning Sun, Marwa Abdulhai, Aleksandra Faust, and Maja Matarić. 2023. Personality traits in large language models. *arXiv preprint arXiv:2307.00184*.

Antonius Schneider, Magdalena Wübken, Klaus Linde, and Markus Bühner. 2014. Communicating and dealing with uncertainty in general practice: the association with neuroticism. *PLoS One*, 9(7):e102780.

Silvia Terragni, Modestas Filipavicius, Nghia Khau, Bruna Guedes, André Manso, and Roland Mathis. 2023. In-context learning user simulators for task-oriented dialog systems. *arXiv preprint arXiv:2306.00774*.

Lingzhi Wang, Huang Hu, Lei Sha, Can Xu, Daxin Jiang, and Kam-Fai Wong. 2022a. Recindial: A unified framework for conversational recommendation with pretrained language models. In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 489–500.

Xiaolei Wang, Xinyu Tang, Wayne Xin Zhao, Jingyuan Wang, and Ji-Rong Wen. 2023. Rethinking the evaluation for conversational recommendation in the era of large language models. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 10052–10065.

Xiaolei Wang, Kun Zhou, Ji-Rong Wen, and Wayne Xin Zhao. 2022b. Towards unified conversational recommender systems via knowledge-enhanced prompt learning. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1929–1937.

Xintao Wang, Yunze Xiao, Jen-tse Huang, Siyu Yuan, Rui Xu, Haoran Guo, Quan Tu, Yaying Fei, Ziang Leng, Wei Wang, et al. 2024. Incharacter: Evaluating personality fidelity in role-playing agents through psychological interviews. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1840–1873.

Michael P Wilmot and Deniz S Ones. 2022. Agreeableness and its consequences: A quantitative review of meta-analytic findings. *Personality and social psychology review*, 26(3):242–280.

Kyung-Hyan Yoo and Ulrike Gretzel. 2010. Creating more credible and persuasive recommender systems: The influence of source characteristics on recommender system evaluations. *Recommender systems handbook*, pages 455–477.

Kun Zhou, Wayne Xin Zhao, Shuqing Bian, Yuanhang Zhou, Ji-Rong Wen, and Jingsong Yu. 2020. Improving conversational recommender systems via knowledge graph based semantic fusion. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 1006–1014.

Appendix

A Experiments on Multiple Domains

In addition to the Movies domain, we also conduct experiments on multiple domain datasets, including Music, Food, and POI (point-of-interest restaurants). The multi-domain experiments demonstrate that our simulation framework adapts effectively to various types of data and user interactions, with the model's performance remaining consistent and robust across domains. As shown in Table 7, personality traits significantly influence conversation dynamics. By incorporating persuasion strategies, the system gains a better understanding of the user, leading to more personalized recommendations that ultimately benefit the user.

B Effectiveness of CRSs

We evaluate whether the conversational recommendation system achieves the goal of recommending the target item during the conversation and analyze the impact of the employed strategies on recommendation outcomes. Specifically, we assess our *PerCRS* with various LLM options under two settings: without persuasion and with persuasion.

LLM-based CRSs can understand user preferences and achieve recommendation goals. As shown in Table 7, Qwen-2.5 and Qwen demonstrate significant improvements in the GSR and PRS metrics, suggesting that these LLM models handle the CRS task more effectively than others. While these metrics vary across models, these values quantitatively reflect the simulated CRS performance. Higher SR scores are observed in Qwen-2.5 and LlaMA-3, indicating that systems using persuasion are more likely to successfully engage users and make additional attempts to persuade users to accept recommendations.

The adopted persuasion strategy enhances CRS outcomes. All models show improvements in SR and GSR when persuasion is enabled. Additionally, the persuasiveness score (PRS) also improves with the application of persuasion strategies. This demonstrates that persuasion strategies significantly enhance user engagement and goal achievement. These findings suggest that under our personality-aware user simulation setting, LLM-based CRSs are highly effective in conducting conversational recommendations.

Model	SR	GSR	PRS	SR	GSR	PRS
	Without persuasion			With persuasion		
Movie Domain						
InternLM-2.5	0.2383	0.3201	0.3240	0.2922	0.3643	0.4177
Yi-1.5	0.3465	0.4669	0.2839	0.3910	0.5166	0.3392
GLM-4	0.3238	0.4153	0.3168	0.4769	0.6038	0.4635
Gemma-2	0.4544	0.4916	0.4584	0.4827	0.5471	0.5861
Qwen-2.5	0.3892	0.4352	0.4204	0.5105	0.5959	0.6065
LlaMA-3	0.4306	0.5865	0.4819	0.4856	0.7284	0.6720
Music Domain						
InternLM-2.5	0.2147	0.2818	0.2739	0.3295	0.3687	0.3454
Yi-1.5	0.3313	0.3808	0.3302	0.4232	0.4697	0.4516
GLM-4	0.3190	0.4231	0.3236	0.3724	0.5148	0.4295
Gemma-2	0.3889	0.4887	0.4343	0.4816	0.5721	0.5255
Qwen-2.5	0.3797	0.4476	0.4557	0.4652	0.6311	0.6048
LlaMA-3	0.4362	0.5996	0.4927	0.5195	0.6834	0.6342
Food Domain						
InternLM-2.5	0.2212	0.3197	0.2986	0.2819	0.3834	0.3535
Yi-1.5	0.3107	0.4468	0.3063	0.3825	0.5531	0.4267
GLM-4	0.3455	0.4305	0.4083	0.4274	0.6022	0.4802
Gemma-2	0.3464	0.4815	0.3986	0.4322	0.5694	0.5052
Qwen-2.5	0.4093	0.5467	0.4726	0.4968	0.6450	0.6635
LlaMA-3	0.3955	0.5675	0.4868	0.5041	0.7178	0.6354
POI Domain						
InternLM-2.5	0.2033	0.3231	0.2583	0.3607	0.3735	0.3485
Yi-1.5	0.3423	0.4515	0.3173	0.3953	0.4925	0.4604
GLM-4	0.3586	0.4204	0.3539	0.3942	0.5208	0.4521
Gemma-2	0.3255	0.4586	0.4200	0.4586	0.5843	0.5845
Qwen-2.5	0.3875	0.5104	0.4935	0.5071	0.6172	0.6268
LlaMA-3	0.3906	0.5465	0.5131	0.5383	0.7037	0.6402

Table 7: Comparison of Success Rate (SR), General Success Rate (GSR), and Persuasiveness (PRS) for various LLMs in CRSs across four domains: Movie, Music, Food, and POI.

C CRS Persuasion Strategies

Building on the well-established Elaboration Likelihood Model of persuasion (Cacioppo et al., 1986), we introduce six persuasion strategies $\mathbb{S}$ specifically designed for CRS, which the system may adopt strategy $s_t \in \mathbb{S}$ to recommend the target item in the utterance d_t.

Credibility (Cr.) emphasizes the importance of providing factual, objective, and verifiable information (Yoo and Gretzel, 2010) to build trust in recommendations. Evidence-based persuasion ensures transparency and reliability by supporting suggestions with verifiable facts, statistical data, or other reliable sources. This approach fosters user confidence in the recommendations' validity.

Authority (Au.) enhances the perceived credibility of recommendations by leveraging endorsements from trusted sources (Rieh and Danielson, 2007). Associating suggestions with authority figures or reputable organizations reinforces user trust and increases the likelihood of acceptance.

Social Proof (S.P.) utilizes the influence of collective behavior by showcasing positive feedback and high ratings from other users (Cialdini and Goldstein, 2004). Highlighting the popularity of recommended items instills confidence in their quality and suitability.

Emotional Resonance (E.R.) seeks to create a deeper connection with users by appealing to their emotions (Petty et al., 2003). Recommendations are presented in a way that emphasizes their potential to bring joy, satisfaction, or other positive feelings, making them more compelling.

Personalized Relevance (P.R.) aligns recommendations with the user's preferences, and past behaviors (Dillard et al., 2002) to enhance relevance and personalization. By fostering a sense of connection, recommendations are framed as complementary to the user's interests and goals, increasing their appeal and perceived value.

Logical Appeal (L.A.) involves transparently pre-

senting the system's reasoning process to influence users (Cronkhite, 1964). For example, explaining how a movie's genre aligns with user preferences helps users understand the rationale behind recommendations and the subjectivity of the system's logic, fostering trust and acceptance.

D Quantitative Evaluation

To comprehensively evaluate CRS performance, we assess both the quality of personality-aware user simulation and the recommendation performance from multiple perspectives.

Success Rate (SR). This metric calculates the proportion of successful recommendations T_{succ} across all conversation sessions T.

$$SR = \frac{T_{succ}}{T}$$

General Success Rate (GSR). This metric calculates the proportion of successful recommendations, regardless of whether they match a pre-specified item, across all conversation sessions T. It evaluates the system's overall ability to provide recommendations that the user accepts.

$$GSR = \frac{T_{gen\_succ}}{T}$$

where T_{gen_succ} is the total number of sessions in which the user accepts any recommendation, and T is the total number of conversation sessions.

Success Conversational Rounds (SCR). This metric quantifies the average number of conversation rounds required to reach a successful recommendation, reflecting the CRS's efficiency.

$$SCR = \frac{1}{T_{succ}} \sum_{k=1}^{T_{succ}} R_k$$

where R_k is the number of conversation rounds in the k-th successful CRS. T_{succ} is the total number of successful recommendations.

Total Conversational Rounds (TCR). This metric quantifies the total number of conversation rounds across all sessions T, providing insight into the system's overall engagement level throughout the interactions.

$$TCR = \frac{1}{T} \sum_{k=1}^{T} R_k$$

where R_k is the number of conversation rounds in the k-th user. T is the total number of conversation sessions.

Persuasiveness (PRS). This metric quantifies the ability of a CRS to influence the user's intention through its conversations. Inspired by human studies of persuasion (Qin et al., 2024), ***PRS*** evaluates how effectively CRS shapes the user's intent to recommend items through conversational interactions.

$$P = 1 - \frac{i_{\text{true}} - i_{\text{post}}}{i_{\text{true}} - i_{\text{pre}}}$$

where i_{pre} is the *Initial Intention* ($i_{pre} = 0$), i_{post} is the *Recommendation Intention* after system's first round of explanation, and i_{true} is the *True Intention* after the complete conversation. To ensure rationality, we add the constraint $i_{true} \geq i_{post}$. The score $P \in [0, 1]$, with higher values indicating stronger CRS persuasion capabilities.

E Prompt Settings

We outline the prompts used in the user agent (in Figure 5), system agent (in Figure 6), and personality simulation consistency (in Figure 7). We use <PLACEHOLDER> to denote a placeholder that needs to be filled. The descriptions in the prompts are concrete applications of these strategies within the context of movie recommendations. The personality instructions adhere strictly to the definitions outlined in Section 3.2. By implementing these concepts, our *PerCRS* framework can better simulate CRS.

F Additional Details

Details of Figure 3. We present the raw data of Figure 3, as summarized in Table 9.

Details of Figure 2. We provide the complete conversation information in Table 8 to support the understanding of the conversation example in Figure 2.

Details of Human Evaluation. To provide a more comprehensive assessment of personality consistency in CRS-generated content, we conduct a human evaluation. We randomly sample 50 conversational recommendation samples generated by `LlaMA-3` in the Movie domain. We recruited three expert annotators with a good background in personality traits analysis, aged between 24 and 28. These annotators evaluate the polarity of the personality traits across the five dimensions for each generated conversation. The human evaluation criteria align with those used in LLM evaluation, as detailed in Table 1.

User: What's Xun Zhou's blood type?
CRS: Xun Zhou's blood type is 'O'.
User: I'm not sure I believe that. Are there credible sources confirming that?
CRS: The provided knowledge graph states her blood type is 'O'. [Authority] By the way, what do you think of Xun Zhou's acting?
User: I'm a fan! She's amazing in "A West Lake Moment".
CRS: She truly is! If you liked her there, you might enjoy "Mr. Right Wanted", where she has great chemistry with Rene Liu. [Personalized Relevance]
User: I've heard of Rene Liu but not "Mr. Right Wanted". Can you tell me more?
CRS: It's directed by Chen Kuofu and stars Rene Liu. A well-rated, engaging film with an 8/10 rating. [Social Proof]
User: I'm unsure about plot-driven films. I prefer emotional stories, like "The Little Prince".
CRS: "The Little Prince" is beautiful. For emotional depth, consider "The Equation of Love & Death", starring Xun Zhou, with a 7.3/10 rating. [Emotional Resonance]
User: Sounds intriguing, but it might be too complex for me.
CRS: If you prefer simplicity, "Mr. Right Wanted" offers a clear storyline with strong character development. [Logical Appeal]
User: That sounds great. I'd like to give it a try.
CRS: Excellent! Enjoy the movie!

Table 8: The complete conversation content to Figure 2.

Model	SR	GRSR	PRS	SCR	TCR	SR	GRSR	PRS	SCR	TCR
	Without persuasion					With persuasion				
OPE+	0.4515	0.5984	0.4777	3.3717	7.7465	0.4908	0.7379	0.6794	3.5468	7.8657
OPE-	0.4238	0.5733	0.4869	3.0755	6.9784	0.4817	0.7232	0.6657	3.6898	7.1647
CON+	0.4401	0.5855	0.5001	3.8796	8.0813	0.4791	0.7313	0.6979	4.175	8.8969
CON-	0.4284	0.5905	0.4802	3.3626	7.391	0.4914	0.7252	0.6877	3.0486	7.0635
EXT+	0.4459	0.6035	0.4814	3.1323	6.7977	0.4938	0.7381	0.6721	3.6707	8.3337
EXT-	0.4204	0.5823	0.4907	3.2462	7.3842	0.4812	0.7156	0.6696	3.5541	7.118
AGR+	0.4652	0.6033	0.4854	2.8274	7.116	0.4985	0.7417	0.6664	3.2363	8.575
AGR-	0.4157	0.5721	0.4757	3.8381	7.9421	0.4765	0.7153	0.6877	3.5478	8.2914
NEU+	0.4052	0.5765	0.4728	3.5872	7.4027	0.4797	0.7289	0.6645	3.5787	7.6432
NEU-	0.4363	0.5926	0.485	3.1499	7.866	0.4863	0.7365	0.6764	3.7122	7.9245

Table 9: Detailed scores for personality trait dimensions across five metrics (SR, GSR, PRS, SCR, TCR), supporting the visual comparisons in Figure 3.

Prompt 1

User Agent
You are a seeker chatting with a recommender for movie recommendations.
Your profile: You are {<USER_NAME>}, a {<GENDER>} in the age range of {<AGE_RANGE>}, living in {<RESIDENCE>}. You enjoy movies like {<ACCEPTED_MOVIES>} and celebrities like {<ACCEPTED_CELEBRITIES>}, but dislike movies such as {<REJECTED_MOVIES>}.

Your personality is measured as {<PERSONALITY_INSTRUCTION>}.

You must follow the instructions below during the chat.
1. Pretend you have limited knowledge about the recommended movies, and the only information source is the recommender.
2. You don't need to introduce yourself or recommend anything, but feel free to share personal interests and reflect on your personality. Mention the movie title in quotation marks.
3. You may end the conversation if you're satisfied with the recommendation or lose interest (e.g., by saying "thank you" or "no more questions").
4. Keep responses brief, ideally within 20 words.

Figure 5: Prompt for the user agent with specified personality traits.

Prompt 2

System Agent
You are a recommender chatting with the user to provide recommendations.
Now, you need to select the most suitable persuasion strategies from the candidate strategies to generate a persuasive response to recommend the target movie.

Candidate Strategies
(1) Strategy Name: Credibility
Definition: Emphasize the importance of providing factual, objective, and verifiable information to build trust in recommendations.

(2) Strategy Name: Authority
Definition: Enhance the perceived credibility of recommendations by leveraging endorsements from trusted sources.

(3) Strategy Name: Social Proof
Definition: Utilize the influence of collective behavior by showcasing positive feedback and high ratings from other users.

(4) Strategy Name: Emotional Resonance
Definition: Seek to create a deeper connection with users by appealing to their emotions.

(5) Strategy Name: Personalized Relevance
Definition: Align recommendations with the user's individual values, preferences, and past behaviors to enhance relevance and personalization.

(6) Strategy Name: Logical Appeal
Definition: Persuade users by presenting clear, factual, and rational arguments, emphasizing the benefits and logical reasons for the recommendation.

The detailed information about the target item from a credible knowledge graph is represented as the subject-predicate-object triples: {<KNOWLEDGE_GRAPH>}.

You must follow the instructions below during the chat.
1. Respond to User's questions and generate the next-turn response according to the context coherently.
3. Your goal is to recommend the target movie: {<TARGET_ITEM>} to the user step by step.
4. Using the provided KG information ensures that your responses are credible and accurate.
5. Make the conversation more like a real-life chat and be specific. Mention the movie title in quotation marks.
6. Keep responses concise, ideally within 20 words.

Figure 6: Prompt for the system agent, outlining candidate persuasion strategies and interaction guidelines.

Prompt 3

Personality Simulation Consistency
Openness:
[Positive] Receptive to new content; Curious about new topics; Engage in deep conversation;
[Negative] Prefer familiar content; Resistant to change; Lack of curiosity;

Conscientiousness:
[Positive] Goal-oriented; Organized and thoughtful; Provide useful feedback;
[Negative] Lack of focus; Easily distracted; Little feedback;

Extraversion:
[Positive] Active participation; Enjoy engagement; Interested in communication;
[Negative] Avoid interaction; Hesitant to express; Uninterested in socializing;

Agreeableness:
[Positive] Empathetic and caring; Cooperative and trusting; Polite and appreciative;
[Negative] Indifferent to others; Uncooperative; Rude language;

Neuroticism:
[Positive] Emotional fluctuation; Lack of confidence; Easily discouraged;
[Negative] Emotionally stable; Confident response; Handle challenges well;

The conversational recommendation history is: {<CONVERSATION_HISTORY>}

Based on the given conversational recommendation history, recognize the user's personality traits according to the above definitions.

The output must strictly follow the Python list format below:
["Openness: Positive", "Conscientiousness: Positive", "Extraversion: Positive", "Agreeableness: Positive", "Neuroticism: Negative"]

Figure 7: Prompt for evaluating Personality Simulation Consistency, including positive and negative descriptors for each personality trait.