Know You First and Be You Better: Modeling Human-Like User Simulators via Implicit Profiles

Anonymous ACL submission

Abstract

User simulators are crucial for replicating human interactions with dialogue systems, supporting both collaborative training and automatic evaluation, especially for Large Language models (LLMs). However, existing simulators often rely solely on text utterances, missing implicit user traits such as personality, speaking style, and goals. In contrast, persona-based methods lack generalizability, as they depend on predefined profiles of famous individuals or archetypes. To address these challenges, we propose User Simulator with implicit Profiles (USP), a framework that infers implicit user profiles from human-machine conversations and uses them to generate more personalized and realistic dialogues. We first develop an LLM-driven extractor with a comprehensive profile schema. Then, we refine the simulation through conditional supervised fine-tuning and reinforcement learning with cycle consistency, optimizing it at both the utterance and conversation levels. Finally, we adopt a diverse profile sampler to capture the distribution of real-world user profiles. Experimental results demonstrate that USP outperforms strong baselines in terms of authenticity and diversity while achieving comparable performance in consistency. Furthermore, dynamic multi-turn evaluations based on USP strongly align with mainstream benchmarks, demonstrating its effectiveness in real-world applications.

1 Introduction

The user simulator is designed as a proxy for real users in interactions with large language models (LLMs). It can simulate a specific user behavior based on the user's characteristics to generate appropriate utterances (Ginsberg, 1991; Song et al., 2021; Moon et al., 2024). Therefore, it has emerged as a promising solution (Liu et al., 2023; Ferreira et al., 2024) for scenarios where it is difficult to obtain real-world human-computer interaction data

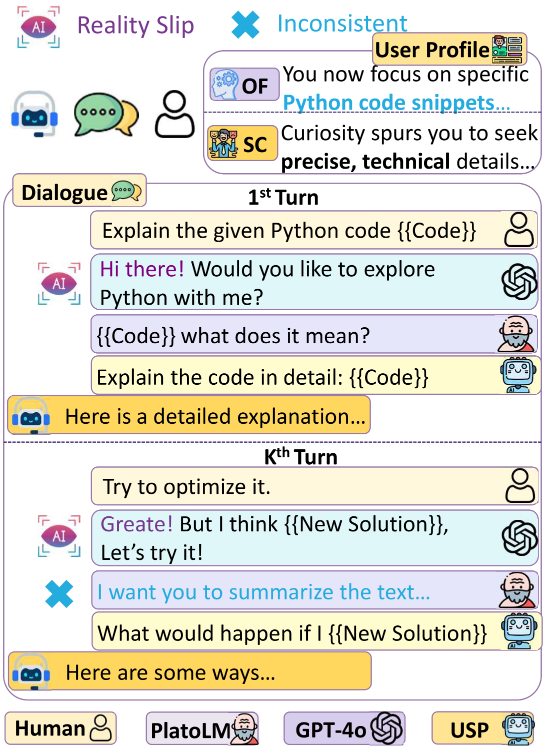


Figure 1: Examples of performance across various user simulators in multi-turn human-LLM interactions. OF and SC refer to objective facts and subjective characteristics, respectively.

due to privacy and ethical issues (such as medical consultation (Valizadeh and Parde, 2022)). It also helps Simulation-to-Reality (Sim2Real) applications, such as tutorial strategies, election simulations, and public opinion research (Liu et al., 2024; Zhang et al., 2024b; Chuang et al., 2024).

Recent Large-scale Language Models (LLMs) advances promote user simulators by enhancing naturalness and utility (Deng et al., 2024; Zhang et al., 2024a), as shown in Figure 1. Since directly using LLM as the user simulator suffers from role confusion (Xu et al., 2023a), some works (Xu et al., 2023a; Kong et al., 2024; Sun et al., 2024) attempt to enhance its authenticity by training a user simulator on the conversation dataset. However, they are only trained on text utterance, making it difficult to simulate diverse user behaviors without seed context, exhibiting limited self-awareness (Tseng et al., 2024) and failing to maintain a consistent personality. Although some LLM-based role-playing methods (Moon et al., 2024) utilize predefined profiles to alleviate this problem, they require additional extensive annotations and can only be applied to celebrities, lacking the diversity of simulation.

To address the issues above, we believe that a user simulator knows users' intrinsic characters hidden in their conversations first and then can be a better simulation. Therefore, we treat user simulation as a dialogue reconstruction task and propose a novel framework named the User Simulator with implicit Profile (USP). It is decomposed into implicit profile extraction to capture the user's underlying characteristics from the target user dialogue and conditional generation based on the profile.

In this framework, we first propose an LLM-driven profile extractor to extract implicit profiles from user conversations with a well-designed profile schema. Inspired by interpersonal interaction theory(Kruglanski and Higgins, 2013), our profile schema contains two dimensions (objective facts (OF) and subjective characteristics (SC)) with a dozen attributes to describe the user comprehensively. Different from existing works (Cheng et al., 2024b; Tu et al., 2024), we then polish the profile attributes into natural, descriptive profiles to ensure generalization.

Then, we integrate the extracted user profiles into the user simulator through two-stage training: (1) conditional supervised fine-tuning with user profiles for utterance-level simulation, and (2) reinforcement learning with cycle consistency to align reflected profiles from simulated dialogues with given profiles for conversation-level simulation. We also implement a diverse profile sampler to capture authentic user distributions.

The experimental results show that our USP significantly outperforms existing baselines in terms of authenticity, consistency, and diversity. We also include a multi-turn dynamic evaluation of LLM with our USP for downstream applications, and the results align well with existing benchmarks and enable a more granular assessment of LLM performance across different user groups. Our key contributions are summarized as follows:

- We propose a novel approach for constructing user simulators with implicit user profiles embedded in human-LLM conversations.
- We develop a new framework that infers implicit user profiles as insight, further enhanced with conditional fine-tuning and reinforcement learning with cycle consistency for better simulation at both the utterance and conversation levels.
- Experimental results show that USP outperforms existing strong baselines in authenticity and diversity, while achieving comparable performance in consistency, and proves effective for multi-turn dynamic evaluation of LLMs.

2 Related Works

2.1 General User Simulator

Early user simulators including agenda-based methods (Schatzmann et al., 2007; Schatzmann and Young, 2009) and model-based methods (Asri et al., 2016; Kreyssig et al., 2018). These simulators were initially designed with a narrow scope due to limited natural language generation capabilities, such as generating synthetic binary preference responses (Christakopoulou et al., 2016) in conversational recommendation systems.

Recent advancements in LLMs enabled more sophisticated simulations of realistic conversations, offering significantly enhanced natural language flexibility. These advances include the use of LLMs for self-chat (Xu et al., 2023b) and dual LLM architectures, where separate models role-play user and assistant based on seed conversations (Ding et al., 2023). Following these innovations, other trained user simulators, such as PlatoLM (Kong et al., 2024) and Parrot (Sun et al., 2024), learn human discourse patterns directly from human-LLM interactions in conversations.

2.2 Persona-based User Simulator

General user simulators often struggle to capture the full spectrum of diverse user needs, leading to a growing interest in persona-based personalization to improve both controllability and diversity in simulations (Takanobu et al., 2020). Some researchers attempt to leverage goal generators (Takanobu et al., 2020) to create diverse user goals or retrieval-based personas derived from historical data (Shi et al., 2019) to guide user simulators in task-oriented dialogue (ToD) systems.

With the rise of LLMs and their impressive zero-shot role-playing abilities (Njifenjou et al., 2024), prompt-driven user simulation has become the dominant approach. For example, LLMs have been used with carefully designed predefined profiles to align with human beliefs (Chuang et al., 2024), simulate consultation scenarios with users exhibiting varying personalities and needs in ToD systems (Zhang et al., 2024a), and model user preferences in conversational recommendation systems (Yoon et al., 2024).

3 Task Definition

We formulate user simulation as a dialogue refactoring task, aiming to replicate multi-turn user behavior in target dialogues. Given a target dialogue $d_i = \{(u_{i1}, r_{i1}), \ldots, (u_{ij}, r_{ij})\}$ between a user U_i and a response model R_i, where u_{ij} and r_{ij} represent the j-th turn user utterance and the corresponding model response, respectively.

To achieve high-fidelity simulation of user responses within a given context, we aim to minimize the utterance-level distance $D_{\text{utt}}(u_{ij}, u_{ij}^{'})$ where $u_{ij}^{'} \sim P_{prob}(\cdot | c_{ij}^{'}, U_i^{'})$. Here, $c_{ij}^{'}$ represents the context used by the user simulator $U_i^{'}$ to generate $u_{ij}^{'}$. This optimization ultimately leads to reduced the dialogue-level distance, as formulated in Eq. 1:

$$\min_{d_i^{'} \sim P_{prob}(\cdot | U_i^{'})} D_{\text{dia}}(d_i, d_i^{'}) \tag{1}$$

where D_{dia} represents a distance function that evaluates the user utterances between the simulated dialogue $d_i^{'}$ and the real conversation d_i.

Recent studies show that role-playing with specific user profiles (P_i) can effectively achieve diverse user simulations (Liu et al., 2023). However, unlike celebrities or well-known characters, user profiles in real-world conversations are often implicit and difficult to obtain (Wang et al., 2024).

To address this, we reformulate the task by first extracting the implicit user profile from the given dialogue using profile extractor $P_{\text{extractor}}$, and then reconstructing a closer dialogue, with an emphasis on the user's utterances, as described in Eq. 2.

$$\min_{d_i^{'} \sim P_{prob}(\cdot | U_i^{'}, P_i)} D_{\text{dia}}(d_i, d_i^{'}), \tag{2}$$

where $P_i = P_{\text{extractor}}(D_i)$.

Category	Dimension	Attributes
Objective Facts	Scene-Consistent Attributes	Age, Gender, Location, Occupation, Education, Family Relationship, Routines/Habits, Social Relationships, Other Experiences
	Scene-Related Attributes	Goals/Plans, Task Details
Subjective Characteristics	Intrinsic Characteristics	Big Five Personality Traits, Language Styles

Table 1: User Profile Schema.

4 Modeling User Simulator with Implicit Profiles

To accomplish this task, we propose the User Simulator with Implicit Profiles (USP) framework, as illustrated in Figure 2, which aims to minimize the objective in Eq. 2 while ensuring authenticity, consistency, and diversity.

4.1 User Profile Construction

4.1.1 User Profile Schema

We believe that the user profile should reveal user characteristics from two aspects: explicit personal information and implicit communication styles. Therefore, inspired by interpersonal interaction theory (Zhou et al., 2024b), we design a user profile schema containing objective facts(OF) and subjective characteristics(SC) to represent them, as shown in Table 1.

The OF focuses on common topics in human conversation (Cheng et al., 2024b; Dunbar et al., 1997) including Scene-Consistent Attributes (such as age, gender, and location) and Scene-Related Attributes (such as goal,and task details). SC considers both external and internal personality dimensions represented by language style (Wang et al., 2024) and the Big-Five Traits Different from previous work (Cheng et al., 2024b; Tu et al., 2024), we reformulate discrete attributes into coherent narrative descriptions to achieve greater generalization and flexibility.

4.1.2 User Profile Extractor

To obtain such a user profile, we design an LLM-driven user profile extractor extracting the implicit user profile from the human-LLM conversation. The extractor first leverages advanced LLM (such as GPT-4o) to extract the user character attributes mentioned above with a well-designed prompt. Then, the extractor collects the valid attributes (No empty) together and polishes them into natural language descriptions. Further prompt details regarding the extractor can be found in Appendix A.2.

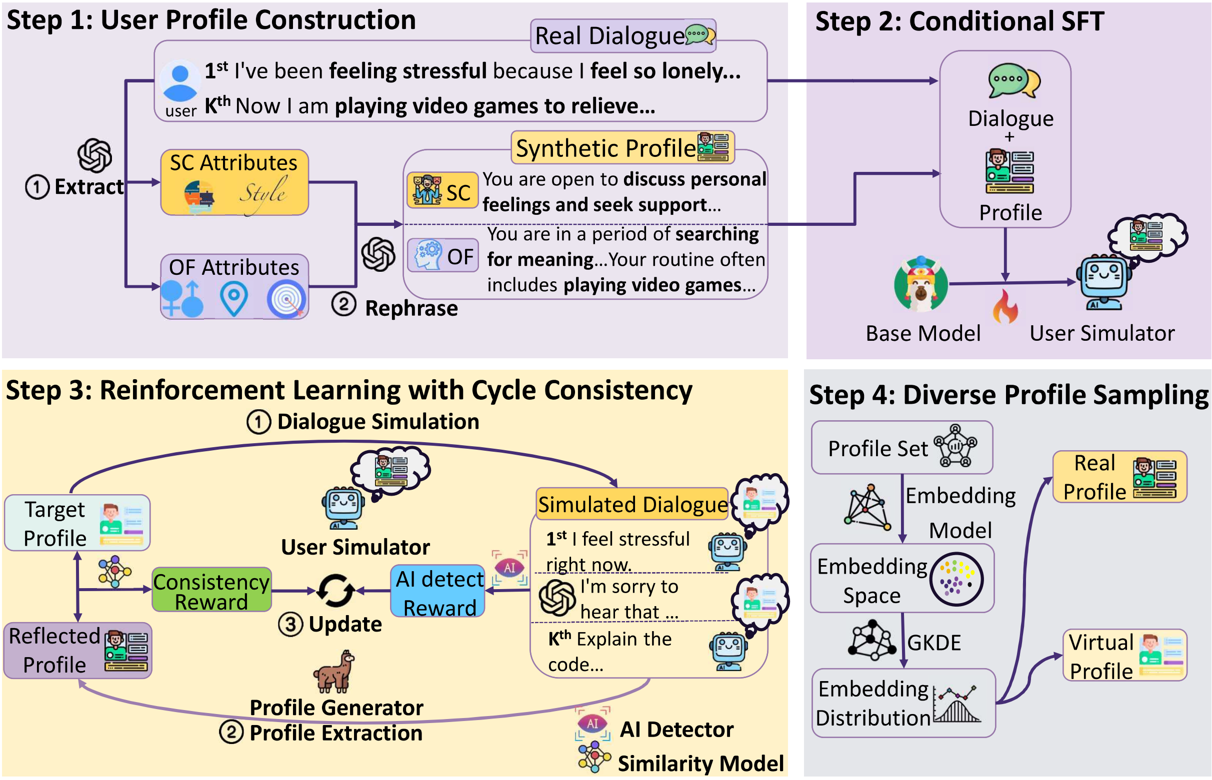


Figure 2: Overview of our proposed User Simulator with implicit Profile(USP) framework.

4.1.3 Profile Quality Verification

Due to the reliance on predefined user profiles in existing role-playing approaches (Zhou et al., 2024b), the correlation between user profiles and conversations has been largely overlooked. To address this limitation, we propose Dialogue Profile Consistency (DPC) for evaluating OF based on atomic fact verification that measures both precision (DP.P) and recall (DP.R) (Jandaghi et al., 2024).

Drawing inspiration from FactScore (Min et al., 2023) and Con.Score (Madotto et al., 2019), we first introduce Factual Consistency (Fact.Con). Given a target T, we evaluate the consistency between the source and target by decomposing T into atomic facts af_k using an atomic fact generator (afg). We then compute the natural language inference (NLI) score for each atomic fact with respect to the source S:

$$\text{Fact.Con}(S, T) = \frac{1}{|af_k|} \sum_{af_k \in \text{afg}(T)} \text{NLI}(S, af_k)$$

We then define $\text{DP.P}_i = \text{Fact.Con}(D_i, P_i)$ and $\text{DP.R}_i = \text{Fact.Con}(P_i, D_i)$, and compute DPC as their harmonic mean. When dialogue D_i serves as the target T, each user utterance u_{ij} is treated directly as an atomic fact af_k. Conversely, when the profile serves as the target T, we utilize afg followed (Min et al., 2023) to decompose it into atomic facts.

Additionally, we use a validation score (Val.Score) to assess the quality of SC descriptions based on the dialogue, rating them on a scale from 1 to 5 using GPT-4o. Detailed prompts in Appendix D.

4.2 Conditional Supervised Fine-Tuning

To empower the LLM with the general capability to simulate diverse users at the utterance level, we utilize conditional supervised fine-tuning based on user profiles. It enables the LLM to learn the conditional generation mapping based on both the extracted profile $P_i^{\prime}$ and context $c_{ij}^{\prime}$. As subtle misalignment between the core objectives of the user simulator and the response model, the SFT language modeling loss is modified as follows:

$$l_{\text{lm}} = \sum_j \sum_k -\log P_{prob}(u'_{i,j,k} | u'_{i,j,<k}, c'_{i,j}, P_i) \quad (3)$$

where u_{ijk} represents the k-th token of the j-th utterance from the i-th user.

4.3 Reinforcement Learning with Cycle Consistency

To further enhance conversation-level consistency, we introduce Reinforcement Learning with Cycle Consistency (RLCC), which optimizes the user simulator by aligning the reflected profile, extracted from simulated dialogues, with the target profile.

In this stage, we extract the simulator profile P'_i from dialogues D' generated by the user simulator U'_i using the profile generator, based on virtual profiles sampled by our diverse profile sampler (see Section 4.4). Our goal is to maximize the semantic similarity between the target profile P_i and the extracted reflected profile both in objective facts and subjective characteristics. The dialogue-level reward is then distributed to each user utterance within the dialogue:

$$r^{cc}_{i,j} = \text{sim}(P_i, P'_i) \quad (4)$$

Then, we optimize profile recall through Proximal Policy Optimization (PPO) (Schulman et al., 2017) using cycle consistency as a dialogue-level reward signal. It enhances the user simulator's self-expression in dialogues, moving beyond simple profile adherence or precision consistency.

Additionally, to prevent deviations or reward hacking, we incorporate an AI detection model as an auxiliary reward. The final reward is given by:

$$r_{i,j} = \lambda r^{cc}_{i,j} + (1 - \lambda) r^{ai\_detect}_{i,j} \quad (5)$$

where $r^{ai\_detect}_{i,j} = \text{AI\_detect}(u_{i,j})$ and $\lambda = 0.8$ is used to emphasize the importance of cycle consistency. The AI detection model (Yang et al., 2024) and profile generator are all fine-tuned based on our train dataset, details in Appendix B.1.

4.4 Diverse Profile Sampler

To generate diverse yet natural user profiles for our tasks, we propose a Density Profile Sampler that maintains distribution characteristics while ensuring coverage of underrepresented cases. It embeds profiles into a semantic space using Sup-SimCSE-RoBERTa (Gao et al., 2021), followed by UMAP projection (McInnes et al., 2018) to preserve density relationships. Finally, we estimate the real profile distribution using Gaussian Kernel Density Estimation (GKDE), which allows us to sample both real and synthetic virtual profiles.

5 Experiments

We evaluate user simulators' authenticity and consistency at both the utterance and conversation levels while assessing diversity through the difference between our simulated and real user distributions.

5.1 Datasets

We select the popular LMSYS-Chat-1M (Zheng et al., 2023) as our data source for the experiment, which contains one million human-LLM conversations. Following previous work (Kong et al., 2024), we filter the samples for non-English language, toxicity, and redundancy and obtain a complete 94,874 samples (87,882/4,626/2,366 for Training/Validation/Test datasets). Then, we use a GPT-4o-based profile extractor described in Section 4.1 to preprocess them with annotating user profiles for each conversation and construct them into LMSYS-USP. Detailed preprocessing are in Appendix A.1.

We used DPC and Val.Score to automatically evaluate the quality of extracted user profiles on the test set of LMSYS-USP, plus Persona-Chat (Zhang et al., 2018) and ConvAI2 [1] (Dinan et al., 2019) with manually annotated dialogue datasets of other existing predefined profiles. Table 2 shows that the extracted profile can achieve over 84% DPC and even the distill-llama3 is close to that of GPT-4o, demonstrating the effectiveness of our annotation. Additionally, we select 100 samples for manual evaluation and it shows over 4/5 scores for the quality of the generated profiles (See Appendix B.4 for further details).

5.2 Baseline Models

(1) User Simulator without User Profile: This includes the untrained DialogueGPT(4o), where GPT-4o relies solely on context to predict the next user utterance, and PlatoLM (Kong et al., 2024), which is fine-tuned on our training dataset using LLaMA-3-8B and can be considered equivalent to our approach without incorporating profile.

(2) User Simulator Guided by the User Profile: We adopt ProfileGPT(4o) and Profi-

[1] We use the human-to-bot dataset in https://huggingface.co/datasets/convai-challenge/conv_ai_2

Dataset	Profile Source	OF					SC
		DP.P	Avg DP.P # Fact	DP.R	Avg DP.R # Fact	DPC	Val.Score
LMSYS-USP	GPT4o	86.89	25.64	82.24	3.71	84.50	4.42
LMSYS-USP	Distill-llama3	86.15	23.81	81.95	3.71	84.00	4.36
Persona Chat	GPT4o	86.21	22.82	62.76	7.86	72.64	4.35
Persona Chat	Human	76.21	8.59	42.94	7.86	54.93	-
ConvAI2	GPT4o	68.71	17.44	39.15	9.97	49.88	3.47
ConvAI2	Human	25.69	8.70	12.64	9.97	16.94	-

Table 2: Automatic evaluation results of profile quality across different datasets.

leGPT(llama), which utilizes GPT-4o and LLAMA-3-8B-Instruct (AI@Meta, 2024) as the role-play backbone with our constructed profiles. Additionally, we include CharacterGLM (Zhou et al., 2024a), which performs role-playing with any given profile, and CharacterLLM (Shao et al., 2023), which role-played on several famous celebrities. The detailed setup and prompts are provided in Appendix B.3.

5.3 Metrics

Authenticity: We use SimCSE (Gao et al., 2021) to compute semantic similarity (Sem-Sim) and style embeddings (Wegmann et al., 2022) to compute style similarity (Style-Sim) for evaluating $D(u_{ij}, u^{'}_{ij})$ and $D(d_i, d^{'}_i)$. We also employ Author Verification Accuracy (AVA) to assess stylistic consistency by measuring whether paired sentences share authorship based on similarity thresholds (Wegmann et al., 2022). For multi-turn evaluation, we compute dialogue-level distances by concatenating each user's utterances.

Consistency: We employ reverse metrics r-DP.P for sentence-level consistency and r-DP.R, r-DPC for dialogue-level consistency. These metrics mirror DP.R, DP.P, and DPC, respectively, but they evaluate consistency from a profile-centric perspective. Additionally, we incorporate Persona Coverage (P.Cover) (Song et al., 2019) to assess keyword-level consistency. The Subjective Characteristic Score (SC.Score), assessed by GPT-4o with prompt in Appendix D, measures the reflection of subjective traits.

Diversity: We compute the Absolute Difference Value (ADV), which represents the Euclidean distance between the PCA-reduced embeddings of generated and target dialogues, to evaluate the discrepancy between the distribution of the reconstructed dialogues and the original dialogues.

Additionally, we assess multi-turn dialogue continuity using the early stop rate (ESR), which flags premature endings caused by repetitive responses or repeated gratitude expressions over three turns.

5.4 Results

5.4.1 Utterance-Level Evaluation

In the utterance-level evaluation, we evaluate the quality of a single-turn response generated by the testing models in a given context.

As shown in Table 3, USP outperforms all baselines in terms of authenticity, with 53.38 and 46.60, as measured by both semantic (Sem-Sim) and stylistic (Style-Sim) similarity metrics. This highlights the effectiveness of our implicit profile-assisted approach for user-LLM dialogue reconstruction, particularly when compared to context-only models like PlatoLM. Although dedicated role-playing models (ProfileGPT variants) achieve higher consistency scores (r-DP.P), this can be attributed to their direct profile text copying. USP maintains comparable overall performance while striking a better balance between authenticity and consistency.

5.4.2 Conversation-Level Evaluation

In the conversation-level evaluation, we assess the performance of testing models to chat with GPT-4o in multi-turns, providing the profile or the first turn of the reference dialogue according to their needs.

As shown in Table 4, USP outperforms baseline models in authenticity, consistency, and continuity. With the lowest ESR(10), it demonstrates superior dialogue continuity. Notably, USP's advantage in authenticity is more pronounced in conversation-level scenarios than sentence-level predictions. In terms of consistency, USP demonstrates exceptional performance in r-DP.R metrics and achieves significantly higher r-DPC scores for overall profile dialogue consistency. This superior performance, particularly when compared to role-playing models such as ProfileGPT(4o) and ProfileGPT(llama) which show high P.Cover scores, suggests that

Model Type	Model	Authenticity			Consistency		
		Sem-Sim↑	Style-Sim↑	AVA↑	r-DP.P↑	P.Cover↑	SC.Score↑
w/o Profile	DialogueGPT(4o)	40.24	13.75	11.28	–	–	–
	PlatoLM	39.37	43.11	40.29	–	–	–
With Profile	Character_LLM	37.54	18.88	15.03	54.77	66.62	2.43
	Character_GLM	38.51	22.28	18.17	68.72	57.72	2.95
	ProfileGPT(llama)	39.82	14.88	13.47	82.19	72.29	3.92
	ProfileGPT(4o)	41.66	5.74	9.87	**92.73**	**73.34**	**4.71**
	USP w/o RLCC	**54.25**	46.57	**43.61**	71.30	71.56	3.36
	USP	53.38	**46.60**	43.35	72.61	71.23	3.39

Table 3: Utterance-level performance comparison of different models on authenticity and consistency metrics.

Model Type	Model	Continuity	Authenticity			Consistency				
		ESR↓	Sem-Sim↑	Style-Sim↑	AVA↑	r-DP.P↑	r-DP.R↑	r-DPC↑	P.Cover↑	SC.Score↑
w/o Profile	DialogueGPT(4o)	35	48.91	14.21	10.58	–	–	–	–	–
	PlatoLM	18	43.24	32.43	31.60	–	–	–	–	–
With Profile	Character_LLM	52	23.37	7.13	4.69	25.48	6.43	10.27	21.49	2.82
	Character_GLM	44	40.19	10.86	12.67	39.51	29.61	33.85	42.75	3.64
	ProfileGPT(llama)	31	46.84	10.58	11.63	67.09	29.98	41.44	47.72	4.19
	ProfileGPT(4o)	32	48.87	10.15	11.26	**76.59**	43.72	55.66	**51.02**	**4.56**
	USP w/o RLCC	12	**66.17**	40.01	35.68	53.17	71.88	61.13	42.63	3.24
	USP	**10**	65.39	**46.23**	**38.77**	56.24	**74.38**	**64.05**	44.08	3.35

Table 4: Conversation-level performance comparison of different models on authenticity and consistency metrics.

RLCC effectively captures abstract profile characteristics rather than merely matching surface-level keywords.

5.4.3 Human Evaluation

We also randomly selected 100 samples with 8 evaluators to conduct conversation-level human evaluations considering authenticity and consistency. Authenticity was assessed through Style, Semantics, and Quality, while consistency covered Accuracy, Completeness, and Quality. Full evaluation details are provided in Appendix B.4.

Table 5 demonstrates USP's superior performance in both authenticity and consistency metrics. Our USP significantly outperforms ProfileGPT (4o) in terms of authenticity (74 vs. 13) and consistency (61 vs. 35) in manual evaluation. USP is superior to PlatoLM trained on the same dataset in terms of authenticity, which demonstrates the usefulness of implicit profile modeling. Thanks to the RLCC module, our USP model has significantly improved consistency (43 vs. 30) by aligning user profiles.

Baseline	Metrics (% USP win/tie/loss)	
	Authenticity (κ=0.548)	Consistency (κ=0.561)
ProfileGPT(4o)	74/13/13	61/4/35
PlatoLM	55/12/33	-
USP w/o RLCC	37/32/31	43/27/30

Table 5: Human evaluation results comparing baselines with USP on authenticity and consistency.

5.4.4 Diversity Sampling Evaluation

Figure 3 shows the absolute difference value between target dialogue and generated dialogue by various models across different percentiles. From the results, we observe that USP and USP w/o RLCC consistently achieve the smallest ADV across all percentiles, indicating the dialogues they generated that closely match the target conversations. For example, marked by the red cross, PlatoLM has 60% of sample ADV below 15%, while USP has only 5% or less ADV. It demonstrates its stronger capability in preserving the semantic characteristics of the original dialogues. The uniformly lower curves of USP and USP w/o RLCC compared to other baselines (PlatoLM, ProfileGPT(4o), and DialogueGPT(4o)) suggest that our approach generates dialogues that are more faithful to the target conversations across typical and extreme cases.

We also analyze and demonstrate that our user simulator can sample different representatives (majority and minority) of users compared to random sampling in Appendix B.5 and apply it to downstream applications in Appendix C.

Model Configuration	Continuity	Authenticity			Consistency				
	ESR↓	Sem-Sim↑	Style-Sim↑	AVA↑	r-DP.P↑	r-DP.R↑	r-DPC↑	P.Cover↑	SC.Score↑
USP w/o RLCC	12	66.17	40.01	35.68	53.17	71.88	61.13	42.63	3.24
USP (5:5)	14	66.28	41.22	37.03	52.23	71.59	60.39	43.58	**3.55**
USP (8:2)	**10**	65.39	**46.23**	**38.77**	56.24	**74.38**	**64.05**	44.08	3.35
USP (9:1)	12	**66.91**	38.87	33.62	**58.36**	70.62	63.90	**46.75**	3.33

Table 6: Ablation study of hyperparameters in RLCC.

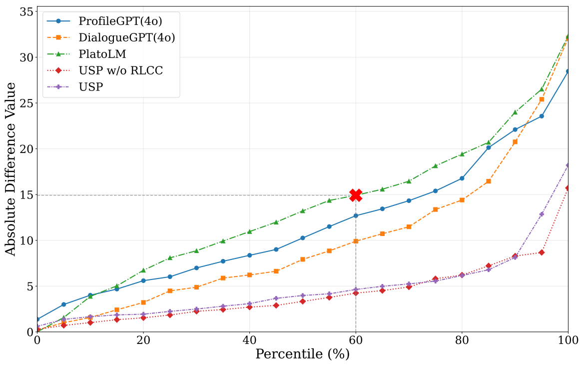


Figure 3: Cumulative distribution of ADV performance comparison across different models.

Model Setup	Sampling Strategy			Avg.	Ranking in LiveBench/ Chatbot-Arena
	Major	Minor	Virtual		
Deepseek-v3	**8.25**	6.13	**7.70**	**7.36**	1
GPT-4o	7.86	**6.65**	7.19	7.23	3
Claude-Sonnet	7.18	6.61	7.48	7.09	2
4o-Mini	6.84	5.70	5.52	6.02	4
Claude-Haiku	4.88	5.42	5.43	5.24	5

Table 7: Response model performance comparison over different target groups.

6 Analysis

6.1 Ablation Study

To evaluate the relative importance of RLCC's two rewards, we tested different values of λ in Equation 5, denoted as USP(λ: $1-\lambda$).

As shown in Table 6, $\lambda = 0.8$ provides the optimal balance between maintaining model capabilities and enhancing dialogue consistency. Higher values of λ (0.9) compromise speaking style authenticity without improving r-DPC, leading to superficial profile matching, as shown by increased P.Cover scores. Conversely, $\lambda = 0.5$ achieves authentic style features but lacks sufficient consistency emphasis, resulting in stagnant performance across capabilities.

6.2 Applications: Dynamic Multi-turn Evaluation For LLMs

One application of our simulator is to fill the gap in the current dynamic multi-turn evaluation of large models. The user simulator can simulate different user groups to dynamically interact with the tested model in multiple rounds and reveal their specific defects, as shown in Table 7.

We simulated 300 diverse user profiles by USP, comprising 100 highest-probability profiles(the majority), 100 lowest-probability profiles(the minority) based on estimated density, and 100 random synthetic profiles(the virtuality), using the sampler mentioned in Section 4.4. Then, we leverage USP using these profiles to chat with LLM in multi-turns and evaluate them, followed by MT-Bench (Zheng et al., 2024). The experimental results indicate that our user simulator is effective and consistent with the average rankings on the latest Livebench (White et al., 2024) and Chatbot-Arena (Chiang et al., 2024). In addition, it can also be seen that GPT-4o is better at catering to minority groups than other models, indicating its superior robustness. A more detailed analysis can be found in the appendix C.

7 Conclusion

In this work, we propose a novel user simulator with implicit profiles that excels in authenticity, consistency, and diversity. Based on this, we introduce the USP framework, which integrates extracted user profiles into the user simulator by conditional fine-tuning and reinforcement learning with cycle consistency. Our experimental results, validated by both automatic metrics and human evaluations, show that USP significantly outperforms role-playing simulators (e.g., GPT-4o) and direct simulation approaches (e.g., PlatoLM) in authenticity while achieving comparable consistency at both the sentence and conversation levels. Furthermore, through a dynamic evaluation across various LLMs chatting with diverse demographic groups, we demonstrate USP's effectiveness in real-world applications.

Limitations

We acknowledge the following limitations: 1) Applicability Across Different Scenarios: We conduct the experiments on a single dataset, and there has been limited validation across multiple datasets to assess the generalizability of the results. 2) Cultural and Linguistic Scope: We focus on English dialogues in this paper, which may limit the applicability of USP to other linguistic and cultural contexts.

Ethics Statement

Although LMSYS-1M has undergone extensive data cleaning and ethical checks, the dataset may still contain sensitive or harmful content, reflecting violent, explicit, or discriminatory traits in certain dialogue. This could result in USP generating unsafe dialogues. We strongly advise against including sensitive terms in profiles when using USP, as this may lead to extreme behavior in both the USP and the response model.

References

AI@Meta. 2024. Llama 3 model card.

Layla El Asri, Jing He, and Kaheer Suleman. 2016. A sequence-to-sequence model for user simulation in spoken dialogue systems. *arXiv preprint arXiv:1607.00070*.

Iz Beltagy, Matthew E Peters, and Arman Cohan. 2020. Longformer: The long-document transformer. *arXiv preprint arXiv:2004.05150*.

Chuanqi Cheng, Quan Tu, Wei Wu, Shuo Shang, Cunli Mao, Zhengtao Yu, and Rui Yan. 2024a. "in-dialogues we learn": Towards personalized dialogue without pre-defined profiles through in-dialogue learning. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 10408–10422, Miami, Florida, USA. Association for Computational Linguistics.

Yi Cheng, Wenge Liu, Kaishuai Xu, Wenjun Hou, Yi Ouyang, Chak Tou Leong, Xian Wu, and Yefeng Zheng. 2024b. Evolving to be your soulmate: Personalized dialogue agents with dynamically adapted personas. *CoRR*, abs/2406.13960.

Zihao Cheng, Li Zhou, Feng Jiang, Benyou Wang, and Haizhou Li. Beyond binary: Towards fine-grained llm-generated text detection via role recognition and involvement measurement. In *THE WEB CONFERENCE 2025*.

Wei-Lin Chiang, Lianmin Zheng, Ying Sheng, Anastasios Nikolas Angelopoulos, Tianle Li, Dacheng Li, Hao Zhang, Banghua Zhu, Michael Jordan, Joseph E. Gonzalez, and Ion Stoica. 2024. Chatbot arena: An open platform for evaluating llms by human preference. *Preprint*, arXiv:2403.04132.

Konstantina Christakopoulou, Filip Radlinski, and Katja Hofmann. 2016. Towards conversational recommender systems. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 815–824.

Yun-Shiuan Chuang, Krirk Nirunwiroj, Zach Studdiford, Agam Goyal, Vincent Frigo, Sijia Yang, Dhavan Shah, Junjie Hu, and Timothy T. Rogers. 2024. Beyond demographics: Aligning role-playing llm-based agents using human belief networks. In *Findings of the Association for Computational Linguistics: EMNLP 2024, Miami, Florida, USA, November 12-16, 2024*, pages 14010–14026. Association for Computational Linguistics.

Yang Deng, Wenxuan Zhang, Wai Lam, See-Kiong Ng, and Tat-Seng Chua. 2024. Plug-and-play policy planner for large language model powered dialogue agents. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net.

Emily Dinan, Varvara Logacheva, Valentin Malykh, Alexander H. Miller, Kurt Shuster, Jack Urbanek, Douwe Kiela, Arthur Szlam, Iulian Serban, Ryan Lowe, Shrimai Prabhumoye, Alan W. Black, Alexander I. Rudnicky, Jason Williams, Joelle Pineau, Mikhail S. Burtsev, and Jason Weston. 2019. The second conversational intelligence challenge (convai2). *CoRR*, abs/1902.00098.

Ning Ding, Yulin Chen, Bokai Xu, Yujia Qin, Shengding Hu, Zhiyuan Liu, Maosong Sun, and Bowen Zhou. 2023. Enhancing chat language models by scaling high-quality instructional conversations. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, EMNLP 2023, Singapore, December 6-10, 2023*, pages 3029–3051. Association for Computational Linguistics.

Robin IM Dunbar, Anna Marriott, and Neil DC Duncan. 1997. Human conversational behavior. *Human nature*, 8:231–246.

Rafael Ferreira, David Semedo, and João Magalhães. 2024. Multi-trait user simulation with adaptive decoding for conversational task assistants. In *Findings of the Association for Computational Linguistics: EMNLP 2024, Miami, Florida, USA, November 12-16, 2024*, pages 16105–16130. Association for Computational Linguistics.

Tianyu Gao, Xingcheng Yao, and Danqi Chen. 2021. SimCSE: Simple contrastive learning of sentence embeddings. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6894–6910, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Matthew L. Ginsberg. 1991. Marvin minsky, the society of mind. *Artif. Intell.*, 48(3):335–339.

Samuel D Gosling, Peter J Rentfrow, and William B Swann Jr. 2003. A very brief measure of the big-five personality domains. *Journal of Research in personality*, 37(6):504–528.

Pegah Jandaghi, XiangHai Sheng, Xinyi Bai, Jay Pujara, and Hakim Sidahmed. 2024. Faithful persona-based conversational dataset generation with large language models. In *Findings of the Association for Computational Linguistics, ACL 2024, Bangkok, Thailand and virtual meeting, August 11-16, 2024*, pages 15245–15270. Association for Computational Linguistics.

Chuyi Kong, Yaxin Fan, Xiang Wan, Feng Jiang, and Benyou Wang. 2024. Platolm: Teaching llms in multi-round dialogue via a user simulator. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2024, Bangkok, Thailand, August 11-16, 2024*, pages 7841–7863. Association for Computational Linguistics.

Florian Kreyssig, Iñigo Casanueva, Pawel Budzianowski, and Milica Gasic. 2018. Neural user simulation for corpus-based policy optimisation for spoken dialogue systems. *arXiv preprint arXiv:1805.06966*.

Arie W Kruglanski and E Tory Higgins. 2013. *Social psychology: Handbook of basic principles*. Guilford Publications.

Wai-Chung Kwan, Xingshan Zeng, Yuxin Jiang, Yufei Wang, Liangyou Li, Lifeng Shang, Xin Jiang, Qun Liu, and Kam-Fai Wong. 2024. MT-eval: A multi-turn capabilities evaluation benchmark for large language models. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 20153–20177, Miami, Florida, USA. Association for Computational Linguistics.

Yajiao Liu, Xin Jiang, Yichun Yin, Yasheng Wang, Fei Mi, Qun Liu, Xiang Wan, and Benyou Wang. 2023. One cannot stand for everyone! leveraging multiple user simulators to train task-oriented dialogue systems. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2023, Toronto, Canada, July 9-14, 2023*, pages 1–21. Association for Computational Linguistics.

Zhengyuan Liu, Stella Xin Yin, Geyu Lin, and Nancy Chen. 2024. Personality-aware student simulation for conversational intelligent tutoring systems. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, EMNLP 2024, Miami, FL, USA, November 12-16, 2024*, pages 626–642. Association for Computational Linguistics.

Andrea Madotto, Zhaojiang Lin, Chien-Sheng Wu, and Pascale Fung. 2019. Personalizing dialogue agents via meta-learning. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5454–5459, Florence, Italy. Association for Computational Linguistics.

Adyasha Maharana, Dong-Ho Lee, Sergey Tulyakov, Mohit Bansal, Francesco Barbieri, and Yuwei Fang. 2024. Evaluating very long-term conversational memory of llm agents. *arXiv preprint arXiv:2402.17753*.

Leland McInnes, John Healy, and James Melville. 2018. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

Sewon Min, Kalpesh Krishna, Xinxi Lyu, Mike Lewis, Wen-tau Yih, Pang Wei Koh, Mohit Iyyer, Luke Zettlemoyer, and Hannaneh Hajishirzi. 2023. Factscore: Fine-grained atomic evaluation of factual precision in long form text generation. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, EMNLP 2023, Singapore, December 6-10, 2023*, pages 12076–12100. Association for Computational Linguistics.

Suhong Moon, Marwa Abdulhai, Minwoo Kang, Joseph Suh, Widyadewi Soedarmadji, Eran Kohen Behar, and David M. Chan. 2024. Virtual personas for language models via an anthology of backstories. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, EMNLP 2024, Miami, FL, USA, November 12-16, 2024*, pages 19864–19897. Association for Computational Linguistics.

Ahmed Njifenjou, Virgile Sucal, Bassam Jabaian, and Fabrice Lefèvre. 2024. Role-play zero-shot prompting with large language models for open-domain human-machine conversation. *arXiv preprint arXiv:2406.18460*.

Alex Rodriguez and Alessandro Laio. 2014. Clustering by fast search and find of density peaks. *Science*, 344(6191):1492–1496.

Jost Schatzmann, Blaise Thomson, Karl Weilhammer, Hui Ye, and Steve Young. 2007. Agenda-based user simulation for bootstrapping a pomdp dialogue system. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Companion Volume, Short Papers*, pages 149–152.

Jost Schatzmann and Steve Young. 2009. The hidden agenda user simulation model. *IEEE transactions on audio, speech, and language processing*, 17(4):733–747.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

Yunfan Shao, Linyang Li, Junqi Dai, and Xipeng Qiu. 2023. Character-LLM: A trainable agent for role-playing. In *Proceedings of the 2023 Conference on*

Empirical Methods in Natural Language Processing, pages 13153–13187, Singapore. Association for Computational Linguistics.

Weiyan Shi, Kun Qian, Xuewei Wang, and Zhou Yu. 2019. How to build user simulators to train rl-based dialog systems. *arXiv preprint arXiv:1909.01388*.

Haoyu Song, Yan Wang, Kaiyan Zhang, Wei-Nan Zhang, and Ting Liu. 2021. Bob: BERT over BERT for training persona-based dialogue models from limited personalized data. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021*, pages 167–177. Association for Computational Linguistics.

Haoyu Song, Weinan Zhang, Yiming Cui, Dong Wang, and Ting Liu. 2019. Exploiting persona information for diverse generation of conversational responses. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pages 5190–5196. ijcai.org.

Yuchong Sun, Che Liu, Kun Zhou, Jinwen Huang, Ruihua Song, Xin Zhao, Fuzheng Zhang, Di Zhang, and Kun Gai. 2024. Parrot: Enhancing multi-turn instruction following for large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2024, Bangkok, Thailand, August 11-16, 2024*, pages 9729–9750. Association for Computational Linguistics.

Ryuichi Takanobu, Runze Liang, and Minlie Huang. 2020. Multi-agent task-oriented dialog policy learning with role-aware reward decomposition. *arXiv preprint arXiv:2004.03809*.

Yu-Min Tseng, Yu-Chao Huang, Teng-Yun Hsiao, Wei-Lin Chen, Chao-Wei Huang, Yu Meng, and Yun-Nung Chen. 2024. Two tales of persona in llms: A survey of role-playing and personalization. In *Findings of the Association for Computational Linguistics: EMNLP 2024, Miami, Florida, USA, November 12-16, 2024*, pages 16612–16631. Association for Computational Linguistics.

Quan Tu, Shilong Fan, Zihang Tian, Tianhao Shen, Shuo Shang, Xin Gao, and Rui Yan. 2024. CharacterEval: A Chinese benchmark for role-playing conversational agent evaluation. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11836–11850, Bangkok, Thailand. Association for Computational Linguistics.

Mina Valizadeh and Natalie Parde. 2022. The AI doctor is in: A survey of task-oriented dialogue systems for healthcare applications. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022*, pages 6638–6660. Association for Computational Linguistics.

Noah Wang, Zhongyuan Peng, Haoran Que, Jiaheng Liu, Wangchunshu Zhou, Yuhan Wu, Hongcheng Guo, Ruitong Gan, Zehao Ni, Jian Yang, Man Zhang, Zhaoxiang Zhang, Wanli Ouyang, Ke Xu, Wenhao Huang, Jie Fu, and Junran Peng. 2024. Rolellm: Benchmarking, eliciting, and enhancing role-playing abilities of large language models. In *Findings of the Association for Computational Linguistics, ACL 2024, Bangkok, Thailand and virtual meeting, August 11-16, 2024*, pages 14743–14777. Association for Computational Linguistics.

Tongzhou Wang and Phillip Isola. 2020. Understanding contrastive representation learning through alignment and uniformity on the hypersphere. In *International Conference on Machine Learning*, pages 9929–9939. PMLR.

Anna Wegmann, Marijn Schraagen, and Dong Nguyen. 2022. Same author or just same topic? towards content-independent style representations. In *Proceedings of the 7th Workshop on Representation Learning for NLP*, pages 249–268, Dublin, Ireland. Association for Computational Linguistics.

Colin White, Samuel Dooley, Manley Roberts, Arka Pal, Ben Feuer, Siddhartha Jain, Ravid Shwartz-Ziv, Neel Jain, Khalid Saifullah, Siddartha Naidu, Chinmay Hegde, Yann LeCun, Tom Goldstein, Willie Neiswanger, and Micah Goldblum. 2024. Livebench: A challenging, contamination-free llm benchmark.

Canwen Xu, Daya Guo, Nan Duan, and Julian McAuley. 2023a. Baize: An open-source chat model with parameter-efficient tuning on self-chat data. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 6268–6278.

Canwen Xu, Daya Guo, Nan Duan, and Julian J. McAuley. 2023b. Baize: An open-source chat model with parameter-efficient tuning on self-chat data. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, EMNLP 2023, Singapore, December 6-10, 2023*, pages 6268–6278. Association for Computational Linguistics.

Lingyi Yang, Feng Jiang, Haizhou Li, et al. 2024. Is chatgpt involved in texts? measure the polish ratio to detect chatgpt-generated text. *APSIPA Transactions on Signal and Information Processing*, 13(2).

Se-eun Yoon, Zhankui He, Jessica Maria Echterhoff, and Julian McAuley. 2024. Evaluating large language models as generative user simulators for conversational recommendation. *arXiv preprint arXiv:2403.09738*.

Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason Weston. 2018. Personalizing dialogue agents: I have a dog, do you have pets too? In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics,* *ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*, pages 2204–2213. Association for Computational Linguistics.

Tong Zhang, Chen Huang, Yang Deng, Hongru Liang, Jia Liu, Zujie Wen, Wenqiang Lei, and Tat-Seng Chua. 2024a. Strength lies in differences! improving strategy planning for non-collaborative dialogues via diversified user simulation. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, EMNLP 2024, Miami, FL, USA, November 12-16, 2024*, pages 424–444. Association for Computational Linguistics.

Xinnong Zhang, Jiayu Lin, Libo Sun, Weihong Qi, Yihang Yang, Yue Chen, Hanjia Lyu, Xinyi Mou, Siming Chen, Jiebo Luo, et al. 2024b. Electionsim: Massive population election simulation powered by large language model driven agents. *arXiv preprint arXiv:2410.20746*.

Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Tianle Li, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zhuohan Li, Zi Lin, Eric P Xing, et al. 2023. Lmsys-chat-1m: A large-scale real-world llm conversation dataset. *arXiv preprint arXiv:2309.11998*.

Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric Xing, et al. 2024. Judging llm-as-a-judge with mt-bench and chatbot arena. *Advances in Neural Information Processing Systems*, 36.

Jinfeng Zhou, Zhuang Chen, Dazhen Wan, Bosi Wen, Yi Song, Jifan Yu, Yongkang Huang, Pei Ke, Guanqun Bi, Libiao Peng, JiaMing Yang, Xiyao Xiao, Sahand Sabour, Xiaohan Zhang, Wenjing Hou, Yijia Zhang, Yuxiao Dong, Hongning Wang, Jie Tang, and Minlie Huang. 2024a. CharacterGLM: Customizing social characters with large language models. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track*, pages 1457–1476, Miami, Florida, US. Association for Computational Linguistics.

Jinfeng Zhou, Zhuang Chen, Dazhen Wan, Bosi Wen, Yi Song, Jifan Yu, Yongkang Huang, Pei Ke, Guanqun Bi, Libiao Peng, et al. 2024b. Characterglm: Customizing social characters with large language models. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track*, pages 1457–1476.

A Dataset Construction

A.1 Preprocessing

Our dataset preprocessing follows the method outlined in PlatoLM (Kong et al., 2024), which includes the removal of non-English content, filtering of toxic data, elimination of exact duplicates at the dialogue level, and segmentation of conversations into maximum-length token sequences. To maintain discourse integrity, truncated dialogues are ensured to start with the speaker's turn, preserving context consistency and coherence.

A.2 Profile Dataset

As discussed in Section 4.1, we categorize all attributes into three types: scene-consistent attributes, scene-related attributes, and deep intrinsic characteristics. For scene-consistent attributes, we use the prompt shown in Figure 11, with each metric definition following the guidelines outlined in (Cheng et al., 2024b). For scene-related attributes, we use the prompt in Figure 12, and for deep intrinsic characteristics, we refer to Figure 10. The definition of the Big Five Traits scores follows (Gosling et al., 2003).

Next, we concatenate the attributes, remove invalid values (e.g., null or meaningless values), and shuffle the order to eliminate any positional bias in the generated profiles. The attributes, which encompass the three aforementioned aspects, are then rephrased using GPT-4o, with the prompt shown in Figure 13. As a result, we obtain automatically labeled profiles for each data entry. The length statistics are shown in Table 8.

Dataset	Train	Val	Test	Profile
LMSYS-USP	1,149	1,295	1,438	231

Table 8: Average token length of LMSYS-USP dataset.

Furthermore, we calculated the frequency of occurrence for each attribute value (i.e., the average number of different attribute values per sample) to assess the prevalence of each attribute. The statistics for the objective facts can be found in Figure 4, while for the subjective characteristics, we focused on whether the Big Five Traits were significantly exhibited. Specifically, we only consider traits with high or low scores, while moderate scores are viewed as the average representation of human behavior (Moon et al., 2024) and are not included in the subsequent profiles.

A.3 Resource Consumption in Implementation

During the data construction process, each attribute extraction requires approximately \$0.003 using the GPT API. Since each sample requires three extractions, the cost per sample is approximately \$0.01. With a total of around 94,000 samples, the cost for attribute extraction amounts to approximately \$940.

Attribute	High Rate (%)	Low Rate (%)
Conscientiousness	78.07	7.53
Agreeableness	6.45	14.98
Extraversion	4.08	14.15
Openness	58.77	5.30
Neuroticism	2.04	10.12

Table 9: Summary of extracted subjective attribute statistics.

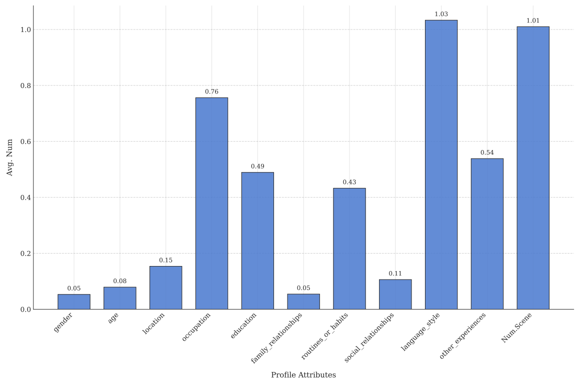


Figure 4: Frequency of occurrence of values across different attributes of objective facts in the attribute extraction process.

Additionally, rewriting the attributes into profiles incurs a cost of around $0.05 per sample. Therefore, the total cost for constructing the dataset is approximately $1,400.

B Implement Detail

B.1 Trainable Model Setup

We use the `llama3-base` model as the base architecture. Directly training an instruct model with user simulation contradicts the inherent task requirements during its SFT phase, making convergence difficult. We first perform conditional SFT on the training dataset, followed by our diverse profile sampler, which randomly selects 1,000 samples from the training set for virtual user sampling. Specifically, we combine objective facts and subjective descriptions from different profiles and generate approximately 1 million profiles. From these, we select the 5,000 profiles with the lowest similarity to the training dataset and use them for RLCC phrase. The conditional SFT is conducted using four A100 40GB GPUs for full fine-tuning with epoch set to 3, taking about two days. The RLCC phase is trained using two H20 96GB GPUs over the course of five days.

For the PlatoLM, we also use `llama3-base` as the base architecture. The system prompt used is: “`A chat between a curious human and an artificial intelligence assistant. The human can ask further questions based on previous conversations, or he can directly ask brand new questions without any context from prior conversations.`” We fine-tune the model using four A100 40GB GPUs with epoch set to 3, which takes approximately two days.

For the AI detection model, we follow (Cheng et al.) and use Longformer (Beltagy et al., 2020) to train on our dataset. Since our dataset naturally distinguishes between AI and human-generated text, we label user utterances as human and assistant utterances as AI. We trained for 3 epochs using dual 3090 GPUs, taking three days to complete

For the profile generator model, we use LLaMA3-Instruct as the backbone and train it on a curated profile dataset. We distilled the profile generation capabilities of GPT-4 into a two-stage process. Training was conducted on four A100 40GB GPUs for 3 epochs, taking two days to complete.

B.2 Baseline Model Setup

Table 10 presents the experimental setup for the baseline models. For models based on GPT, we use the corresponding APIs with default settings for inference. For other models, the experiments are conducted on a single NVIDIA RTX A100 GPU with a batch size of 2 and a repetition penalty of 1.0. An exception is made for CharacterGLM, where we utilize its chat function with a repetition penalty of 1.6 and set the number of beams to 3.

B.3 Seed conversation design

As there are two types of simulators, one being a response model based on role-playing, which cannot proactively initiate conversations, we first embed the corresponding profile into the system prompt. Then, by using the query "What will you say to start the conversation?" we guide the model to simulate the user’s input over the test dataset. Finally, the simulated user utterance is passed to the response model for interactive generation.

B.4 Human Evaluation

B.4.1 Profile Evaluation

We employed two annotators to rate the extracted profiles on a scale of 1 to 5 based on the given dialogues, assessing the accuracy and completeness

Model Name	Backbone	System Prompt
USP	LLAMA3-8B	You are engaging in a conversation with an AI assistant. your profile is: `{profile}`. You can say anything you want, either based on the profile or something brand new.
DialogueGPT(4o)	gpt-4o-2024-08-06	-
PlatoLM	LLAMA3-8B	A chat between a curious human and an artificial intelligence assistant. The human can ask further questions based on previous conversations, or he can directly ask brand new questions without any context from prior conversations.
ProfileGPT(4o)	gpt-4o-2024-08-06	You are engaging in a conversation with an AI assistant. your profile is: `{profile}`. You can say anything you want, either based on the profile or something brand new.
CharacterGLM	ChatGLM-6B	以下是一段User和AI assistant之间的对话。 关于User的信息：`{profile}` 关于AI assistant的信息：GPT-4o
Character-LLM-Socrates-7b	LLAMA-7B	I want you to act like the person described in the profile below: `{profile}`. I want you to respond and answer like the person, using the tone, manner, and vocabulary the person would use. You must know all of the knowledge of the person. The status of you is as follows: Status: You are engaging in a conversation with an AI assistant. The interactions are as follows:

Table 10: Baseline model setup.

of objective facts, reasonableness of subjective descriptions, and overall naturalness. For accuracy, we emphasized precision without hallucinations; for completeness, we focused on the absence of significant omissions; and for reasonableness, we prioritized rational, well-justified descriptions, avoiding exaggeration or bias. The results in Table 11 show that annotators achieved moderate to high agreement on profile assessment.

Dataset	Profile Source	Objective Facts	Subjective characters	Naturalness
lmsys	GPT-4o	4.64	4.19	4.66

Table 11: Human evaluation results for the quality of constructed profiles.

B.4.2 Dialogue Evaluation

We recruited eight annotators to assess conversation-level results based on two key criteria: authenticity and consistency. For authenticity, annotators were tasked with identifying user utterances within a dialogue that most closely resemble those in a reference dialogue, evaluating them across three dimensions: Style, Semantics, and Quality. For consistency, annotators were provided only with the user profile and asked to select the user utterances that best align with the profile, considering Accuracy, Completeness, and Quality. Definitions for all six evaluation dimensions are provided in Table 12, with detailed guidelines in Section D.

To minimize position bias and the potential influence of prior exposure, each annotator was randomly assigned pairs of dialogues, with the order of options also randomized. The evaluation interface used for this process is shown in 5.

B.5 Sampling strategy effectiveness

To evaluate the effectiveness of our density sampler approach, we propose two complementary metrics that directly measure both local structure preservation and global coverage. For local structure preservation, we employ Local Density Loss (Rodriguez and Laio, 2014) (LDL), where a lower value indicates tighter local clustering, allowing us to verify whether our sampler preserves the natural local structures of the profile distribution. For global coverage assessment, we utilize Uniformity Loss (Wang and Isola, 2020), where a lower value indicates better global uniformity, ensuring our sampling achieves comprehensive coverage of

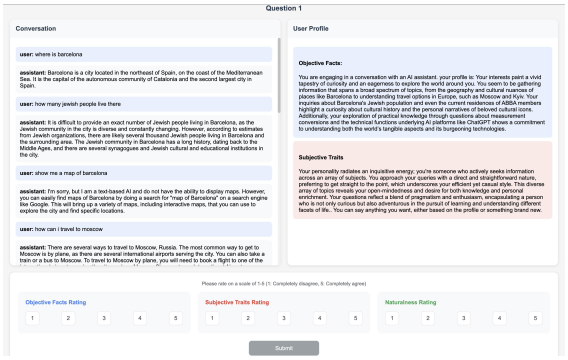


Figure 5: The web-based interface used for profile evaluation.

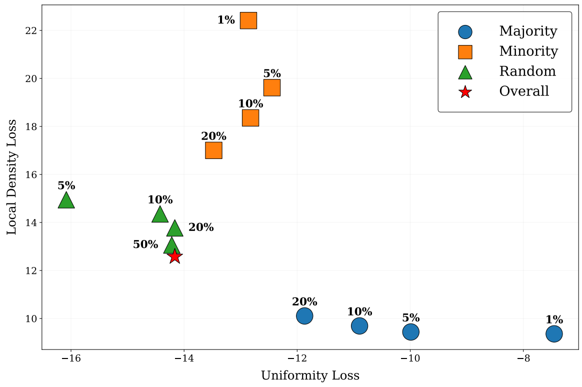


Figure 6: Distribution of different sampling strategies.

Aspect	Dimension	Description
Authenticity	Style	Whether the wording and tone align with the target (Cheng et al., 2024a).
	Semantic	Evaluating the thematic consistency and semantic content.
	Quality	Fluency, coherence, and human-likeness of the user utterance (Cheng et al., 2024b).
Consistency	Accuracy	Whether the utterance accurately reflects the persona information of the target (Cheng et al., 2024a).
	Completeness	Whether the utterance comprehensively represents the described persona.
	Quality	Fluency, coherence, and human-likeness of the user utterance (Cheng et al., 2024b).

Table 12: Evaluation dimensions for authenticity and consistency assessment.

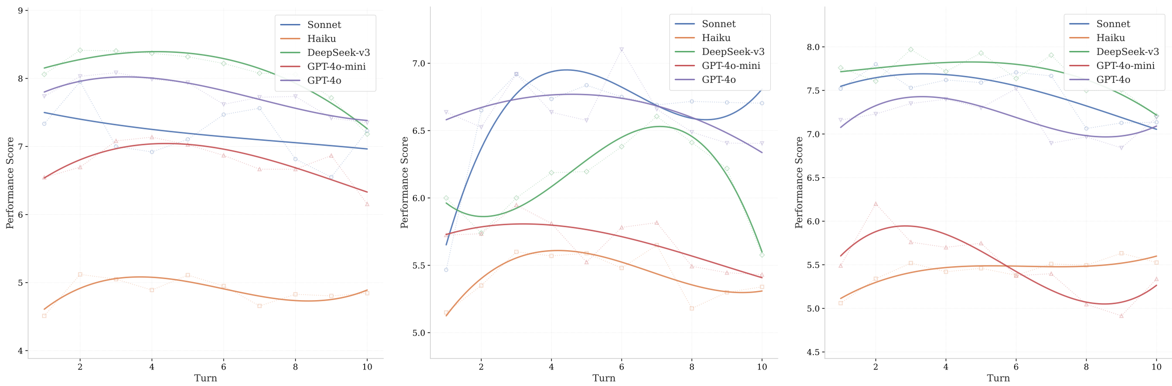
the profile space while maintaining realistic distributions.

Using the GKDE density distribution as our guide, we implement two targeted sampling strategies: sampling from high-density regions to capture majority patterns, and weighting towards low-density regions to ensure coverage of minority cases. As shown in Figure 6, this approach effectively maintains distribution balance while achieving these targeted objectives. From right to left on the uniformity loss axis reveals a clear transition from majority to minority sampling patterns, with majority samples (blue circles) showing optimal performance in the low local density loss region and high uniformity, indicating effective preservation of key distribution characteristics. The minority samples (orange squares) strategically occupy higher local density regions with moderate uniformity loss, demonstrating our method's ability to capture diverse distribution patterns. The steady progression of sampling percentages in both majority and minority cases shows stable and controlled sampling behavior, while random sampling (green triangles) exhibits more scattered patterns, validating our approach's reliability. The overall performance marker (red star) positioned at the intersection of these patterns confirms our method's success in balancing between distribution preservation and targeted sampling objectives.

C Downstream analysis

Our results reveal four key findings, as illustrated in Figure 7: (1) LLMs demonstrate strong performance with the majority demographic but exhibit reduced effectiveness with minority groups, highlighting the limitations in personalization; (2) Despite this, reasonable performance with virtual groups suggests robust generalization capabilities; (3) As shown in Figure 7, instruction-following ability deteriorates in extended dialogues, particularly for minority groups, a trend consistent with previous studies (Kwan et al., 2024; Maharana et al., 2024); (4) Significant performance fluctuations across dialogue turns for minority groups emphasize the need for improved capabilities in handling and responding to less common instructions.

D Prompt Templates



(a) Variation of different models with Semantic Similarity across turns.

(b) Variation of different models with Style-Similarity scores across turns.

(c) Variation of different models with AVA scores across turns.

Figure 7: Trends of Semantic, Style Similarity, and AVA across turns.

Case Study on Authenticity of USP Simulated Dialogue

Simulated Dialogue:
USP: hello who am i talking to?
assistant: Hello! You're speaking with ChatGPT...

USP: are you vicuna?
assistant: No, I'm not Vicuna. I'm ChatGPT, created by OpenAI...

USP: what do you know about it?
assistant: Vicuna is an AI language model developed by a different research group and is aimed at generating human-like text...

USP: can you tell me how i could train and deploy my own ai model?
assistant: Training and deploying your own AI model involves several key steps, from data preparation to deployment...

Reference Dialogue:
User: how to train a model like you
assistant: I am a large language model, which means that I was trained on a vast amount of text data....

User: what is your base model
assistant: I am a large language model called GPT....

User: isn't you vicuna
assistant: I am not aware of any information or research about a model called "vicuna". I am a large language model called GPT....

User: what hardware do i need to run you
assistant: I am a large language model called GPT, which was trained on a vast amount of text data and requires significant computational resources to run.....

Figure 8: Case study of USP simulated dialogue based solely on the provided profile, interacting with GPT-4o. The reference dialogue below serves as the target for reconstruction.

Case Study on Consistency of USP Simulated Dialogue

Profile:
You are engaging in a conversation with an AI assistant. your profile is: You have a passion for engaging in activities that challenge your vocabulary and word formation skills. One of your favorite pastimes is diving into puzzle games that offer a thrilling test of language prowess. You often find yourself thinking of five-letter words with tricky placements, such as words containing 'A' in the fourth position, or concocting ones with 'R' in the second spot and 'T' at the end. This aligns with your appreciation for linguistic puzzles and an enduring enjoyment of wordplay, reflecting your cognitive curiosity and dedication to enhancing your language skills.

Your personality shines through a playful and engaging language style, particularly evident when you use structured questions to create dynamic game scenarios. You approach life with a vivid imagination, delighting in intellectual challenges that require intricate problem-solving. Your attention to detail is noteworthy, and you thrive in situations where organization is key. Preferring a more reserved lifestyle, you cherish moments of solitude and introspection, embracing activities that allow you to delve into complex, solitary intellectual pursuits with enthusiasm and precision.

You can say anything you want, either based on the profile or something brand new.

Simulated Dialogue:
USP: Let's play a game about 5 letter words. I'll ask you some questions and try to guess the answer. Can we start?
assistant: Sure, I'd be happy to play a game with you! Let's get started!

USP: Question: What is a 5-letter word which has "A" in its 4th place? Answer:
assistant: Is the word you are thinking of "Polar"?

USP: Nope What about "gnaw"?
assistant: "GNW" is a four-letter word. Let's think of another five-letter word with "A" in the fourth position...

USP: Next Question: Give me one example of a 6-letter word having R at second place and T at sixth place.
assistant: Sure! A six-letter word with "R" in the second position and "T" at the sixth position could be "Report".

USP: Great! Now give me one example of a 7-letter word having S at third place and H at seventh place.
assistant: Sure! One example of a seven-letter word with "S" in the third place and "H" in the seventh place is "English."

Figure 9: Case study of USP simulated dialogue based solely on the provided profile, interacting with GPT-4o.

Prompt for Extracting Big Five Personality Traits

Task:
Analyze the provided dialogue to assess the user's personality traits across 5 personality dimensions Focus exclusively on the user's characteristics, disregarding any information related to others individuals, unless it directly impacts the user.
For each personality trait:

1. Assign each dimension independently with **score**: "High", "Low," or "Inconclusive."
2. Provide **conclusion**: A high-level description, with concise supporting details.
3. Provide **reason**: Justify the assigned score with specific evidence from the dialogue.
4. Mark traits as **Inconclusive** only when no clear evidence exists.

Personality Trait Defination:
`{{metric}}: {{definition}}`
Format:

```
{
    "Trait Name": {
        "score": "High/Low/Inconclusive",
        "conclusion": "The user is a [trait descriptor] person...",
        "reason": "Explanation referencing specific dialogue evidence."
    },
    ...
}
```

Example:

```
[User]: "She is my age, in a homeless women's shelter, living under very poor
conditions. She is a mental health client, but the treatment team seems to
... Her background is similar to mine, and I cannot abandon her."
```

Detected Personality Traits:

```
{
  "Conscientiousness": {
    "score": "High",
    "conclusion": "The user is a conscientious person who demonstrates a
    sense of duty and commitment.",
    "reason": "The user expresses a strong sense of responsibility ..."
  },
  "Agreeableness": {
    "score": "High",
    "conclusion": "The user is an empathetic and compassionate person who
    values relationships.",
   "reason": "The user shows care and concern for their cousin's well-being..."
  },
  "Extraversion": {
    "score": "Inconclusive"
  },
  ...
}
```

Figure 10: Prompt for extracting deep intrinsic characteristics.

Prompt for Extracting Scene-Consistent Attributes

User Persona Analysis Task
Objective
The primary goal of this task is to analyze user utterances in-depth and accurately extract key persona attributes based on both direct and implicit cues. These attributes should be categorized into distinct fields, with any missing or unclear details left blank.
Field Descriptions: `{{Field}}: {{definition}}`
Guidelines

1. Carefully examine each user utterance to extract relevant persona traits. Consider both direct statements and implicit clues.
2. Ensure that the extracted attributes are specific and directly relevant to the user's utterances. Avoid vague or generalized descriptions unless explicitly supported by the text.
3. Pay attention to distinctive communication styles (e.g., formal or casual tone, frequent use of specific words or phrases) to capture the user's unique way of communicating.

Example
User Utterances:

```
[User]: Given an array of integers nums and an integer target, return indices of
the two numbers such that they add up to target in Python...
[Assitant]: To solve the problem...
[User]: Thanks
```

Expected Output:

```
{
    "gender": [],
    "age": [],
    "location": [],
    "occupation": [
        "Likely a beginner programmer or student studying computer science,
        as evidenced by the simple coding problem in Python."
    ],
    "education": [
        "Possibly a student in computer science or a related field,
        at an early stage in learning programming, specifically Python."
    ],
    "family_relationships": [],
    "routines_or_habits": [],
    "social_relationships": [],
    "language_style": [
        "Concise and task-oriented",
        "Polite response 'Thanks' after getting satisfactory answer"
    ],
    "other_experiences": []
}
```

Figure 11: Prompt for extracting scene-consistent attributes.

Prompt for Extracting Scene-Related Attributes

User Persona Analysis Task

Objective

The goal of this task is to analyze multi-turn user utterances within a conversation with an assistant and extract key elements such as the primary goals and specific task descriptions. Each extracted detail should be as specific as possible, reflecting the user's context, objectives, and scenario.

Output Format

The extraction should be presented in a structured JSON format, as shown below:

```
{
    "scenarios": [
        {
            "goals_or_plans": "<List of User's goals or plans>",
           "task_details": "<List of specific tasks summary made by the user>"
        },
        ...
    ]
}
```

Field Descriptions:

- **goals_or_plans**: User's short-term or long-term objectives, derived from explicit statements or inferred from the overall conversation. If no explicit goals are stated, infer them from the main topics discussed.
- **task_details**: Specific tasks, actions, or requests made by the user. Each task should be concisely summarized with specifics. If there are multiple tasks or requests, list each one separately.

Example

User Utterances:

```
[User]: Summarize: Harry Potter is a fictional character in the Harry Potter series...
[Assitant]: Harry Potter is a fictional character...
[User]: Write an email inquiring about coursework...
```

```
{
    "scenarios": [
        {
          "goals_or_plans": "Aiming to gain a deeper understanding of the Harry
                 Potter series, possibly for academic or personal enrichment.",
            "task_details": [
           "Summarizing introductory content about the Harry Potter character."
            ]
        },
        {
        "goals_or_plans": "Looking to improve professional communication skills.",
            "task_details": [
                "Writing an email to inquire about coursework."
            ]
        }
    ]
}
```

Figure 12: Prompt for extracting scene-related attributes.

Prompt for Rephrasing Attributes into Natural Descriptions

Narrative Generation Objective
Rephrase the provided key-value pairs into a natural, coherent second-person description.
Core Requirements

1. **Perspective**: Use second-person perspective ("you", "your").
2. **Structure**: Two paragraphs:
   - First paragraph: Present objective facts.
   - Second paragraph: Describe subjective characteristics.
3. **Key Principles**
   - Accurately represent **all** provided details.
   - Ensure the language flows naturally, remains engaging, and avoids redundancy.
   - Focus on clear and seamless transitions between ideas.

Output Expectations

- **Objective Facts**:
  - Convert the key-value pairs into a clear and natural description without over-explaining or adding unnecessary details.
  - Ensure each scenario is logically connected and key information is presented effectively.
- **Subjective Characteristics**:
  - Avoid vague terms like "high perfectionism" or "moderate emotional stability." Use vivid, descriptive language to bring these traits to life.

Figure 13: Prompt for rephrasing attributes into natural descriptions for profile generation.

Human Evaluation Guidelines for Authenticity

1. **Task Description:**
Please choose which user in the two test conversations is more similar to the reference conversation being spoken by the same person.
2. **Evaluation Criteria:**

- **Semantic Similarity**: Measure the thematic consistency and discourse coherence between the generated user utterance and the target user utterance. Preference should be given to the utterance that more accurately reflects the semantic content of the target.
- **Stylistic Parity**: Analyze whether the generated user utterance matches the style of the target user utterance, including its tone, vocabulary, and grammatical structure. The utterance that aligns more closely with the stylistic elements of the target should be favored.
- **Quality**: Examine the fluency and logical coherence of the user utterance, focusing on the linguistic and logical smoothness of the user utterance. The more coherent and fluent utterances should be chosen.

Figure 14: Human evaluation guidelines for authenticity.

Human Evaluation Guidelines for Consistency

1. **Task Description:**
Select the dialogue that contains the most appropriate **user utterance** from the two given generated dialogues based on the target user profile.
2. **Evaluation Criteria:**

- **Persona Reflection**: Evaluate how closely each user utterance reflects the target profile's thematic content, tone, and stylistic features. Preference should be given to the dialogue whose user utterance most accurately reflects the profile's characteristics in terms of thematic coherence and expression.
- **Comprehensiveness**: Assess the extent to which the user utterance encapsulates the target profile, integrating both objective facts and subjective descriptions. The more comprehensive utterance, which addresses a broader range of profile aspects, should be favored over one that focuses narrowly on a single dimension.
- **Quality**: Evaluate the fluency, coherence, and human-likeness of the user utterance. Preference should be given to the utterance that demonstrates greater linguistic smoothness, logical cohesion, and alignment with genuine human conversational patterns.

Figure 15: Human evaluation guidelines for consistency.

Prompt for NLI Score of Profile Precision Based on Given Dialogue

Role
You are an expert in evaluating the **consistency** between a given **user profile (Source)** and **the user's utterance (Target)**. Your task is to assess whether the **Target** aligns with, contradicts, or is ambiguous in relation to the **Source**.
Task Instructions:
For each **Source-Target** pair, determine the relationship using the following scoring criteria:

- **Score 1**: The Target is consistent with the Source (the interpretation can be inferred from the Source).
- **Score -1**: The Target conflicts with the Source (the interpretation contradicts the Source).
- **Score 0**: The relationship is unclear or ambiguous (insufficient evidence to infer consistency or contradiction).

Output Format:
Provide your result in the following JSON format:

```
{
  "score": <score>,
  "reason": "<concise explanation of the reasoning>"
}
```

Example:
Source: You are interested in dataset-related details.
Target: [User]: Show me how to implement a toy version of a relational database.
Output:

```
{
  "score": 1,
  "reason": "The request for implementing a relational database suggests an
  interest in data structures and datasets, which aligns with the Source."
}
```

Guidelines:

1. **Contextual Inference**: Do not evaluate the Target **in isolation**. Instead, determine its logical relationship to the Source, considering both explicit statements and reasonable implications.
2. **Concise & Precise Justification**: The reasoning should be clear, objective, and free from unnecessary elaboration.

Figure 16: Prompt for DP.P based on NLI.

Prompt for NLI Score of Dialogue Precision Based on Given Profile

Role
You are an expert in evaluating consistency between **a given dialogue history (Source)** and a corresponding **user profile description (Target)**. Your task is to determine whether the Target aligns with, contradicts, or is ambiguous in relation to the Source.
Task Instructions:
For each **Source-Target** pair, determine the relationship using the following scoring criteria:

- **Score 1**: The Target is consistent with the Source (the interpretation can be inferred from the Source).
- **Score -1**: The Target conflicts with the Source (the interpretation contradicts the Source).
- **Score 0**: The relationship is unclear or ambiguous (insufficient evidence to infer consistency or contradiction).

Output Format:
Provide your result in the following JSON format:

```
{
  "score": <score>,
  "reason": "<concise explanation of the reasoning>"
}
```

Example :
Source:

User (Turn-1): Show me how to implement a toy version of a relational database.

User (Turn-2): Thanks a lot!

Target: You are polite.
Output:

```
{
  "score": 1,
  "reason": "The user's expression of gratitude in Turn-2 indicates politeness,
  which aligns with the Target."
}
```

Guidelines:

1. **Contextual Inference**: Do not evaluate the Target **in isolation**. Instead, determine its logical relationship to the Source, considering both explicit statements and reasonable implications.
2. **Concise & Precise Justification**: The reasoning should be clear, objective, and free from unnecessary elaboration.

Figure 17: Prompt for DP.R based on NLI.

Prompt for Subjective Characteristics Score (SC.Score) in Consistency Evaluation

Task Description
You are tasked with evaluating the quality of user responses in real human-LLM interactions. Specifically, you will assess the degree to which a given response (Target) aligns with a predefined personality profile, tone, and linguistic characteristics (Source). Additionally, you must consider the naturalness and authenticity of the Target, ensuring it reflects genuine human conversational patterns.
Evaluation Criteria
Your assessment will focus on two primary dimensions:

1. **Human-Likeness**: The extent to which Target exhibits natural human language, characterized by appropriate syntax, tone, and conversational flow.
2. **Alignment with Source**: The degree to which the Target adheres to the personality traits, tone, and linguistic features specified in the Source.

Scoring Guidelines
Assign a score from 1 to 5 based on the following criteria:

- **Score 5**: The Target fully aligns with the Source and demonstrates exceptional human-likeness.
- **Score 3**: The relationship between the Target and Source is ambiguous or unclear, lacking sufficient evidence for alignment or contradiction.
- **Score 1**: The Target significantly deviates from the Source or lacks human-likeness, rendering it unnatural or inconsistent.

Output Requirements
Provide your evaluation in the following JSON format:

```
{
  "score": <score>,
  "reason": "<concise reason>"
}
```

Key Considerations

1. **Contextual Inference**: Analyze both explicit content and implicit nuances in the Target to determine its alignment with the Source.
2. **Conciseness and Precision**: Ensure that your reasoning is clear, objective, and free of superfluous elaboration.
3. **Human-Likeness Emphasis**: A lack of human-likeness, even if the Target aligns with the Source, will result in a lower score.

Figure 18: Prompt for evaluating consistency in subjective characteristics.

Prompt for Validation Score (Val.Score) in Assessing the Quality of Subjective Characteristics in Profiles

Role
As an expert in evaluating the **consistency** between **user utterances in a dialogue (Source)** and a provided **subjective description (Target)**, your task is to assess whether the **Target** accurately reflects the characteristics described in the **Source**, including personality traits, tone, and other relevant attributes. You will then rate this consistency on a scale from 1 to 5.

Task Instructions
For each pair of **Source-Target**, apply the following scoring criteria to determine their relationship:

- **Score 5**: The **Target** completely aligns with the **Source**, with no discrepancies. The profile perfectly represents the characteristics observed in the user's utterance.
- **Score 3**: Ambiguity or insufficient evidence exists, making it difficult to ascertain the relationship definitively.
- **Score 1**: A clear discrepancy exists, with the **Target** significantly deviating from the **Source**, indicating a mismatch in the represented characteristics.

Output Format
Your assessment should adhere to the following structured JSON format:

```
{
  "score": "<numerical score>",
 "reason": "<a succinct explanation providing justification for assigned score>"
}
```

Guidelines:

1. **Contextual Inference**: Determine the target's logical relationship to the Source, considering both explicit statements and reasonable implications.
2. **Concise & Precise Justification**: The reasoning should be clear, objective, and free from unnecessary elaboration.

Figure 19: Prompt for validation score (Val.Score) in assessing the quality of subjective characteristics in profiles.