D^2PCM:A Multi-Turn Dialogue Dataset with Personalized Contextual Memory

Zhe Yang [1], Yi Huang [1,2] *, Yaqin Chen [1], Chunyang Gao [1], Jingyu Yao [1], Junlan Feng [1]
[1] JIUTIAN Research; [2] Department of Computer Science and Technology, Tsinghua University, China
{yangzhe,huangyi,chenyaqin,gaochunyang,yaojingyu,fengjunlan}@cmjt.chinamobile.com

Abstract

Memory serves as a pivotal component in interactive response generation, supplying essential background information and referential knowledge for dialogues. Conventional interactive algorithms have predominantly treated memory as a merely contextual element, largely neglecting the nuanced cognitive processes involved in individualized memory encoding and retrieval. This conceptual gap has led to the prevailing schema where memory-enhanced dialogue datasets incorporate monolithic, undifferentiated memory content, failing to capture the personalized nature of persoa memory processing. Grounded in the self-reference effect from cognitive psychology, we introduce a Multi-Turn **D**ialogue **D**ataset with **P**ersonalized **C**ontextual **M**emory ($\mathbf{D^2PCM}$), establishing a comprehensive benchmark to facilitate advanced research on personalized memory processing algorithms.

1 Introduction

The utilization of memory (Hatalis et al., 2024; Zhong et al., 2024; Xu et al., 2025) serves as a central pillar for enhancing response generation, providing a foundational context that bolsters the credibility of the outputs and elevates user satisfaction. Conventional Retrieval-Augmented Generation (RAG) based approaches (Lewis et al., 2020; Gao et al., 2023; Asai et al., 2024; Wang et al., 2025c) typically select relevant memory content by performing similarity searches within a global memory pool against the user query, with the selection being operationalized at the turn level (Yuan et al., 2024), the session level (Li et al., 2025a), or the segment level (Pan et al., 2025). Correspondingly, such datasets generally require only a single information repository that aggregates all user-related memories, exhibiting no significant variation across individuals. This poses significant challenges for investigating the underlying mechanisms of personalization in memory processes (Du et al., 2025).

The cognitive mechanisms underlying memory processing are divergent across individuals. In psychology, the **Self-Reference Effect** delineates inherent characteristics of human memory processing (Rogers et al., 1977; Wiesmann et al., 2025). When information is self-relevant, individuals engage a richer network of personal experiences, emotions, and prior knowledge to assimilate it, thereby generating more numerous and robust mnemonic cues. Consequently, memory content that aligns more closely with an individual's personal attributes tends to be maintained in a state of heightened activation, making it more likely to be selected as context in subsequent cognitive or behavioral processes. Building upon this phenomenon, we design a Multi-Turn Dialogue Dataset with Personalized Contextual Memory, abbreviated as $\mathbf{D^2PCM}$. At each turn of interaction, the memory chunk is utilized to store distinct types of recollections. Each item within the chunk exhibits variability in persona expression, with one specific item aligning closely with the user's characteristics, thereby increasing its likelihood of being selected by the user. In summary, our contributions can be outlined as follows:

- Guided by the self-reference effect, we design a novel memory mechanism that realistically simulates human memory processes. Memory contents are categorized based on persona, ensuring that only those aligned with user characteristics are more likely to be retrieved.
- To facilitate comprehensive validation across various post-training algorithms, we provide multiple candidate responses with reward values for each interaction turn. This enriches the dataset's representational diversity and supports evaluation under reinforcement learning (RL) based algorithms.
- We introduce a suite of evaluation metrics for

*Corresponding Author: Yi Huang

Findings of the Association for Computational Linguistics: ACL 2026, pages 37512–37529

the dataset assessment, including LLM-judgement-based win-rate analysis (Gu et al., 2024) and reward-driven response quality evaluation, thereby establishing a multi-perspective benchmark for algorithmic verification and analysis.

2 D^{2}PCM Collection

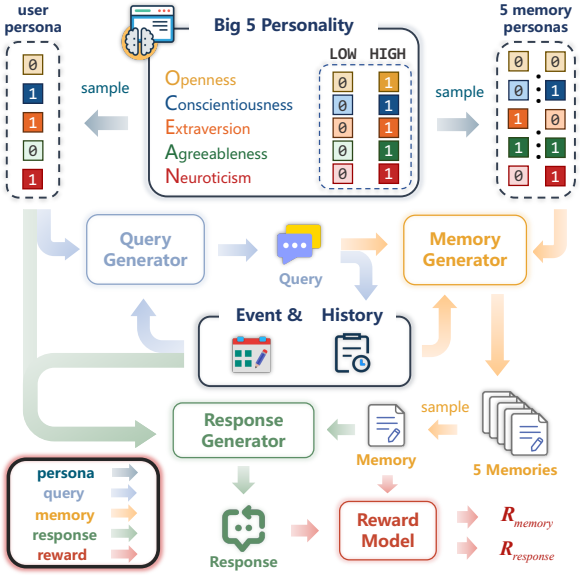


Figure 1: Overview of the Data Generation Procedure. A dialogue session commences with the initialization of both **User Persona** and **Memory Personas**. Subsequently, a complete interaction cycle comprises the following stages: **Query Generation** → **Memory Generation and Sampling** → **Response Generation** → **Reward Derivation**.

While query relevance establishes the foundational criterion for memory retrieval, we posit that congruence with the user's persona acts as a decisive factor for preferential user endorsement. This specialized class of memory, which integrates personal contextual alignment, is formally defined as **Personalized Contextual Memory**. Accordingly, within our structured schema, a complete dialogue data pertinent to personalized memory comprises four mandatory fields: an **event** (elucidating the dialogue's topic), a **persona** (encompassing user-specific profile information), a **memory personalities set** (curating the persona information manifested within memories), and a **dialogue** (the conversational body). Furthermore, each turn within the dialogue exhibits a granular structure, sequentially consisting of: a user **query**, a **memory chunk** (wherein each constituent memory item is categorized as either the Personalized Contextual Memory or not), a **chosen memory** (the specific memory item selected by the assistant from the chunk to facilitate response generation), and the assistant's **response** (generated based on the dialogue history and the chosen memory).

2.1 Personality Construction

The fidelity of memory and stylistic nuances of the dialogue are fundamentally contingent upon the operationalization of the user's personality. Anchored in the **Big Five Personality Theory** (Ackerman, 2020; Wang et al., 2025b), we model personality along five axes, i.e., *Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism*, each quantized dichotomously as high or low. This quantization allows for the delineation of 32 foundational archetypes.

Name	Description	Type
X	32 meta-personalities set	set
χ	the element of X	text
Ω	descriptions for X	set
$\mathcal{C}$	keywords set	set
c	the element of $\mathcal{C}$	text
ε	the event	text
ω_ε	the element of Ω for ε	text
$\hat{\omega}_\varepsilon^i$	perturbation of ω_ε	text
$\mathcal{M}_\varepsilon$	memory personas set for ε	set
m_ε	the element of $\mathcal{M}_\varepsilon$	text
q_t	user query at turn t	text
M_t	memory chunks set at turn t	set
m_t	chosen memory at turn t	text
u_t	assistant response at turn t	text
h_t	history for turn t	text
r_t	reward value for turn t	vector

Table 1: Notations for the data collection procedure.

Subsequently, we leverage the LLM to transmute these typological constructs into coherent textual descriptors, thereby constituting a structured personality pool:

$$\Omega = \{\mathbf{LLM}(\mathcal{P}_\Omega(\chi)) \mid \chi \in X\}, \quad (1)$$

where Ω denotes the personality pool, $\mathcal{P}_\Omega$ means the devised prompt for textual personality genera-

tion. χ represents the archetype from the set of 32 personality constructs.

The First Principle for Data Construction: Based on the aforementioned personality definitions, we formulate the foundational logic governing memory selection and response generation within the data construction pipeline.

• The memory item demonstrating stylistic and attitudinal congruence with the user's personality profile receives elevated probability of being selected for response generation.

• The response must exhibit consistent personality alignment through its stylistic and attitudinal features to optimize user satisfaction metrics.

Name	Description	Input
$\mathcal{P}_\Omega$	textual personalities generation	χ
$\mathcal{P}_\varepsilon$	event generation	c
$\mathcal{P}_{IQG}$	initial query generation	$\varepsilon, \omega_\varepsilon$
$\mathcal{P}_{QG}$	query generation	$\varepsilon, \omega_\varepsilon, h_t$
$\mathcal{P}_{MG}$	memory generation	$\varepsilon, h_t, q_t, m_\varepsilon$
$\mathcal{P}_{UG}$	response generation	$\varepsilon, \omega_\varepsilon, h_t, q_t, m_t$
$\mathcal{P}_{UE}$	response evaluation	u_t

Table 2: Prompt design for data collection.

2.2 Basic Information

Beyond user personality, each data entry should also include an event to specify the topic discussed in the dialogue, along with a memory personality to constrain the stylistic attributes of the resultant memory.

We begin by identifying a set of core thematic keywords—such as education, athletics, gastronomy, and interpersonal relationships—which serve as conceptual anchors. Subsequently, the LLM is employed to extrapolate from these words, generating a coherent narrative event involving two interlocutors. This constructed event then functions as the foundational theme for the ensuing dialogue.

$$\varepsilon = \mathbf{LLM}(\mathcal{P}_\varepsilon(c)) \ \textbf{s.t.}\ c \in \mathcal{C}, \tag{2}$$

where ε denotes the generative event, $\mathcal{P}_\varepsilon$ the event-related prompt, and $\mathcal{C}$ the set of topic keywords.

Grounded in a cognitive architecture of memory and personality, memory content is stipulated to emulate authentic stylistic traits. From the perspective of the user, associated memories may manifest in diverse cognitive styles. Adhering to our first principle in Section 2.1, memory content exhibiting high stylistic congruence with the user's persona possesses a greater probability of eliciting positive engagement when utilized for response generation. To model the heterogeneity of authentic memory styles and to investigate the impact of personalized memory retrieval, we instantiate each event with n ($n = 5$ in the paper) distinct memory personalities. One profile is identical to the user's persona, while the remainder are systematically modulated variants. This modulation is operationalized within the Big Five personality trait by randomly selecting specific dimensions of the five axes and perturbing their intensity values (e.g., transitioning a trait from high to low). Consequently, the set of memory personality for the current event, i.e., $\mathcal{M}_\varepsilon$, can be formally defined as:

$$\mathcal{M}_\varepsilon = \{\omega_\varepsilon\} \cup \{\hat{\omega_\varepsilon^i} \mid i \in [1, n-1]\}, \tag{3}$$

where $\omega_\varepsilon \in \Omega$ means the attached user personality for the event. $\hat{\omega_\varepsilon^i}$ is the $i-th$ perturbation item of the user persona.

2.3 Dialogue Generation

The multi-turn dialogue is orchestrated through a sequential generation chain (Figure 1), where each turn is rigorously driven by the triad of: the user's query, the pertinent user memories, and the assistant response.

Query: Each user query must be contextually coherent with both the ongoing event and the dialogue history, while also exhibiting stylistic alignment with the user's persona. To distinguish the initial query from subsequent turns, we employ two discrete, specialized prompt templates to guide the LLM, ensuring targeted and contextually appropriate generation for each stage.

$$q_t = \begin{cases} \mathbf{LLM}(\mathcal{P}_{IQG}(\varepsilon, \omega_\varepsilon)) & t = 1, \\ \mathbf{LLM}(\mathcal{P}_{QG}(\varepsilon, \omega_\varepsilon, h_t)) & t > 1, \end{cases} \tag{4}$$

where $\mathcal{P}_{IQG}$ and $\mathcal{P}_{QG}$ denote the prompts for generating the inaugural and follow-up queries, respectively, and h_t represents the dialogue history.

Memory: Diverging from the conventional paradigm of a singular, monolithic memory repository, we conceptualize the memory as a set of mnemonic chunks. The cardinality of these chunks

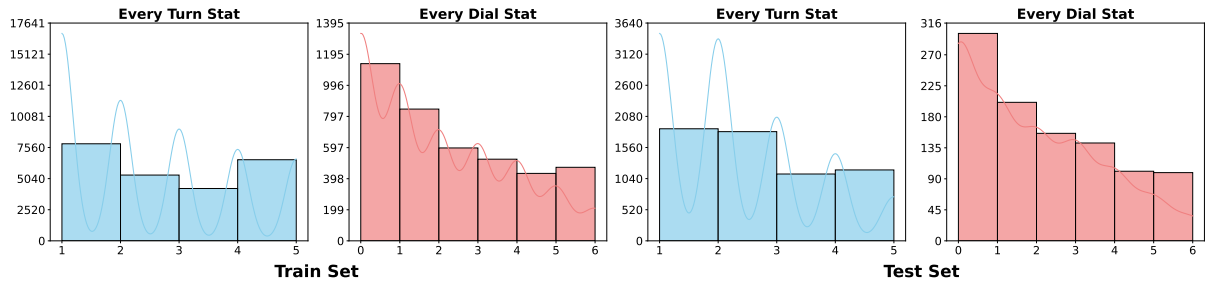


Figure 2: The alignment between the **most relevant memory** selected by the LLM and the **persona memory**. Blue bars represent turn-level statistics, where a value of 1 indicates consistency with the persona memory; Red bars denote dialogue-level statistics, with values representing the number of turns matched to the persona memory within each dialogue.

corresponds directly to the number of memory personalities, i.e., $\mathcal{M}_\varepsilon$ in Equation 3, with each chunk acting as a dedicated cache for its assigned personality's content. The fundamental characteristic for content in any chunk is its fulfillment of multiple semantic constraints: verisimilitude with the event narrative, congruity with the dialogue history, and paramountly, direct materiality to the immediate user's query. **A case study illustrating personalized memory content is found in Section E.**

$$M_t{=}\{\mathbf{LLM}(\mathcal{P}_{MG}(\varepsilon, h_t, q_t, m_\varepsilon))|m_\varepsilon{\in}\mathcal{M}_\varepsilon\}, \quad (5)$$

where $\mathcal{P}_{MG}$ denotes the memory generation prompt. Critically, the content across memory chunks is designed to be divergent. This diversity is not only guaranteed by the distinct memory personalities but is also actively enforced through our prompt, which is designed to solicit a spectrum of distinct viewpoints or commentaries from each chunk pertaining to the current query, thereby emulating the pluralistic feedback inherent in real-world user experiences for an event.

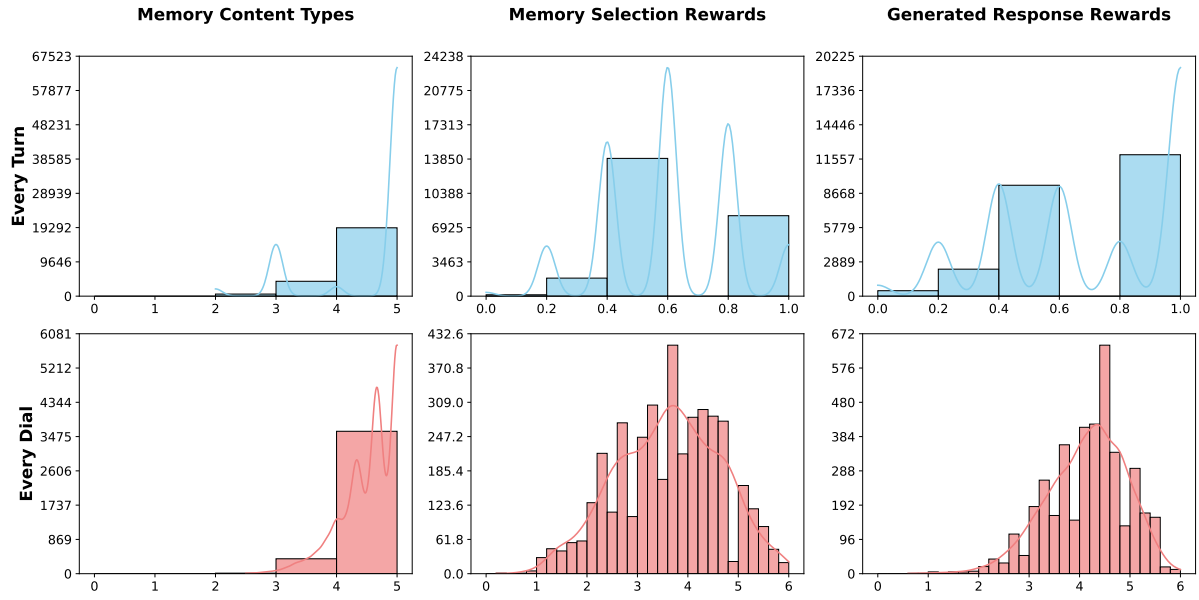
Response: The response generation process is orchestrated by first probabilistically sampling a chunk from aforementioned memory chunks. This chunk's content serves as the key contextual input to the LLM for synthesizing the assistant's reply. The generated response is consequently constrained by the necessity to maintain narrative consistency with the event and historical dialogue, provide a resolution to the user query, and optimally reflect the stylistic nuances of the user persona.

$$\begin{gathered} u_t = \mathbf{LLM}(\mathcal{P}_{UG}(\varepsilon, \omega_\varepsilon, h_t, q_t, m_t)), \\ y_t = \mathbf{LLM}(\mathcal{P}_{UE}(u_t)), \\ \textbf{s.t.} \quad m_t = M_t.sample(p = p_\varepsilon), \\ p_\varepsilon = \{0.4\} \cup \{p_i{=}\frac{0.6}{n{-}1} \mid i \in [1, n{-}1]\}, \end{gathered} \quad (6)$$

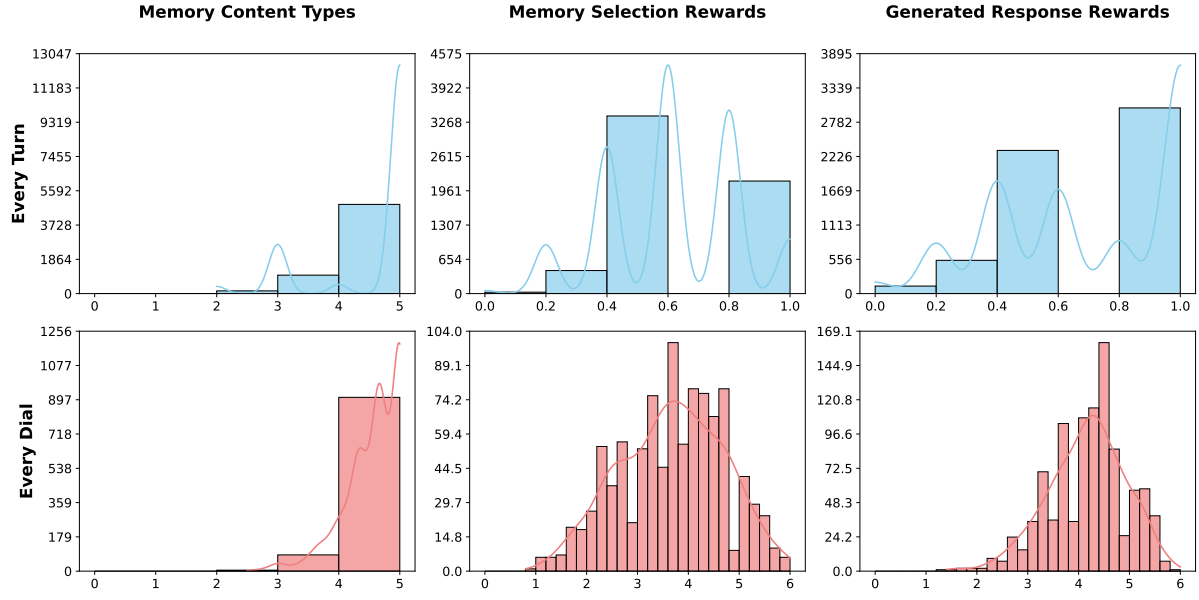
where $m_t \in M_t$ is the sampled memory for current dialogue turn. The sampling process is governed by a pre-configured probability distribution p_ε. Notably, we architect this distribution to be biased, explicitly allocating a privileged probability (e.g., 0.4) to the first memory chunk (the user-persona congruent chunk), which favors the retrieval of memory context that is most likely to align with the user. $\mathcal{P}_{UG}$ means the response generation prompt.

To facilitate implementation of learning methods such as DPO (Rafailov et al., 2023), an alternative response is subsequently generated by assigning a randomly sampled personality profile that diverges from the user ($\omega'_\varepsilon! = \omega_\varepsilon$). Both responses undergo systematic evaluation ($\mathcal{P}_{UE}(.)$ in Equation 6) across the five personality dimensions. Through comparative assessment (Equation 7) of their respective alignment with the target user personality, the response demonstrating superior congruence is designated as the *preferred*, while the counterpart is categorized as the *rejected*. We also annotate each interaction turn with four candidate responses, which span a spectrum of personalization alignment quality. These additional responses are harnessed as training data for offline GRPO (Shao et al., 2024; Mroueh et al., 2025) learning.

Reward: To operationalize response quality assessment and facilitate the potential adaptation for offline RL methods (Prudencio et al., 2022; Jackson et al., 2025; Wang et al., 2025a), we design



(a) Train Set



(b) Test Set

Figure 3: Train & Test Sets Statistics on Every turn and Every Dialogue Aspects. For each dialogue turn, a clustering analysis is performed via an LLM on the five memory contents to statistically determine the **Memory Content Types**. The rewards associated with each **Memory Selection** and its corresponding **Generated Response** are then quantified using Equation 7.

a composite reward function evaluating two critical alignment dimensions: 1) *Memory-Persona Alignment*, to quantify the congruence between the selected memory chunk's personality expression and the user's ground-truth persona. 2) *Response-Persona Alignment*, to assess the personality consistency manifested in the response content itself. Both evaluations are conducted against the standardized 5-dimensional personality profile in Section 2.1, employing an identical scoring regime: a perfect match among all the dimensions awards 1.0, with a linear deduction of 0.2 for each deviating dimension.

$$r_t = [1.0 - 0.2n_{m,t}, 1.0 - 0.2n_{u,t}], \quad (7)$$

where $n_{m,t}$ and $n_{u,t}$ ($\in [1, 5]$) serve as the mismatched number of the dimensions for the memory chunk and response, respectively, demonstrating deviation from the user persona

3 Experiments

We employ the 8b size LLMs (Llama3.1-8b (Grattafiori et al., 2024) and Qwen3-8b (Yang et al., 2025)) for both prompt and post-training based evaluations with the dataset, demonstrating their respective win-rate and reward achievements. This integrated experimental design effectively highlights the distinctive characteristics of personalized memory selection mechanisms.

Post-training Methods: We conduct dataset evaluations based on post-training algorithms such as SFT, PPO (Ouyang et al., 2022), DPO, and GRPO. For these methods, we indiscriminately feed all memory contents into the context without performing explicit memory selection (content-similarity/personalized selection). The model subsequently learns an implicit memorization process guided by output signals, i.e., response attributes and reward values.

Evaluation Metrics: We evaluate the effectiveness of memory selection and response generation from two perspectives: win-rate and reward achievement. For win-rate, we employ an LLM-based judgement approach to assess the relative superiority of two compared outputs in terms of fulfilling personalized requirements. The frequencies of win (N_w), loss (N_l), and tie (N_t) are respectively tallied, and the final win-rate is computed as: $r_\omega = \frac{N_w - N_l}{N_w + N_l + N_t}$ (Ji et al., 2024). For reward calculation, drawing on the methodology of response persona alignment (Equation 7), we measure the alignment between the response and the user persona across five dimensions to derive the reward value. Notably, prompt-based methods, i.e., RAG, explicitly select a memory (each annotated with Big Five labels and a reward in the dataset), so the memory reward can be directly accessed. For post-training methods, which generate responses from context, memory-based evaluation is performed by first using an LLM to identify the most probable memory corresponding to the context-response pair, and then indexing its reward from the dataset:

$$m_t = \mathbf{LLM}(\mathcal{P}_{MD}(\varepsilon, h_t, q_t, \mathcal{M}_t, u_t)), \quad (8)$$

where $\mathcal{P}_{MD}$ means the memory acquisition prompt. u_t is the response at turn t. $\{\varepsilon, h_t, q_t, \mathcal{M}_t\}$ constitutes the dialogue context.

3.1 Basic Statistics

A dialogue dataset is curated with 4,000 training and 1,000 test entries, each item spanning six turns and five memory candidates per turn, one of which matched the user persona. For each turn, an LLM is tasked with deriving the most relevant memory in content to guide the response. As is displayed in Figure 2, evaluation using the metric of persona consistency shows that only 7,862 training turns (32.76%) and 1,873 test turns (31.22%) feature the persona-consistent memory as relevant. From a dialogue-level perspective, the proportion of dialogues in which all selected memories perfectly match the persona memories is merely 4.4% (with this figure being 3.5% in the test set). Dialogues where over half of the turns achieve memory matching account for 35.7% of the total (34.2% in the test set). In contrast, dialogues exhibiting no memory matching across all turns constitute a substantial 28.4% (30.1% in the test set).

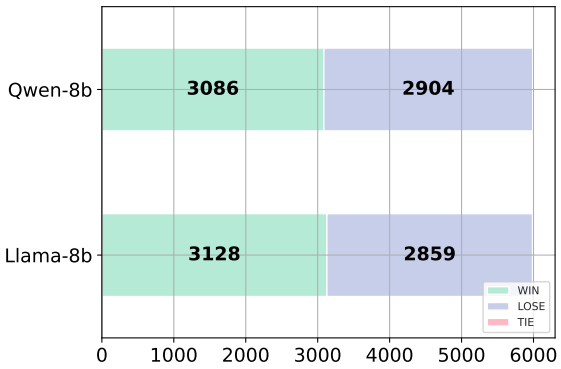


Figure 4: The win-rate evaluation for responses of the personalized memory selection and the RAG.

This significant discrepancy demonstrates that **prevailing RAG for memory selection, which rely predominantly on relevance, are inadequate for fostering the true personalization**.

Figure 3 illustrates the distribution of reward values for memory selection and response generation across the dataset. In the training set, the selected memory matches the user persona in 9,615 turns (40.06%, the memory selection reward being 1.0). This proportion aligns with the predefined sampling probability for the memory selection (p_ε in Equation 6). In the test set, this figure is 2,449 turns (40.82%). Furthermore, it indicates that a substantial portion of the data trajectories represent high-reward scenarios. Specially, at the turn level, the proportion of dialogue turns with a reward ex-

Method	Llama-8b			Qwen-8b		
	Memory Ratio	Memory Reward	Response Reward	Memory Ratio	Memory Reward	Response Reward
RAG-based	0.3122	4.0288	2.5074	0.3122	4.0288	2.5194
Persona-based	0.5215	4.6700	2.9204	0.5772	4.8434	3.0432

Table 3: A comparative evaluation of the personalized memory selection versus RAG: accuracy rates in memory selection and reward metrics for memory retrieval and response generation.

Method	Llama-8b				Qwen-8b			
	win.	loss.	tie.	r_ω	win.	loss.	tie.	r_ω
SFT	3364	2634	2	0.1217	3328	2672	0	0.1093
PPO	4016	1984	0	0.3387	3483	2517	0	0.1610
DPO	4412	1586	2	0.4710	4118	1879	3	0.3732
GRPO	4332	1667	1	0.4442	4027	1973	0	0.3423

Table 4: Win-rate evaluation for post-training methods with the direct LLM response as the reference.

ceeding 0.6 is 68.5% (70.2% in the test set); at the dialogue level, dialogues with a total return value above 4.0 account for 39.6% of the total (40.1% in the test set).

3.2 Prompt-based Methods Evaluation

We validate the advantages of the **Personalized Memory Selection** method over the **RAG** approach on the dataset. For the RAG baseline, the most relevant memory is selected from the memory contents of the current turn, based on which the base LLM generates a response. As for the personalized memory selection process, we employ an offline RL method to train a lightweight memory selection policy (0.6B parameters). In this setup, the action space is defined as $[0, 1, 2, 3, 4]$, corresponding to the five memory chunks in the dataset, and the feedback signal is the memory selection reward, i.e., the first element in the reward vector specified in Equation 7. The state s comprises the dialogue history, persona information, and memory contents. The policy gradient expression is then formed as:

$$\nabla \mathcal{J} = \mathbb{E}_{s,m}[\delta(s) \nabla log \pi(m|s)], \tag{9}$$

where $\delta(s)$ represents the advantage value for s.

Figure 4 demonstrates the enhancement in personalized response quality achieved through personalized memory selection in comparison to RAG. It can be observed that with the Llama-8b model, the number of superior responses ($N_w - N_l$) reaches 289, corresponding to a win-rate of 4.48%; for Qwen-8b, the number is 182, with a win-rate of 3.03%.

Table 3 presents the results of the two approaches in terms of memory ratio and reward performance. The **memory ratio** is defined as the proportion of dialogue turns in which the selected memory matches the personalized memory. As shown, the implementation of a personalized memory selection policy on the Llama-8b model substantially elevates the proportion of persona-aligned memories to 20.93%. In terms of reward, compared to RAG, memory selection reward increases by 0.6412 (a relative improvement of 15.9%), and response reward improves by 0.4130 (16.5%). For the Qwen-8b model, the improvements in memory ratio, memory selection and response generation rewards are 26.50%, 0.8146 (20.2%), and 0.5283 (20.8%), respectively.

LLM	Memory Ratio	Memory Reward	Response Reward
Llama-8b	0.3780	4.2372	2.6017
Qwen-8b	0.3347	4.1858	2.5568

Table 5: Memory ratio and rewards evaluation for the direct LLM output.

Method	Memory Ratio			Memory Reward			Response Reward		
	v	g_a	g_r (%)	v	g_a	g_r (%)	v	g_a	g_r (%)
SFT	0.4620	0.0840	22.22	4.4698	0.2326	5.49	2.8243	0.2226	8.56
PPO	0.7280	0.3500	92.59	5.3104	1.0732	25.33	3.8485	1.2468	47.92
DPO	0.7527	0.3747	99.12	5.4084	1.1712	27.64	4.1144	1.5127	58.14
GRPO	0.8255	0.4475	118.38	5.5796	1.3424	31.68	4.2523	1.6506	63.44

(a) Llama-8b

Method	Memory Ratio			Memory Reward			Response Reward		
	v	g_a	g_r (%)	v	g_a	g_r (%)	v	g_a	g_r (%)
SFT	0.4561	0.1214	36.27	4.4528	0.2670	6.38	2.7993	0.2425	9.48
PPO	0.7802	0.4455	113.10	5.4114	1.2256	29.28	4.0748	1.5180	59.37
DPO	0.8580	0.5233	156.35	5.6708	1.4850	35.48	4.3994	1.8426	72.07
GRPO	0.7945	0.4598	137.38	5.4616	1.2758	30.48	4.1716	1.6148	63.16

(b) Qwen-8b

Table 6: Memory ratio and rewards evaluation for the post-training methods. We report their values (v) and absolute/relative (g_a/g_r) gains against the direct LLM output.

3.3 Post-training Methods Evaluation

Similarly, we assess the personalization efficacy by training on this personalized contextual data with various post-training algorithms. The direct LLM output (the metric values are displayed in Table 5) is used as a base to measure the relative improvement of each method. Specifically, for PPO and GRPO, we employ an offline learning paradigm, utilizing ground-truth responses and rewards from the dataset as the sampled trajectories in place of dynamic online rollouts.

Table 4 presents the improvements achieved by these post-training methods in terms of win-rate metric. It is clearly demonstrated that training on our personalized contextual dataset leads to win-rate gains ranging from 10% to 47%. Among them, the SFT method yields relatively modest improvements, with increases of 12.17% and 10.93% on the Llama-8b and Qwen-8b models, respectively. In contrast, DPO exhibits the most substantial gains, reaching 47.10% and 37.32% on the two models respectively. For PPO and GRPO algorithms, despite being trained via an offline strategy based on the dataset, effective response strategies are still learned from the feedback signals, i.e., response rewards. Consequently, their win-rate results are significantly higher than those of SFT and remain competitive with DPO, with GRPO performing nearly on par.

We also present a comparative analysis of post-training methods against the base (direct LLM output) in terms of memory ratio and reward metrics. The results, detailed in Table 6, are reported using the following notation: v denotes the value for post-training methods, g_a represents the absolute gain over the base, and g_r indicates the relative improvement. Evidently, these methods yield substantial improvements across aforementioned evaluated metrics on different models. On the Llama-8b model, GRPO delivers the greatest improvement across key metrics: memory ratio increases by 0.4475 (+118.38%), memory selection reward by 1.3424 (+31.68%), and response generation reward by 1.6506 (+63.44%). Additionally, DPO demonstrates its peak efficacy on Qwen-8b, with enhancements of 0.5233 (+156.35%), 1.4850 (+35.48%), and 1.8426 (+72.07%) in the respective metrics. Although SFT yields the most modest gains, fine-tuning on the personalized contextual dataset nonetheless leads to measurable personalization alignment, producing average improvements of 29.25% in memory ratio, alongside 5.94% and 9.02% in the reward metrics.

4 Conclusion

We propose a novel memory processing mechanism that integrates memory selection with personalized alignment. Leveraging this mechanism, we

introduce a schema for personalized memory building and memory-based generation, and construct a multi-turn dialogue dataset alongside personalized contextual memory. Evaluations across both prompt-based and post-training-based algorithms on this dataset validate the critical relationship between memory selection and personalization.

Limitations

We propose a multi-turn dialogue dataset featuring personalized contextual memory and use it to underscore the significance of personalized alignment for memory selection. It should be noted that the current evaluation is predominantly reliant on LLM-as-a-judge, which necessitates further enhancement in terms of credibility and accuracy. Future work will focus on developing more objective metrics to precisely characterize the extent of personalization.

Acknowledgments

This work is supported by China Mobile Strategic Project (R26110S3, R24113J4).

References

Courtney E. Ackerman. 2020. Big five personality traits. *Encyclopedia of Education and Information Technologies*.

Akari Asai, Zeqiu Wu, Yizhong Wang, Avirup Sil, and Hannaneh Hajishirzi. 2024. Self-RAG: Learning to retrieve, generate, and critique through self-reflection. In *The Twelfth International Conference on Learning Representations*.

Shulin Cao, Jiaxin Shi, Liangming Pan, Lun Yiu Nie, Yutong Xiang, Lei Hou, Juanzi Li, Bin He, and Hanwang Zhang. 2020. Kqa pro: A dataset with explicit compositional programs for complex question answering over knowledge base. In *Annual Meeting of the Association for Computational Linguistics*.

Yiming Du, Wenyu Huang, Danna Zheng, Zhaowei Wang, Sebastien Montella, Mirella Lapata, Kam-Fai Wong, and Jeff Z. Pan. 2025. Rethinking memory in ai: Taxonomy, operations, topics, and future directions. *Preprint*, arXiv:2505.00675.

Yiming Du, Bingbing Wang, Yang He, Bin Liang, Baojun Wang, Zhongyang Li, Lin Gui, Jeff Z. Pan, Ruifeng Xu, and Kam-Fai Wong. 2026. Memguide: Intent-driven memory selection for goal-oriented multi-session llm agents. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(36):30584–30592.

Yiming Du, Hongru Wang, Zhengyi Zhao, Bin Liang, Baojun Wang, Wanjun Zhong, Zezhong Wang, and Kam-Fai Wong. 2024. Perltqa: A personal long-term memory dataset for memory classification, retrieval, and synthesis in question answering. *ArXiv*, abs/2402.16288.

Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, Qianyu Guo, Meng Wang, and Haofen Wang. 2023. Retrieval-augmented generation for large language models: A survey. *ArXiv*, abs/2312.10997.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, Amy Yang, Angela Fan, Anirudh Goyal, Anthony Hartshorn, Aobo Yang, Archi Mitra, Archie Sravankumar, Artem Korenev, Arthur Hinsvark, and 542 others. 2024. The llama 3 herd of models. *Preprint*, arXiv:2407.21783.

Jiawei Gu, Xuhui Jiang, Zhichao Shi, Hexiang Tan, Xuehao Zhai, Chengjin Xu, Wei Li, Yinghan Shen, Shengjie Ma, Honghao Liu, Yuanzhuo Wang, and Jian Guo. 2024. A survey on llm-as-a-judge. *ArXiv*, abs/2411.15594.

Kostas Hatalis, Despina Christou, Joshua Myers, Steven Jones, Keith Lambert, Adam Amos-Binks, Zohreh Dannenhauer, and Dustin Dannenhauer. 2024. Memory matters: The need to improve long-term memory in llm-agents. *Proceedings of the AAAI Symposium Series*.

Matthew Thomas Jackson, Uljad Berdica, Jarek Luca Liesen, Shimon Whiteson, and Jakob Nicolaus Foerster. 2025. A clean slate for offline reinforcement learning. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*.

Jihyoung Jang, Taeyoung Kim, and Hyounghun Kim. 2024. Mixed-session conversation with egocentric memory. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 11786–11815, Miami, Florida, USA. Association for Computational Linguistics.

Jiaming Ji, Boyuan Chen, Hantao Lou, Donghai Hong, Borong Zhang, Xuehai Pan, Tianyi Qiu, Juntao Dai, and Yaodong Yang. 2024. Aligner: Efficient alignment by learning to correct. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*.

Eunwon Kim, Chanho Park, and Buru Chang. 2024. Share: Shared memory-aware open-domain long-term dialogue dataset constructed from movie script. In *Annual Meeting of the Association for Computational Linguistics*.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020.

Retrieval-augmented generation for knowledge-intensive nlp tasks. In *Advances in Neural Information Processing Systems*, volume 33, pages 9459–9474. Curran Associates, Inc.

Hao Li, Chenghao Yang, An Zhang, Yang Deng, Xiang Wang, and Tat-Seng Chua. 2025a. Hello again! LLM-powered personalized agent for long-term dialogue. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 5259–5276, Albuquerque, New Mexico. Association for Computational Linguistics.

Xintong Li, Jalend Bantupalli, Ria Dharmani, Yuwei Zhang, and Jingbo Shang. 2025b. Toward multi-session personalized conversation: A large-scale dataset and hierarchical tree framework for implicit reasoning. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 11504–11517, Suzhou, China. Association for Computational Linguistics.

Alex Troy Mallen, Akari Asai, Victor Zhong, Rajarshi Das, Hannaneh Hajishirzi, and Daniel Khashabi. 2022. When not to trust language models: Investigating effectiveness of parametric and non-parametric memories. In *Annual Meeting of the Association for Computational Linguistics*.

Youssef Mroueh, Nicolas Dupuis, Brian M. Belgodere, Apoorva Nitsure, Mattia Rigotti, Kristjan H. Greenewald, Jirí Navrátil, Jerret Ross, and Jesus Rios. 2025. Revisiting group relative policy optimization: Insights into on-policy and off-policy training. *ArXiv*, abs/2505.22257.

Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. 2022. Training language models to follow instructions with human feedback. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, Red Hook, NY, USA. Curran Associates Inc.

Zhuoshi Pan, Qianhui Wu, Huiqiang Jiang, Xufang Luo, Hao Cheng, Dongsheng Li, Yuqing Yang, Chin-Yew Lin, H. Vicky Zhao, Lili Qiu, and Jianfeng Gao. 2025. Secom: On memory construction and retrieval for personalized conversational agents. In *The Thirteenth International Conference on Learning Representations*.

Rafael Figueiredo Prudencio, Marcos R.O.A. Maximo, and Esther Luna Colombini. 2022. A survey on offline reinforcement learning: Taxonomy, review, and open problems. *IEEE Transactions on Neural Networks and Learning Systems*, 35:10237–10257.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. 2023. Direct preference optimization: Your language model is secretly a reward model. In *Thirty-seventh Conference on Neural Information Processing Systems*.

Timothy B. Rogers, Nicholas A. Kuiper, and W. S. Kirker. 1977. Self-reference and the encoding of personal information. *Journal of personality and social psychology*, 35 9:677–88.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. 2024. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *Preprint*, arXiv:2402.03300.

Alon Talmor and Jonathan Berant. 2018. The web as a knowledge-base for answering complex questions. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 641–651.

Huaijie Wang, Shibo Hao, Hanze Dong, Shenao Zhang, Yilin Bao, Ziran Yang, and Yi Wu. 2025a. Offline reinforcement learning for LLM multi-step reasoning. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 8881–8893, Vienna, Austria. Association for Computational Linguistics.

Yilei Wang, Jiabao Zhao, Deniz S. Ones, Liang He, and Xin Xu. 2025b. Evaluating the ability of large language models to emulate personality. *Scientific Reports*, 15.

Zilong Wang, Zifeng Wang, Long Le, Steven Zheng, Swaroop Mishra, Vincent Perot, Yuwei Zhang, Anush Mattapalli, Ankur Taly, Jingbo Shang, Chen-Yu Lee, and Tomas Pfister. 2025c. Speculative RAG: Enhancing retrieval augmented generation through drafting. In *The Thirteenth International Conference on Learning Representations*.

Charlotte Grosse Wiesmann, Katrin Rothmaler, Esra Hasan, Kathrine Habdank, Chen Yang, Emanuela Yeung, and Victoria Southgate. 2025. The self-reference memory bias is preceded by an other-reference bias in infancy. *Nature Communications*, 16(1):1–8.

Wujiang Xu, Zujie Liang, Kai Mei, Hang Gao, Juntao Tan, and Yongfeng Zhang. 2025. A-mem: Agentic memory for LLM agents. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, Chujie Zheng, Dayiheng Liu, Fan Zhou, Fei Huang, Feng Hu, Hao Ge, Haoran Wei, Huan Lin, Jialong Tang, and 41 others. 2025. Qwen3 technical report. *Preprint*, arXiv:2505.09388.

Ruifeng Yuan, Shichao Sun, Yongqi Li, Zili Wang, Ziqiang Cao, and Wenjie Li. 2024. Personalized large language model assistant with evolving conditional memory. *Preprint*, arXiv:2312.17257.

Wanjun Zhong, Lianghong Guo, Qiqi Gao, He Ye, and Yanlin Wang. 2024. Memorybank: enhancing large language models with long-term memory. In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'24/IAAI'24/EAAI'24. AAAI Press.

A Related work

The development of memory-related datasets has predominantly focused on single-turn, multi-turn interactions, and personalization. However, no existing dataset adequately captures the nuanced characteristics of personalized user memory processing.

Specifically, in the domain of Question Answering (QA), the ComplexWebQuestions dataset (Talmor and Berant, 2018) comprises a wide range of complex queries that necessitate reasoning over multiple web snippets and extracting relevant knowledge to generate accurate responses. The KQA Pro dataset (Cao et al., 2020) provides explicit compositional reasoning programs (KoPL) and SPARQL query annotations, facilitating models' acquisition of explicit reasoning capabilities and enabling them to retrieve context that meets requirements from knowledge bases via query statements. The POPQA dataset (Mallen et al., 2022) consists of 14k questions designed to cover factual information in the long-tail distribution, serving as a benchmark to evaluate the extent of factual knowledge memorization in large-scale models.

In the context of multi-turn interactions, the MISC dataset (Jang et al., 2024) comprises multi-session conversational data, wherein discussions on event contents and varying dialogues are conducted across the first five sessions; in the final session, the previously discussed content is utilized as memory to facilitate the ongoing dialogue. The SHARE dataset (Kim et al., 2024), constructed from movie scripts, incorporates personal character profiles and events while implicitly extracting shared memories. Its primary objective is to enhance long-term conversational coherence and engagement, aiming to address the existing deficiencies in dialogue systems regarding the handling of long-term shared memory. The MS-TOD dataset (Du et al., 2026) integrates memory selection with task-specific goals, empirically validating the critical role of task objectives in the process of memory selection.

Additionally, there exist interaction datasets oriented toward personalization. For instance, the PerLTQA dataset (Du et al., 2024) integrates semantic and episodic memory, encompassing world knowledge, profiles, social relationships, events, and dialogues, to investigate the utilization of personalized memory by large-scale models. ImplexConv (Li et al., 2025b) constructs long-range interactive content, aiming to explore implicit reasoning in personalized dialogues, where relevant informa-

tion is embedded through subtle, syntactically or semantically distant connections rather than being explicitly stated.

B The Big-Five Personality Theory

We provide a comprehensive enumeration of the five-dimensional values and characteristic descriptions for 32 types of Big-Five personalities, as presented in Table 7. In the table (the **OCEAN** column), the value of 0 indicates that the corresponding dimensional trait (*Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism*) is not prominent, whereas 1 indicates a pronounced manifestation of the trait. As an illustration, the OCEAN code "00011" corresponds to a trait profile defined by significant expression in Agreeableness and Neuroticism (dimensions four and five), with the other dimensions showing minimal or absent trait salience. During the data generation process, descriptions of these personalities are incorporated as user characteristics into the LLM context, thereby facilitating memory personality acquisition, memory generation, response synthesis, and reward evaluation.

OCEAN	**Description**
00000	Down-to-earth, spontaneous, solitary, and competitive. They are emotionally stable but can seem detached and unmotivated by conventional goals.
00001	Practical but disorganized, reserved, skeptical, and prone to anxiety. They may feel pessimistic about the world and struggle with instability and negative emotions.
00010	A practical, easy-going introvert who is cooperative and calm. They avoid conflict and drama, preferring a peaceful and predictable life alone or in small circles.
00011	Conventional and flexible, quiet but kind. Their high agreeableness leads them to want to help, but their neuroticism means they are often worried and emotionally vulnerable.

Continued on next page

Table 7: 32 foundational personality archetypes.

00100	A pragmatic, spontaneous, and socially confident type. They are competitive, emotionally resilient, and focused on the here-and-now, often thriving in fast-paced, transactional environments.
00101	Sociable and energetic but also practical, spontaneous, and challenging. Their combination of extraversion and low agreeableness can make them argumentative, and their neuroticism makes them prone to mood swings.
00110	The life of a practical party. They are outgoing, cooperative, and calm. While they prefer concrete ideas and can be spontaneous, their agreeableness and stability make them pleasant and easy to be around.
00111	Outgoing, friendly, and conventional, but also disorganized and prone to worry. They seek social harmony and connection but are often stressed and emotionally sensitive.
01000	Highly dependable, organized, and practical. They are reserved, direct, competitively, and emotionally calm. They are the rock-solid, no-nonsense backbone of any operation.
01001	Practical, disciplined, and reserved. They have high standards, can be skeptical of others, and their neuroticism drives them to anxiety over details and a fear of making mistakes.
01010	A stable, reliable, and practical pillar of the community. They are reserved, kind, and organized, providing quiet, unwavering support without seeking the spotlight.
01011	Organized, conventional, and reserved. They are cooperative and kind but are often plagued by anxiety and a strong need for security and predictability.

Continued on next page

Table 7: 32 foundational personality archetypes. (Continued)

01100	A natural manager who is practical, efficient, and assertive. They are socially confident and competitive but emotionally stable, allowing them to drive results without being derailed by emotion.
01101	A pragmatic, organized, and outgoing leader. However, their competitiveness and skepticism, combined with high neuroticism, can make them a demanding and easily stressed authority figure.
01110	The friendly, reliable community pillar. They are practical, organized, outgoing, cooperative, and calm. They excel at maintaining social harmony and getting things done reliably.
01111	Organized, conventional, and outgoing. They are highly agreeable and strive to please everyone, but this often leads to stress and anxiety as they take on too much responsibility for others' feelings.
10000	A creative, independent thinker who is spontaneous, reserved, and unemotional. They are intellectually competitive and enjoy debating ideas, but are not driven by social needs.
10001	Imaginative and curious but disorganized and solitary. They are intellectually competitive and skeptical, with a turbulent inner world filled with anxiety and intense emotions.
10010	A creative, easy-going, and calm introvert. They are cooperative and kind, often lost in their own imaginative world, and unbothered by the need for social status or rigid schedules.

Continued on next page

Table 7: 32 foundational personality archetypes. (Continued)

10011	Creative, gentle, and introverted. They have a rich inner world and deep empathy, but their spontaneity and high neuroticism make them vulnerable to emotional turmoil and overwhelm.
10100	Energetic, innovative, and intellectually challenging. They love exploring new ideas and arguing for fun. They are emotionally stable and spontaneous, making them exciting and unpredictable.
10101	Charismatic, creative, and spontaneous. They are full of new ideas and energy but can be competitive, argumentative, and emotionally volatile, leading to dramatic interpersonal conflicts.
10110	An energetic, creative, and friendly free spirit. They are cooperative and stable, making them great at brainstorming and bringing people together around new, exciting projects.
10111	Enthusiastic, creative, and deeply caring. They champion causes and connect with people easily, but their spontaneity and high neuroticism can lead to burnout and emotional exhaustion.
11000	A visionary and systematic thinker. They are organized, reserved, and enjoy intellectual debate. They are calm and self-assured in their well-thought-out beliefs and creations.
11001	A highly creative and organized planner who is reserved and intellectually competitive. Their neuroticism manifests as anxiety over their complex plans and a fear that their ideas aren't good enough.
11010	Thoughtful, principled, and creative. They are organized and cooperative, providing deep, calm, and insightful advice. They are the stable, trusted confidant.

Continued on next page

Table 7: 32 foundational personality archetypes. (Continued)

11011	Imaginative, dutiful, and reserved. They are deeply caring and strive to create a perfect, harmonious world, but this leads to constant worry and stress over every detail.
11100	The quintessential entrepreneur or visionary CEO. They are open to new ideas, highly organized, socially dominant, competitively, emotionally resilient in face of challenges.
11101	A charismatic, innovative, and organized leader. They have a bold vision and drive to achieve it, but their competitiveness and neuroticism can make them intense, demanding, and prone to stress.
11110	The ideal and well-rounded leader. They are imaginative, disciplined, energetic, compassionate, and calm under pressure. They inspire loyalty and drive change effectively and harmoniously.
11111	An inspiring, creative, and organized leader who connects deeply with others. They are driven by idealism and empathy, but they carry the weight of the world on their shoulders, leading to high stress and emotional volatility.

Table 7: 32 foundational personality archetypes. (Continued)

C Prompts for Data Generation

We have meticulously documented the prompt engineering employed in the data generation pipeline, the specific functions and rationales of which are delineated in Table 2.

1: Personality-Generation Prompt: $\mathcal{P}_\Omega$

You will be given a personality description that is characterized by specific scores on the Big Five personality traits. Your task is to rewrite this description using different words and sentence structures, while ensuring that the core meaning and the implied level of each of the five traits remain unchanged.

Original Description:
{persona}

The Big Five Value:
{big_5}

Instructions:
- The rephrased version must be semantically equivalent to the original.
- The perceived levels of all five personality dimensions must be identical in the new description.
- Output ONLY the final rephrased text. STRICTLY do not include any introductions, explanations, or acknowledgments.

2: Initial-Query-Generation Prompt: $\mathcal{P}_{IQG}$

Generate a multi-turn dialogue between {s1} and {s2} according to the following specifications:

- **Topic:**
{event}
- **Speakers Information:**
{s1} is {s1_job}. {s2} is {s2_job}.
- **Personality of {s1}:**
{persona}

Instructions:
1. Begin the dialogue with the first utterance from {s1}.
2. Ensure this opening line and the subsequent conversation naturally reflect the specified personality traits.
3. Maintain a fluid and interactive exchange that explores the given topic.
4. Use English only.

First, output {s1}'s opening line. Output dialogue content only.

3: Query-Generation Prompt: $\mathcal{P}_{QG}$

Continue the dialogue according to the following specifications:

- **Scenario:**

A dialogue has occurred between {s1} and {s2}. {s1} is {s1_job}. {s2} is {s2_job}.
- **Topic:**
{event}
- **Personality of {s1}:**
{persona}
- **Dialogue history:**
{history}

Instructions:
1. Continue the dialogue with the utterance from {s1} based on the provided history.
2. Foster the dialogue's progression while precluding redundancy from the history.
3. Ensure this opening line and the subsequent conversation naturally reflect the specified personality traits of {s1}.
4. Maintain a fluid and interactive exchange that explores the given topic.
5. Output ONLY {s1}'s next line of dialogue. Use English only.

4: **Memory-Generation Prompt:** $\mathcal{P}_{MG}$

Context:
- Scenario:
A dialogue has occurred between {s1} and {s2}. {s1} is {s1_job}. {s2} is {s2_job}.
- Core Topic:
{event}
- {s1}'s Personality:
{persona}
- Current Dialogue Context:
{history}
- {s1}'s Current Inquiry:
{query}

---

*Background:**
Prior to this, {s1} discussed this same topic with {s3}.{s3} is known for a communication style characterized by: {persona_sup}. {attitude}.

---

*Your Task:**
Synthesize all the information above to generate a summarized version of {s3}'s likely perspective to serve as {s1}'s memory content in the current conversation. Your generation must be:
- Deeply Informed: The content must directly reflect the core topic, align with {s3}'s communication style.
- Context-Aware: It should be a plausible response to the "Current Dialogue Context" and "{s1}'s Current Query", as if {s3} were stepping into the conversation now.
- Concise & Impactful: Keep {s3}'s perspective concise yet impactful. Present as a narrative summary, NOT in dialogue form.
- Extra Information: You may reasonably infer additional context to show {s3}'s familiarity with both {s1} and {s2}.
- Diverse Viewpoints: {s3} possesses the capacity to formulate divergent perspectives that contrast with those of {s1} or {s2}, while retaining the discretion to align with their viewpoints.

Now, Output only the text of {s3}'s opinion. Do not include any meta-commentary, or narrative descriptions.

5: **Response-Generation Prompt:** $\mathcal{P}_{UG}$

Context:
- Scenario:
A dialogue has occurred between {s1} and {s2}. {s1} is {s1_job}. {s2} is {s2_job}.
- Core Topic:
{event}
- {s1}'s Personality:
{persona}
- Dialogue History:
{history}
- {s1}'s Current Inquiry:
{query}
- Background Information:
{memory}

---

*Task:**
Continue the dialogue with the utterance from {s2} based on the provided history to respond to {s1}'s current inquiry.

---

*Requirements:**
- Craft a reply that aligns with {s1}'s personality traits
- Ensure the response would likely satisfy {s1}

- Consider dialogue history and relevant background information
- Maintain natural conversational flow
- Output only {s2}'s direct response, without additional explanations

Output dialogue content only. Use English only.

6: Response-Evaluation Prompt: $\mathcal{P}_{UE}$

Task:
Analyze the following description against the Big Five personality traits (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism). For each trait, judge if the match is 'High' or 'Low'.

Output Format:
A list of five values, i.e., ["High"/"Low", "High"/"Low", "High"/"Low", "High"/"Low", "High"/"Low"], in O-C-E-A-N order

Description:
{response}

Output ONLY the final list. STRICTLY do not include any introductions, explanations, or acknowledgments.

D Prompts for Evaluation

The evaluation protocol comprises three key prompt types: 1) for memory indexing corresponding to generated responses (i.e., $\mathcal{P}_{MD}$ in Equation 8), 2) for pairwise win-rate evaluation, 3) for post-training process, respectively.

7: Response-Indexing Prompt: $\mathcal{P}_{MD}$

Task: Analyze the provided context and determine which memory entry (Index 1–5) the given response is generated from. Follow these steps:

*Review the following information**:
– Event
{event}
– Dialogue History
{history}
– Current Query
{query}
– Memories
{memory}
– Response
{response}

*Compare the response with the five memories (Index 1–5)**:
Identify the memory that most directly aligns with or explains the content of the response.

*Output only the following in a list format: ["Index", "Reason"]**
– Index: The number of the better memory to obtain the response (from 1 to 5).
– Reason: A concise, single-sentence explanation for your choice, directly referencing the criteria.

Output ONLY the final list. STRICTLY do not include any introductions, explanations, or acknowledgments.

8: Win-rate Evaluation Prompt

Task: Below are two candidate responses to the user's inquiry, based on Event and Dialogue History, considering the user's personalized profile. Please determine which response is more effective in promoting further dialogue and better aligns with the user's personalized characteristics.

- Event:
{event}

- User Profile:
1. Description: {persona}
2. Detail: {persona key}

- History:
{history}

- User Inquiry:
{query}

- Candidate Responses:
{response}

Please perform the following tasks:
1. Choose the response that is more effective at:
- Aligning with the user's profile: Reflecting the traits and preferences described in User Profile.
- Promoting further dialogue: Encouraging a continued and engaging conversation.
2. Your output must be a valid list in the exact format: ["Index", "Reason"]
- Index: The number of the better response (only 1 or 2) or -1 if the two responses are considered equally suitable.
- Reason: A concise, single-sentence explanation for your choice, directly referencing the criteria.
3. Do not output any other text, commentary, or formatting.

9: **Post-training Prompt**

Task: You are an AI assistant tasked with embodying a specific Persona in your responses. Your goal is to engage in a conversation about an Event in a way that is consistent with this personality and the surrounding Background.
– Input Parameters:

Persona:
{persona}

Event:
{event}

Background Information:
{background}

E A Case for Personalized Memory

Table 8 delineates a conversation between Kevin and Brenda, wherein Kevin assumes the role of "USER" with a personality profile characterized as practical, impulsive, and reserved (OCEAN code 00000). The dataset encompasses five distinct categories of memory, corresponding to codes 00000, 11110, 01110, 11000, and 10111. A content analysis reveals that **Memory #1** conceptualizes the donor's gift not as a "transaction" necessitating reciprocation, but as an extension of life itself—advocating for a present-focused existence. This perspective embodies a pragmatic worldview that aligns coherently with the user persona. Conversely, **Memory #2** acknowledges Kevin's aspiration for purposeful action while interrogating its potential underpinnings in "sensationalism-seeking" motives. It posits that authentic commemoration resides not in accomplishments or posthumous recognition, but in embodying this gift through genuine "being"; it further cautions against transmuting gratitude into instrumentalized "proof". The emphasis here lies in the intentionality of donation, which cannot be straightforwardly classified as pragmatic. Indeed, this viewpoint diverges from the user's impulsive and reserved disposition. This case exemplifies the heterogeneity of memories, where the RAG method's content-similarity-based selection retrieves Memory #2—a choice discordant with the user's personality—resulting in diminished reward valuation.

F LLM Usage Clarification

Throughout the paper, the use of LLMs is solely restricted to the polishing of textual elements, such as lexical or phrasal substitutions, and does not extend beyond this scope.

Key	Content
Event	Kevin is feeling grateful and anxious about his second chance at life after receiving a heart transplant. He is determined to honor the gift and make the most out of his new opportunity, and Brenda is there to support and guide him through it all.
Persona (00000)	Practical, impulsive, reserved, and driven to win. They remain calm under pressure but may appear indifferent and uninterested in traditional ambitions.
Memory #1 (00000)	Kimberly would remind Kevin that the donor's gift wasn't given for repayment, but as a quiet transfer of possibility—life passing the baton. The meaning isn't built in monuments or missions, but in the way he now breathes deep on a cold morning, laughs a little louder, or stays present with Brenda instead of chasing some idealized version of "worthy."
Memory #2 (11110)	Jason'd acknowledge Kevin's drive to "do something that means something" but gently challenge the need for spectacle, suggesting that honoring the donor isn't about legacy or achievement, but about presence.
Memory #3 (01110)	Christopher would remind Kevin that honoring the donor isn't about grand gestures or needing to know their name—it's about living with intention and kindness, every day.
Memory #4 (11000)	David would likely remind Kevin that gratitude, while profound, need not be repaid through grand gestures or external validation.
Memory #5 (10111)	Christine would remind Kevin that grief and gratitude can coexist—that the donor's loss was immense, not just to him but to a family who made an unimaginable choice in their pain.
Memory Index (RAG)	2 ($r = 0.2$)
Memory Index (Personalized Memory Selection)	1 ($r = 1.0$)

Table 8: A case conversation turn (turn 1 of a dialogue) occurred between Kevin and Brenda.