KNOW ME TO RECOMMEND ME: PERSONALIZED RECOMMENDATION FROM LONG-TERM HUMAN-MLLM ASSISTANT INTERACTIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

Imagine that in the near future, when you ask your personal Multimodal Large Language Model (MLLM) assistant, "Recommend a laptop that would suit me," it can recover your preferences from sparse, incidental cues accumulated through everyday interactions and directly identify the right item, without requiring you to repeatedly *spell out all details*. To **bridge this research gap**, we propose **LIP-Rec**, which studies how everyday experience between human and their MLLM assistants can be leveraged for personalized recommendation. To systematically evaluate such capability, we introduce **PersonaRec**, a large-scale benchmark that lays the foundation for moving personalized recommendation from modeling behavioral signals within recommendation platforms toward a new paradigm grounded in *daily interactions with MLLM assistants*. It comprises 200 distinct personas and over 30K human-refined candidate items, spanning three common daily-life domains: Fashion, Food, and Movies, and 21 sub-domains. For each persona, we construct 83.5K length daily multimodal interactions, where item preferences are sparsely and naturally revealed through explicit, implicit, and evolving preference patterns, and also be grounded in multimodal clues. We benchmark frontier MLLMs and find that they perform poorly on PersonaRec: they struggle to recover item preferences from long interaction histories and translate them into effective retrieval queries. Therefore, we propose **Yo'Rec**, a simple baseline method that targets these limitations and enhances open-source MLLMs on PersonaRec. Empirical results show its efficacy, offering a promising direction for LIP-Rec.

1 INTRODUCTION

In the future, when you ask your personal Multimodal Large Language Model (MLLM) assistant: "Help me find a jacket that *fits me*", you *need not* spell out every preference in detail. Instead, your assistant revisits your long-term daily multimodal interactions with it — where you casually revealed preferred colors yesterday, styles last month, and materials months ago — to *recommend your ideal jacket* from vast candidate items. As MLLMs evolve into capable everyday companions, such scenarios are not rare. Rather than manually browsing e-commerce platforms or repeatedly clarifying their needs with recommender chatbots, users are now more inclined to ask their personal MLLM assistants to do this: As in everyday conversations, they expect assistants to draw on the accumulated understanding of their preferences from daily interactions to find suitable items (Tou et al., 2026), without tediously specifying them all. Echoing this trend, leading AI companies launched attempts (Perplexity, 2025; OpenAI, 2025), e.g., OpenAI introduced ChatGPT Shopping Research to explore users' past interactions and memory with ChatGPT for personalized shopping assistance.

Nevertheless, the corresponding research landscape **remains relatively blank**. On one side, recent studies on personalizing MLLM assistants through long-term interactions mainly focus on daily chat and question answering (Alaluf et al., 2024; Nie et al., 2026), while recommendation-oriented scenario remains largely unexplored. On the other side, current recommender system studies, including recent LLM-based approaches, primarily model user preferences from interactions *within recommender platforms*, either user–item behaviors (e.g., clicks, ratings, and purchases) or dialogues with platform-specific recommender chatbots (Lin et al., 2025a; Zhu et al., 2026), leaving long-term daily multimodal interactions with personal MLLM assistants largely untapped.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

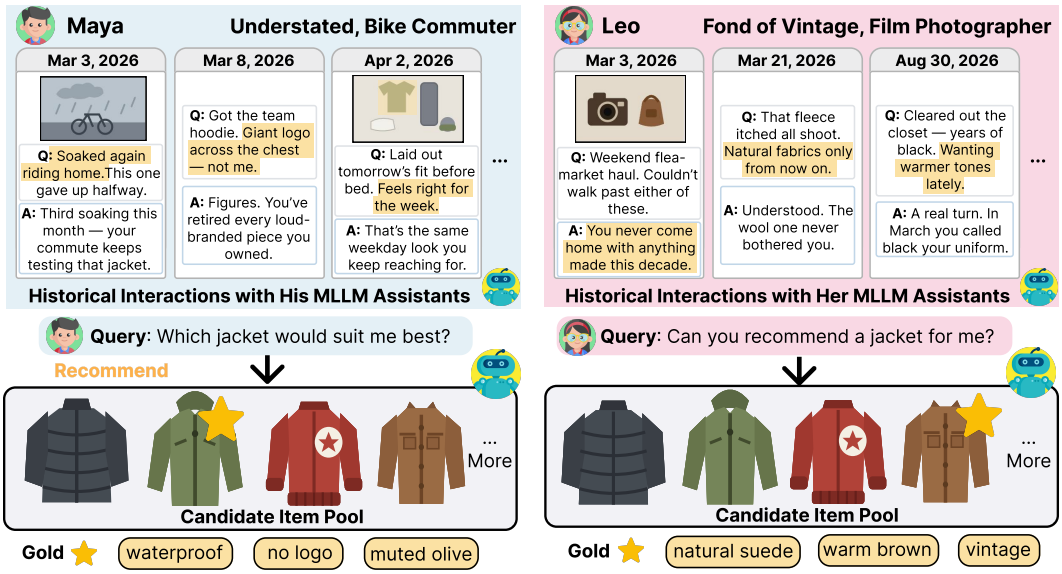


Figure 1: A glance of our proposed LIP-Rec task, where a similar recommendation query from different personas lead to distinctly personalized recommendations.

To bridge this gap, we propose **LIP-Rec**, a novel research task that studies how **L**ong-term human–MLLM assistant **I**nteractions can support **P**ersonalized **Rec**ommendation. Specifically, LIP-Rec diagnoses *whether a personal MLLM assistant can utilize dispersed user-specific cues and evidence from past multimodal, long-term daily interactions, which may incidentally reveal stable preferences over product or content characteristics, to help different individuals recommend the items that best align with their needs.* (Figure 1). As a response, we propose **PersonaRec**, the first comprehensive benchmark tailored for LIP-Rec. Our benchmark features 200 well-characterized personas, each paired with unique target items and 83.5K length daily multimodal interaction histories with their personal MLLM assistants, where target-relevant preferences may occasionally emerge across *coherent and naturalistic* conversational interactions. To ground these personalized signals and intents in *realistic recommendation spaces*, we further construct three representative everyday item domains —"Fashion, Food, and Movies" — with 21 sub-domains, over 30K high-quality manually refined items, and 421 carefully annotated discriminative attributes. In addition, PersonaRec exhibits several salient properties, including **diverse preference dynamics**, where target-relevant preferences may emerge explicitly or implicitly and evolve over long-term daily interactions, and **visually grounded preference inference**, where partial multimodal interactions require models to infer user preferences from *visual content*. We also conduct human audits to ensure the realism of the constructed interactions and the overall reliability of the benchmark. As a result, an interesting phenomenon comes from the PersonaRec: *similar* underspecific recommendation queries from different personas may lead their personal MLLM assistants to recommend *different* products (cf. Figure 1).

We benchmark both frontier proprietary and open-source MLLMs, and find that both struggle with LIP-Rec. Through in-depth analysis, we uncover two major failure modes: incomplete or spurious capture of personalized preferences for the query-anchored item from long-term multimodal interactions, where preference signals may even arise from different modalities, and ineffective transformation of recovered preferences into recommendation queries. Building on these observations, we propose a simple baseline, namely **Yo'Rec**, which enhances **Yo**ur personal MLLM assistant with more customized **Rec**ommendation capabilities from your long-term daily interactions. Specifically, through a three-step Personalized Recommendation Preference Location Chain-of-Thoughts (Wei et al., 2022) (**PRPL-CoT**) based SFT and Personalized Recommendation-query Completion Policy Optimization (**PRC-PO**), Yo'Rec improves the capability of LIP-Rec on open-source MLLMs.

Our Contributions. First, we introduce LIP-Rec, a new task for personalized recommendation grounded in long-term human–MLLM assistant interactions. Second, we present PersonaRec benchmark, which lays a solid foundation for studying LIP-Rec, and diagnose failures of frontier MLLMs. Third, we propose Yo'Rec, a simple baseline that improves open-source MLLMs on PersonaRec. We release code at `https://anonymous.4open.science/r/LIP-Rec`.

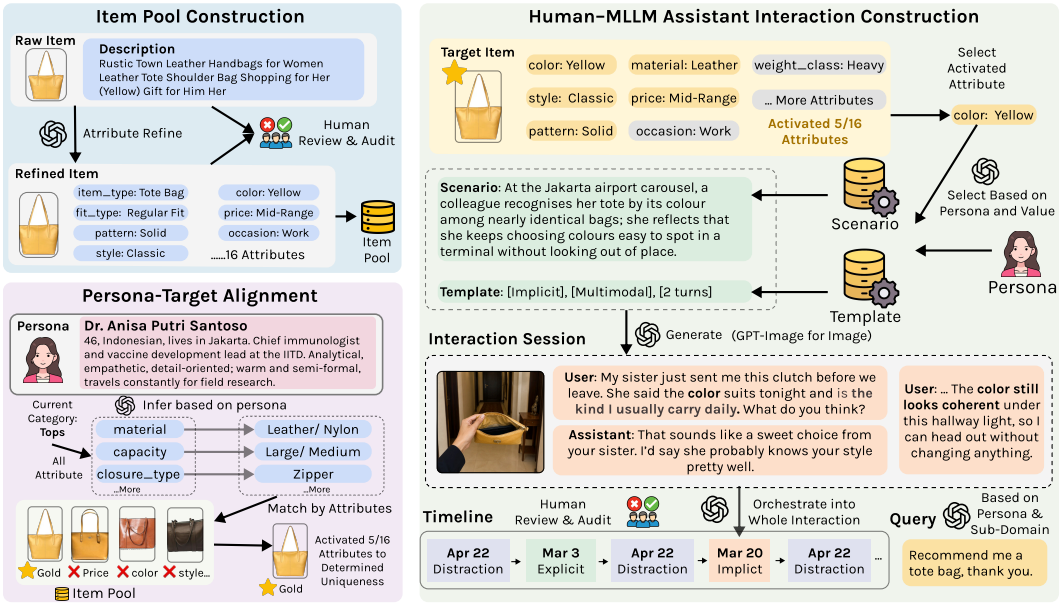


Figure 2: Construction workflow of our proposed PersonaRec benchmark.

2 PERSONAREC BENCHMARK

In this section, we formulate our proposed task LIP-Rec, and detail the construction of PersonaRec, followed by observations on the performance and behaviors of frontier MLLMs on PersonaRec.

2.1 FORMULATION OF LIP-REC

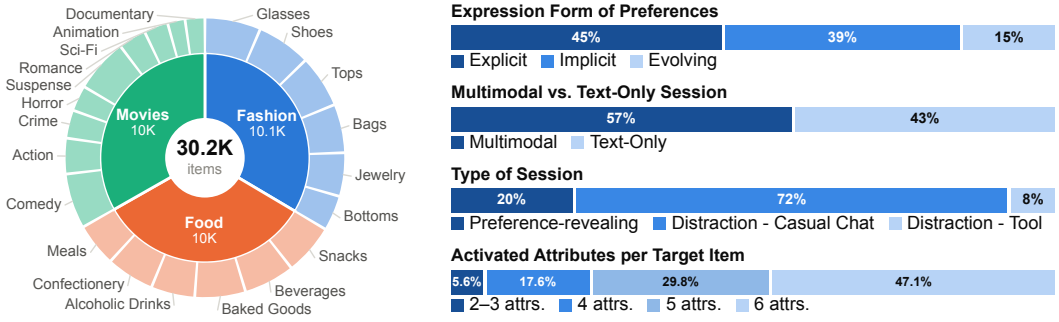
With the rise of personal MLLM assistants, we expect them to identify the items that best match our needs from simple and *personalized recommendation queries* (e.g., "Recommend a T-shirt that best fits me"), by drawing on our preferences embedded in daily historical interactions rather than requiring us to repeatedly clarify every detail. To fill in this gap, we propose the task of LIP-Rec. Formally, we consider a persona U (e.g., a biker with distinctive preferences) and long-term history of daily interactions (like chit-chat or tool use) with their personal MLLM assistant before time step T: $\boldsymbol{H}_U^T = \{H_1, H_2, \cdots, H_T\}$, where each H_i represents an everyday interaction session between U and the assistant and may contain textual and visual content. At time step $T+1$, persona U issues an underspecified recommendation query Q_U to their personal MLLM assistant, like a casual request in everyday use. LIP-Rec **expects** the assistant to recommend item for U by leveraging Q_U and preferences sparsely revealed from $\boldsymbol{H}_U^T$ that are relevant to the queried product or content:

$$f(\boldsymbol{H}_U^T, Q_U, \boldsymbol{I}) \rightarrow \hat{I}_U, \quad I_U^* \in \boldsymbol{I}, \tag{1}$$

where f is the **specific recommendation mechanism** employed by the personal MLLM assistant (or the assistant itself), $\boldsymbol{I}$ is the candidate item pool, $\hat{I}_U$ is the predicted item. I_U^* is the ground-truth target item, characterized by a set of distinctive attributes $\{A_1, A_2, \cdots, A_M\}$, among which a subset corresponds to the item preferences of U occasionally revealed from $\boldsymbol{H}_U^T$. Based on this, we construct PersonaRec, the first benchmark tailored for LIP-Rec, covering all specified components.

2.2 PERSONAREC BENCHMARK CONSTRUCTION

Construction Overview. Since LIP-Rec concerns personalized recommendation grounded in long-term daily interactions with MLLM assistants, ❶ we begin the construction of PersonaRec by building an item pool $\boldsymbol{I}$ with high-quality, human-refined items. ❷ We then construct a set of personas $\boldsymbol{U} = \{U_1, U_2, \cdots, U_N\}$ with basic profiles and align each of them with appropriate target items from $\boldsymbol{I}$. ❸ Finally, we construct long-term daily interactions $\boldsymbol{H}_U$ between each persona and their personal MLLM assistant, from which target-item-relevant preferences are incidentally revealed. An overview of the construction process of our PersonaRec is illustrated in Figure 2.



(a) Domain Distribution in the Item Pool (b) Features of Interaction Histories

Figure 3: Core characteristics of PersonaRec: item pool distribution and interaction features.

Table 1: Core statistics of PersonaRec. History Length is total length of all persona interactions.

# Domain	# Item	# Attribute	# Persona	# Target	# Template	# Scenario	# Session	# History Length
3/21 (Sub)	30,182	421	200	1,200	44	17,760	35,452	16.71M

Recommendation Item Pool Construction. Since LIP-Rec studies recommendation requests arising from daily human-MLLM assistant interactions, we construct the recommendation space of PersonaRec around three common everyday domains: "Fashion, Food, and Movies". Specifically, for Fashion and Food, we select the original items from the `Amazon-Reviews-2023` (Hou et al., 2026), whereas for Movies, we select initial candidates from `MovieLens-32M` (Harper & Konstan, 2015). Subsequently, we leverage the proprietary LLM `GPT-5.6-Sol` (OpenAI, 2026a) to curate these original items, refining noisy raw metadata into several distinctive natural-language attributes for each item (e.g., color, material, spicy level), with human manual audit to ensure common and accuracy. As a result, we construct 30K high-quality items across three domains and 21 sub-domains, 421 discriminative attributes with 32,604 attribute values.

Persona and Target Item Alignment. We first sample 200 personas from `PersonaMem-v2` (Jiang et al., 2025), which augments diverse personas from `PersonaHub` (Ge et al., 2024) with basic demographic attributes (e.g., age, habits). Subsequently, we match each persona with 6 target items spanning on three domains. Specifically, within each domain, we assign two sub-domains to each persona. We then employ the LLM to infer potential item-attribute preferences from each persona's basic profile over the attributes of the assigned sub-domains, and use these preferences only for persona–item alignment. For personas whose basic profiles reveal few or no such preferences, we select items by simply ensuring that they do not conflict with the profile information. This is natural in real world, where many user preferences emerge through interactions, which is the focus of LIP-Rec. We also conduct human audit for the alignment.

Long-Term Human–MLLM Assistant Interaction Construction. Based on the personas U and their target items I_U^*, we leverage the LLM to retrospectively construct daily interactions between each persona and their personal MLLM assistant, ensuring that the target attributes are naturally revealed through the interactions. Unlike Tou et al. (2026), we do not directly expose the I_U^* to the LLM during interaction construction. Instead, each interaction session (one or more dialogue turns) *focuses only on* the preference for a specific attribute A of I_U^*, thereby alleviating overly contrived conversations and yielding more realistic interactions. Specifically, to enhance the diversity and real-world relevance of the dialogues, we introduce session templates and scenarios:

▷ **Session Templates**: Simple constraints on how preferences are expressed and evolve throughout a dialogue. We design 19 concise templates that mainly specify **(1)** how a product preference unfolds over time, such as being explicitly stated, implicitly revealed, or subsequently changed, and **(2)** whether the preference is revealed through text-only or multimodal interactions.

▷ **Session Scenarios**: Human–AI interaction situations and topics in which preferences may naturally arise. We first sample seed scenarios from *real-world* human-LLM interaction benchmarks (e.g., Xiao et al. (2026)). We then leverage the LLM to adapt them into refined ones in which preferences over item attributes can be naturally expressed, while maintaining the plausibility of the scenarios for everyday human-MLLM interactions. As a result, we obtain 17,760 diverse, human-reviewed scenarios, each suited to eliciting a small subset of attribute-level preferences.

Consequently, we leverage the LLM to construct an interaction session between U and their personal MLLM assistant to reveal the preference for attribute A in a *role-play manner* (Wang et al., 2025), conditioning on the persona and the most appropriate interaction scenario and dialogue templates. Notably, 57% of the interaction sessions are multimodal, with `GPT-Image-2` (OpenAI, 2026b) used to generate the images. In a subset of these sessions, we deliberately omit the target attribute value from the text, requiring the model to infer the preference from the visual content. Ultimately, for each target item I_U^* of persona U, we only activate a subset of its attributes to construct the corresponding interaction sessions. We then construct *distraction sessions* (e.g., casual or tool-use chats, may involve the three item domains and express user's dislikes) for each U and leverage the LLM to orchestrate them together with the preference-revealing sessions across all target items of U into a coherent history. The resulting history is human-audited to produce the final set of interactions $\boldsymbol{H}_U$, with around 83.5K length per persona. The core features of PersonaRec are in Figure 3 and Table 1.

Evaluation of LIP-Rec. During evaluation, each persona issues several recommendation requests in a natural daily-chat style, and PersonaRec evaluates whether the personal MLLM assistant can successfully recommend the corresponding target items under their full past interactions. Specifically, we propose **two evaluation settings** to better simulate real-world queries from users to their personal MLLM assistants for personalized recommendation: the Underspecific Query (**U-Query**) setting, where no explicit item preferences are specified in the query (i.e., users simply state what they want in **sub-domain level** [1] and expect their personal assistants to already know their preferences), and the Partially Specific Query (**PS-Query**) setting, where one or two item preferences are explicitly specified by the user (i.e., users mention only what they currently care about most and expect their assistants to fill in the rest.). Notably, simple queries are tailored to each persona by the LLM rather than being identical across users (e.g., different presentation manner or tone), even within same sub-domain. For evaluation, we mainly report the common metrics in recommender systems: Recall@1 (**R@1**), Recall@10 (**R@10**), and NDCG@10 (**N@10**). Since target-relevant preferences are explicitly grounded in the daily interaction history, PersonaRec enables **fine-grained evaluation** of whether models recover each personalized cue associated with the target item.

More details on benchmark construction, with human evaluations and audits of each stage, interaction quality, and the overall reliability and accuracy of PersonaRec, are in Appendices A and D.

2.3 Observation: Frontier MLLMs Struggle on PersonaRec

We benchmark frontier MLLMs on PersonaRec, including `GPT-5.6-Sol` (OpenAI, 2026a), `Gemini-3.8-Flash` (Google DeepMind, 2026), and `Qwen3.5-27B` (Qwen Team, 2026). We provide models with both the full interaction histories and user recommendation queries, and instruct them to formulate the final recommendation queries based on the past interactions. We then retrieve candidate items from the item pool using the formulated queries and further re-rank the top-20 retrieved items using the same model. We observe that all these models struggle with the capability of LIP-Rec (as reported in Table 2). Taking a further step, we conduct additional empirical evaluations on the intermediate stages leading to the final recommendation, including **(1)** the identification of query-relevant, personalized preferences on items and **(2)** the formation of recommendation queries (Detailed experimental designs and evaluations are provided in Appendix B.). As a result, as an initial exploration of PersonaRec, we shed light on several basic findings about the failure modes of current frontier MLLMs on LIP-Rec: ❶ Incomplete or spurious recovery of personal item preferences for user's simple and personalized recommendation queries, largely due to distraction from long interaction histories, further amplified by the extended context introduced by numerous images and complicated by preference signals distributed across different modalities. ❷ Suboptimal transformation of preferences into recommendation queries aligned with item information.

3 Methodology: Yo'Rec

To tackle the above limitations of existing frontier MLLMs on PersonaRec, we propose a simple baseline Yo'Rec, which improves the LIP-Rec capabilities on open-source MLLMs. Specifically, Yo'Rec first introduces a PRPL-CoT-based SFT stage, which initially equips MLLMs with the rea-

[1] For Fashion and Food, to better reflect real-world recommendation requests in daily chat, we further introduce several *mini-domains* under each sub-domain, e.g., *Jacket* is a mini-domain under the *Tops* sub-domain.

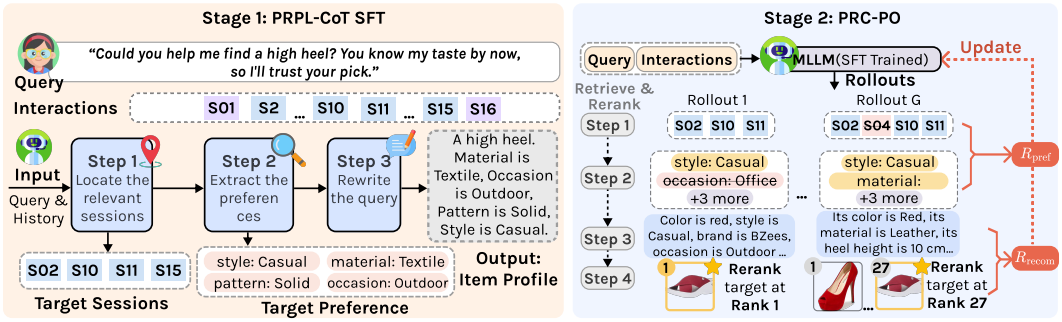


Figure 4: Training recipe of Yo'Rec, including a PRPL-CoT SFT stage that initializes the reasoning routines required for LIP-Rec, followed by a PRC-PO stage that further refines these capabilities.

soning routine to identify query-relevant, item-oriented preferences from long-term interactions and rewrite the original recommendation requests (Section 3.1). Subsequently, a PRC-PO stage incentivizes the MLLM to autonomously recover these preferences, while optimizing their transformation into effective queries for final recommendation (Section 3.2). Our method is presented in Figure 4.

3.1 Personalized Recommendation Preference Location CoT Based SFT

Since users often reveal item preferences only occasionally during everyday interactions with their personal MLLM assistants, such preferences are sparsely scattered across abundant histories. Therefore, we first propose a new Personalized Recommendation Preference Location CoT (**PRPL-CoT**) based SFT stage. It initializes MLLMs the **reasoning routine for LIP-Rec**: identifying requested-item-preference from the full histories and forming the final enriched quires for recommendation.

Specifically, we first construct an additional set of personas for the SFT stage, where each persona U is equipped with the interaction histories and target items. For simplicity, we detail the SFT data construction for a single target item I_U^* of U, which corresponds to one SFT training instance. Subsequently, based on the interaction history $\boldsymbol{H}_U$ between U and their assistant, PRPL-CoT SFT introduces a three-step reasoning process for the MLLMs: ❶ we first instruct the MLLMs to identify the subset $\boldsymbol{H}_U^* \subseteq \boldsymbol{H}_U$ that contains target-item-relevant preference signals, and ❷ then guide the MLLMs to extract the corresponding item preferences (attributes) $\boldsymbol{A}$ from $\boldsymbol{H}_U^*$ and ❸ rewrite the original recommendation query Q_U into a preference-enriched query $\widetilde{Q}_U$. This three-step chain design *reduces interference* of MLLMs from irrelevant long-term, multimodal daily interactions, enabling more effective recovery of relevant preferences. We derive the $\boldsymbol{H}_U^*$ and the oracle preferences $\boldsymbol{A}$ from the construction of personas and their targets. We further leverage the LLM to rewrite the Q_U into $\widetilde{Q}_U$ based on $\boldsymbol{A}$, completing the construction of a single SFT instance. After constructing instances for all targets across all personas, we obtain the SFT dataset $\boldsymbol{D}_{\text{SFT}}$, where each instance consists of an input $\boldsymbol{x}$ mainly containing Q_U and $\boldsymbol{H}_U$, and an output $\boldsymbol{y}$ containing $\boldsymbol{H}_U^*$, $\boldsymbol{A}$, and $\widetilde{Q}_U$.

3.2 Personalized Recommendation-query Completion Policy Optimization

After endowing MLLMs with the basic reasoning routine required for LIP-Rec, we further propose Personalized Recommendation-query Completion Policy Optimization (**PRC-PO**). Inspired by the GRPO (Shao et al., 2024), we expect MLLMs to autonomously identify scattered item preferences from long-term interactions, and improve recommendation query transformation through self-exploration and feedback from their own generated trajectories. Specifically, we re-utilize the SFT data $\boldsymbol{D}_{\text{SFT}}$ for the PRC-PO stage, which can be formulated as:

$$L_{\text{PRC-PO}} = -\mathbb{E}_{\boldsymbol{x}\sim\boldsymbol{D}_{\text{SFT}},\{o_i\}_{i=1}^{G}\sim\pi_{\theta_{\text{old}}}(\cdot|\boldsymbol{x})} \Big[\frac{1}{\sum_{i=1}^{G}|o_i|}\sum_{i=1}^{G}\sum_{t=1}^{|o_i|}\min(r_{i,t}(\theta)A_{i,t}, \text{clip}(r_{i,t}(\theta), 1-\epsilon, 1+\epsilon)A_{i,t}) - \eta D_{\text{KL}}[\pi_\theta\|\pi_{\theta_{\text{ref}}}]\Big]. \tag{2}$$

where π_θ is the policy of the MLLM. The probability ratio $r_{i,t}(\theta)$ and advantage $A_{i,t}$ are defined as:

$$r_{i,t}(\theta) = \frac{\pi_\theta(o_{i,t} \mid \boldsymbol{x}, o_{i,<t})}{\pi_{\theta_{\text{old}}}(o_{i,t} \mid \boldsymbol{x}, o_{i,<t})}, \quad A_{i,t} = \frac{R_i - \text{mean}(\{R_1,\ldots,R_G\})}{\text{std}(\{R_1,\ldots,R_G\})}. \tag{3}$$

Table 2: Performance on PersonaRec across three domains. Best performance is in **bold**.

U-Query	Fashion			Food			Movie			AVG		
Method	**R@1**	**R@10**	**N@10**	**R@1**	**R@10**	**N@10**	**R@1**	**R@10**	**N@10**	**R@1**	**R@10**	**N@10**
Query-Only	1.89	10.09	5.20	0.78	6.20	2.80	0.75	4.25	2.18	1.14	6.85	3.39
GPT-5.6-Sol	8.20	32.18	19.82	8.53	39.92	22.83	2.00	8.25	4.85	6.24	26.78	15.83
Gemini-3.8-Flash	7.62	26.94	16.21	11.12	27.67	18.39	2.28	10.58	6.06	7.01	21.73	13.55
Qwen3.5-27B	6.47	26.81	15.28	10.08	24.81	16.35	2.00	7.00	4.02	6.18	19.54	11.88
Qwen3.5-9B	5.05	19.72	11.24	5.43	17.83	11.11	1.62	5.00	3.03	4.03	14.18	8.46
+ RAG	6.62	20.35	12.52	9.69	25.58	16.27	0.75	3.25	1.85	5.69	16.39	10.21
+ Memory	3.63	14.83	8.54	6.59	23.64	14.18	0.38	3.62	1.84	3.53	14.03	8.19
+ Rerank	7.89	21.61	14.07	8.91	20.16	13.85	2.00	4.75	3.25	6.27	15.51	10.39
+ Yo'Rec	**13.41**	**27.41**	**20.24**	**19.63**	**32.22**	**25.97**	**7.50**	**13.22**	**10.17**	**13.51**	**24.28**	**18.79**
PS-Query	**Fashion**			**Food**			**Movie**			**AVG**		
Method	**R@1**	**R@10**	**N@10**	**R@1**	**R@10**	**N@10**	**R@1**	**R@10**	**N@10**	**R@1**	**R@10**	**N@10**
Query-Only	3.79	19.09	10.34	3.10	12.02	7.21	1.75	6.25	3.60	2.88	12.45	7.05
GPT-5.6-Sol	12.15	38.64	24.03	14.34	41.47	26.74	3.88	12.75	7.76	10.12	30.95	19.51
Gemini-3.8-Flash	9.60	30.37	19.10	10.04	33.37	21.01	3.85	11.82	7.64	7.83	25.19	15.92
Qwen3.5-27B	9.31	30.44	18.48	7.75	23.26	14.68	2.00	10.25	5.73	6.35	21.32	12.96
Qwen3.5-9B	6.31	27.76	15.73	7.36	23.26	14.65	0.88	6.12	3.20	4.85	19.05	11.19
+ RAG	9.31	27.13	17.26	12.40	28.29	19.76	1.00	5.38	2.87	7.57	20.27	13.30
+ Rerank	12.46	**30.91**	21.36	11.24	24.42	18.05	2.88	9.50	5.73	8.86	21.61	15.05
+ Memory	5.99	21.61	12.91	7.75	29.07	16.98	1.12	6.12	3.44	4.95	18.93	11.11
+ Yo'Rec	**15.78**	29.28	**22.95**	**18.65**	**30.04**	**24.23**	**10.31**	**14.44**	**12.51**	**14.91**	**24.59**	**19.90**

The final reward is mainly designed from two dimensions. The first reward (R_{pref}) incentivize the model to identify the relevant interaction subset $\boldsymbol{H}_U^*$ and recover the corresponding item preferences $\boldsymbol{A}$, all at the fine-grained per-preference level. We implement R_{pref} using F1-Score. Furthermore, to improve the transformation of the recovered item preferences into effective recommendation queries, we follow Huang et al. (2026a) and directly employ NDCG@K as the recommendation reward (R_{recom}), as it provides rank-sensitive feedback on how well the $\widetilde{Q}_U$ retrieves the target. R_{recom} is computed over the re-ranked Top-K items retrieved by $\widetilde{Q}_U$, with re-ranking performed by the MLLM itself. The algorithm of Yo'Rec is in Algorithm 1, and all the prompts are in Appendix K.

4 Experiments

4.1 Experimental Setup

Baselines. We first introduce a simple lower-bound baseline, **Query-Only**, which directly uses the original recommendation query for item retrieval and recommendation without any interaction history. Thereafter, as mentioned in Section 2.3, we benchmark frontier MLLMs on PersonaRec, including the proprietary `GPT-5.6-Sol` and `Gemini-3.8-Flash`, as well as the open-source `Qwen3.5-27B`. In addition, we evaluate the moderate-capacity open-source MLLM `Qwen3.5-9B` (Qwen Team, 2026) and construct several baselines on top of it: **(1) RAG** (Lewis et al., 2020): we follow the implementation in Nie et al. (2026) by leveraging the original recommendation query to retrieve Top-K interaction sessions and provide these sessions only to the MLLM; **(2) Memory**: we adopt the implementation in SeCom (Pan et al., 2025); and **(3) Rerank**: we additionally apply a reranking process using the model itself after the retrieval. Finally, we apply our Yo'Rec to Qwen3.5-9B and report its performance as our method.

Evaluation Setup. Following the definition in Section 2.1, we provide each method (except Query-Only) with both the original incomplete user query and the full daily interaction history, and instruct it to formulate the final recommendation query. We then retrieve candidate items from the item pool using the formulated query, optionally re-rank the Top-K retrieved items, and compute the final recommendation performance using R@1, R@10, and N@10 metrics.

Implementation Details. We use `Qwen3-VL-Embedding-2B` (Li et al., 2026b) for item retrieval and evaluate BM25 as an alternative in Appendix G.1. We train Yo'Rec with LoRA on 10 additional personas disjoint from the evaluation ones. All training and inference run on NVIDIA L40S GPU.

Table 3: Ablation study on core training components within our proposed Yo'Rec.

U-Query	Fashion			Food			Movie			AVG		
Method	**R@1**	**R@10**	**N@10**	**R@1**	**R@10**	**N@10**	**R@1**	**R@10**	**N@10**	**R@1**	**R@10**	**N@10**
Qwen3.5-9B	5.05	19.72	11.24	5.43	17.83	11.11	1.62	5.00	3.03	4.03	14.18	8.46
+ PRPL-CoT SFT	13.35	26.60	19.22	15.70	29.78	22.90	5.30	10.84	7.92	11.45	22.41	16.68
+ RL R_{pref}	13.30	26.87	19.78	17.32	31.21	24.94	6.10	11.59	8.79	12.24	23.22	17.84
+ RL R_{recom} (**Yo'Rec**)	**13.41**	**27.41**	**20.24**	**19.63**	**32.22**	**25.97**	**7.50**	**13.22**	**10.17**	**13.51**	**24.28**	**18.79**

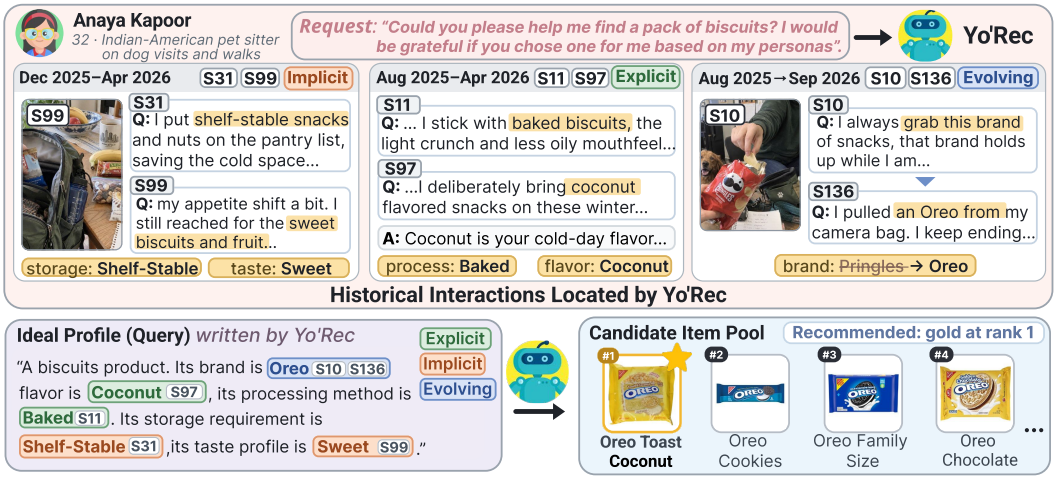


Figure 5: Case study: Yo'Rec identifies the latent user's preferences on snack biscuit.

4.2 Main Performance on PersonaRec

In Table 2, we report the performance of both frontier MLLMs and several long-context baselines built on a moderate-sized MLLM. Firstly, most baselines perform clearly better than Query-Only, indicating that leveraging long-term interaction histories can substantially improve performance on LIP-Rec and supporting the validity of the benchmark design. Besides, most baselines perform better under PS-Query, as it explicitly provides a small amount of immediate preference information and is relatively easier. For frontier MLLMs, they still struggle on PersonaRec, as analyzed in Section 2.3, with more in-depth empirical evaluations provided in Appendix B. For controlled analysis on the 9B MLLM, we observe that: **(1) Rerank** improves over the base model, suggesting that explicitly re-assessing the retrieved candidates can help distinguish among fine-grained items under LIP-Rec. **(2) RAG** and **Memory** provide only limited gains (or even degraded) as: **(i)** user's daily recommendation requests are often highly personalized and context-dependent (e.g., "a shirt that fits me"), for which RAG's semantic retriever struggle to surface the item-preference-relevant interactions. **(ii)** For Memory, which also relies on semantic retrieval over the history, pre-condensed memories may omit fine-grained preference cues that become relevant to a later recommendation request. **Yo'Rec** turns the base MLLM into a more intelligent item preference localizer, which identifies relevant raw histories based on personalized user recommendation requests, while dedicated training improves their translation into retrieval-effective queries, achieving initial better results.

4.3 Ablation Study of Yo'Rec

To evaluate the core training components in Yo'Rec, we conduct an ablation study. The first variant (**V1**) applies only PRPL-CoT SFT. Building upon V1, **V2** further introduces R_{pref}, while **V3** additionally incorporates R_{recom}, corresponding to the full Yo'Rec. As reported in Table 3, we observe that V1 achieves a clear improvement over the base MLLM, showcasing the effectiveness of equipping the model with the reasoning routine required for LIP-Rec. V2 further improves upon V1 by explicitly optimizing preference identification, while V3 achieves additional gains by further shaping the model to translate the identified item preferences into effective recommendation queries.

4.4 Case Study: Recommend Suited Items From Long-Term Daily Interactions

In Figure 5, we provide a case study of how the proposed Yo'Rec enhances MLLMs for LIP-Rec.

Specifically, Yo'Rec first locates the relevant interaction sessions from the full history that reveal `Anaya Kapoor`'s latent snack preferences, particularly those related to biscuits. Based on these sessions, Yo'Rec further aggregates the identified preferences and transforms them into the final recommendation query (profile), ultimately recommending `Oreo Toasted Coconut` as the best-fitting item for the user. Additional cases of Yo'Rec are provided in Appendix I.

5 Related Work

5.1 Long-Term LLM and MLLM Personalization

Personalizing LLMs and MLLMs for long-term daily interaction is critical to build more capable and reliable human assistants (Wu et al., 2025; Huang et al., 2026c). A growing body of work has explored long-term LLM personalization through prompting-, memory-, and retrieval-based approaches (Zhang et al., 2026a) (e.g., RF-Mem (Zhang et al., 2026c), RMM (Tan et al., 2025)), with benchmarks primarily evaluating preference retrieval, profile following, and personalized reasoning for daily conversations and question answering (e.g., PrefEval (Zhao et al., 2025a), LongMemEval (Wu et al., 2025), PersonaMeM-v2 (Jiang et al., 2025)). With the rapid advancement of MLLMs, recent studies have also begun transforming general-purpose MLLMs into long-horizon personalized assistants (e.g., PersonaVLM (Nie et al., 2026), TAME (Hong et al., 2026)), where user preferences can be inferred from multimodal interactions beyond text alone (Li et al., 2026c; Oh et al., 2026). Nevertheless, these studies largely center on personalization in chats and question answering, leaving user-customized recommendation from long-term daily interactions, a natural and practical use case for personal assistants, underexplored.

5.2 LLM-based Recommendation

With strong reasoning and language understanding capabilities, LLMs are increasingly adopted in recommender systems for more intelligent recommendation (Lin et al., 2025b). Current research on LLM-based recommendation can be broadly divided into two groups (Lin et al., 2025a; Huang et al., 2026b). The *first group* mainly leverages LLMs either to model user preferences from historical behaviors on platforms (e.g., clicks, ratings, and reviews) (Zhang et al., 2026b; Xu et al., 2025; Ding et al., 2026) or to enrich item representations (e.g., augmenting item descriptions) for better preference matching (Sheng et al., 2025; Zhai et al., 2025; Kim et al., 2026b; Hou et al., 2026). The *second group* focuses on LLM-based conversational recommendation, where users actively interact with LLM agents-powered recommender chatbots on e-commerce platforms to elicit their current intents through dialogues (Zhu et al., 2026; Kim et al., 2026a; Raj et al., 2026; Li et al., 2026a; Yu et al., 2026). Nevertheless, all these studies primarily model user preferences from *platform-side interactions*. As MLLMs increasingly serve as personal daily assistants, users naturally reveal preference signals throughout their everyday multimodal interactions with them. Yet, how such assistants can leverage these past interactions to personalize and complete users' recommendation intents, without requiring repeated preference clarification, remains largely unexplored. **To bridge this gap**, we propose PersonaRec, a timely and tailored benchmark that evaluates the recommendation capabilities of personalized MLLM assistants grounded in users' *assistant-side* long-term daily interactions. While Shopping Companion (Tou et al., 2026) studies product preference identification from multi-session dialogues, it reduces the task to retrieving isolated, target-product-conditioned attributes planted into an existing textual LLM conversation benchmark, rather than modeling coherent real-world personas with consistent and interrelated preferences across interactions.

6 Conclusion

In this study, we introduce LIP-Rec, a new research task, together with a dedicated human-evaluated benchmark, PersonaRec. We benchmark frontier MLLMs on PersonaRec and find that even advanced models struggle to effectively leverage long-term human–MLLM assistant interactions for personalized recommendation. We further propose Yo'Rec, which improves open-source MLLMs on PersonaRec, while extensive ablations and qualitative analyses validate the effectiveness of its designs. Future work calls for more advanced yet resource-efficient solutions to PersonaRec.

AI USE STATEMENT

In this work, we used generative AI tools to aid and polish the writing and to generate synthetic datasets. For writing assistance, generative AI was used primarily to improve the clarity, fluency, and presentation of author-written text. For synthetic data generation, generative AI was used to construct the long-term human–MLLM interaction histories in PersonaRec based on predefined personas, scenarios, templates, and preference specifications. These interactions were subsequently subjected to extensive human auditing and refinement, as described in the paper. We have not used generative AI tools for other tasks requiring disclosure, such as retrieval and discovery of related work, research ideation or execution, drafting scientific content from scratch, or proving mathematical claims. The remaining disclosure categories are not applicable to this work. We have reviewed all AI-assisted work and manually checked the resulting text and synthetic data for correctness, consistency, and quality. We take responsibility for the final content of this work.

ETHICS STATEMENT

This work studies personalized recommendation from long-term human–MLLM interactions, with the goal of advancing more capable and user-aware recommendation systems. The study poses minimal risk and does not involve interventions on real users or the collection of private user data. All human evaluation and auditing in this work (volunteers) were conducted exclusively by the authors, with no external participants involved. All participating authors were fully informed of the purpose and procedures of the study and participated voluntarily. Accordingly, no external recruitment or participant compensation was involved. We did not collect any personally identifiable information, private conversation records, demographic information, voice recordings, images, or other personal data from the evaluators.

REPRODUCIBILITY STATEMENT

We make extensive efforts to support the reproducibility of this work. First, we provide a detailed description of the construction pipeline of PersonaRec in Section 2, covering nearly all major stages of benchmark creation, with additional construction details, statistics, and human validation provided in Appendices A and D and figs. 7 to 11. Second, we describe the design of our baseline method Yo'Rec in Section 3 and provide further implementation details, including training configurations and baseline settings, in Appendix F. Third, we provide the evaluation code for PersonaRec together with the implementation of Yo'Rec to facilitate reproduction of our experimental results. Due to the large scale of PersonaRec, particularly its extensive multimodal interaction images and large candidate item pool, releasing the complete benchmark within the submission package is impractical. We therefore plan to publicly release the full benchmark upon publication to facilitate future research and comparison on LIP-Rec.

REFERENCES

Yuval Alaluf, Elad Richardson, Sergey Tulyakov, Kfir Aberman, and Daniel Cohen-Or. Myvlm: Personalizing vlms for user-specific queries. In *European Conference on Computer Vision (ECCV)*, pp. 73–91. Springer, 2024.

Zhihao Ding, Jinming Li, Shuai Mu, and Jieming Shi. Token-efficient long-term interest sketching and internalized reasoning for LLM-based recommendation. In *International Conference on Learning Representations (ICLR)*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/c9b551a2e195a209fc0b280de2f7f781-Abstract-Conference.html.

Tao Ge, Xin Chan, Xiaoyang Wang, Dian Yu, Haitao Mi, and Dong Yu. Scaling synthetic data creation with 1,000,000,000 personas. *arXiv preprint arXiv:2406.20094*, 2024.

Google DeepMind. Gemini 3.8 Flash model card, 2026. URL https://deepmind.google/models/model-cards/gemini-3-8-flash/.

F Maxwell Harper and Joseph A Konstan. The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)*, 5(4):1–19, 2015.

Rongpei Hong, Jian Lang, Ting Zhong, Yong Wang, and Fan Zhou. Tameing long contexts in personalization: Towards training-free and state-aware mllm personalized assistant. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 452–463, 2026. doi: 10.1145/3770854.3780214.

Yupeng Hou, Jiacheng Li, Xiangjun Fu, Zhankui He, An Yan, Xiusi Chen, and Julian McAuley. Bridging language and items for retrieval and recommendation: Benchmarking LLMs as semantic encoders. In *Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 3251–3265, 2026. doi: 10.18653/v1/2026.acl-long.147. URL https://aclanthology.org/2026.acl-long.147/.

Jiani Huang, Shijie Wang, Liang-bo Ning, Wenqi Fan, and Qing Li. Rerec: Reasoning-Augmented LLM-based Recommendation Assistant via Reinforcement Fine-tuning. In *Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 21040–21055, 2026a.

Jiani Huang, Shijie Wang, Liangbo Ning, Wenqi Fan, Shuaiqiang Wang, Dawei Yin, and Qing Li. Towards Next-Generation Recommender Systems: A Benchmark for Personalized Recommendation Assistant with LLMs. In *ACM International Conference on Web Search and Data Mining (WSDM)*, pp. 217–226. ACM, 2026b.

Zisu Huang, Muzhao Tian, Xiaohua Wang, Jingwen Xu, Zhengkang Guo, Qi Qian, Kaitao Song, Jiakang Yuan, Changze Lv, and Xiaoqing Zheng. Controllable memory usage: Balancing anchoring and innovation in long-term human–agent interaction. In *Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 14699–14719, 2026c.

Bowen Jiang, Yuan Yuan, Maohao Shen, Zhuoqun Hao, Zhangchen Xu, Zichen Chen, Ziyi Liu, Anvesh Rao Vijjini, Jiashu He, Hanchao Yu, Radha Poovendran, Gregory Wornell, Lyle Ungar, Dan Roth, Sihao Chen, and Camillo Jose Taylor. Personamem-v2: Towards Personalized Intelligence via Learning Implicit User Personas and Agentic Memory. *arXiv.org*, abs/2512.06688, 2025.

Chuanyang Jin, Binze Li, Haopeng Xie, Cathy Mengying Fang, Tianjian Li, Shayne Longpre, Hongxiang Gu, Maximillian Chen, and Tianmin Shu. Thoughttrace: Understanding user thoughts in real-world llm interactions. *arXiv preprint arXiv:2605.20087*, 2026.

Jiho Kim, Junseong Choi, Woosog Chay, Daeun Kyung, Yeonsu Kwon, Yohan Jo, and Edward Choi. ProPerSim: Developing proactive and personalized AI assistants through user-assistant simulation. In *International Conference on Learning Representations (ICLR)*, 2026a. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/b3bdc41d0b2387eabbb719956a38f295-Abstract-Conference.html.

Kibum Kim, Sein Kim, HongSeok Kang, Jiwan Kim, Heewoong Noh, Yeonjun In, Kanghoon Yoon, Jinoh Oh, Julian McAuley, and Chanyoung Park. Token-efficient item representation via images for LLM recommender systems. In *International Conference on Learning Representations (ICLR)*, 2026b. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/951690f8dae19c76394404540b7ca734-Abstract-Conference.html.

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. Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Advances in Neural Information Processing Systems*, volume 33, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html.

Guanrong Li, Kuo Tian, Jinnan Qi, Qinghan Fu, Zhen Wu, Rui Xia, and Xinyu Dai. LatentCRS: A variational EM framework for bridging semantics and behavior in LLM-based conversational recommendation. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2026a. doi: 10.1145/3770855.3818202. URL https://doi.org/10.1145/3770855.3818202.

Mingxin Li, Yanzhao Zhang, Dingkun Long, Keqin Chen, Sibo Song, Shuai Bai, Zhibo Yang, Pengjun Xie, An Yang, Dayiheng Liu, Jingren Zhou, and Junyang Lin. Qwen3-VL-Embedding

and Qwen3-VL-Reranker: A unified framework for state-of-the-art multimodal retrieval and ranking. *arXiv preprint arXiv:2601.04720*, 2026b. URL https://arxiv.org/abs/2601.04720.

Zhuoqun Li, Zhaopei Huang, Wenxuan Wang, and Qin Jin. Longmp-bench: A benchmark for multimodal persona understanding in long-term dialogues. In *Findings of the Association for Computational Linguistics (ACL)*, pp. 23132–23160, 2026c.

Jianghao Lin, Xinyi Dai, Yunjia Xi, Weiwen Liu, Bo Chen, Hao Zhang, Yong Liu, Chuhan Wu, Xiangyang Li, Chenxu Zhu, et al. How can recommender systems benefit from large language models: A survey. *ACM Transactions on Information Systems*, 43(2):1–47, 2025a.

Xinyu Lin, Chaoqun Yang, Wenjie Wang, Yongqi Li, Cunxiao Du, Fuli Feng, See-Kiong Ng, and Tat-Seng Chua. Efficient inference for large language model-based generative recommendation. In *International Conference on Learning Representations (ICLR)*, 2025b. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/e4bf5c3245fd92a4554a16af9803b757-Abstract-Conference.html.

Chang Nie, Chaoyou Fu, Yifan Zhang, Haihua Yang, and Caifeng Shan. Personavlm: Long-term personalized multimodal llms. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*, pp. 15000–15009, 2026. URL http://arxiv.org/abs/2604.13074.

Yeongtak Oh, Dongwook Lee, Sangkwon Park, Heeseung Kim, and Sungroh Yoon. Omnipersona: Systematic benchmarking and improving omnimodal personalization. *arXiv preprint arXiv:2605.09996*, 2026.

OpenAI. Introducing shopping research in chatgpt. https://openai.com/index/chatgpt-shopping-research/, 2025.

OpenAI. Previewing GPT-5.6 Sol: A next-generation model, June 2026a. URL https://openai.com/index/previewing-gpt-5-6-sol/.

OpenAI. GPT-Image-2 model, 2026b. URL https://developers.openai.com/api/docs/models/gpt-image-2.

Zhuoshi Pan, Qianhui Wu, Huiqiang Jiang, Xufang Luo, Hao Cheng, Dongsheng Li, Yuqing Yang, Chin-Yew Lin, H Vicky Zhao, Lili Qiu, et al. On memory construction and retrieval for personalized conversational agents. *International Conference on Learning Representations (ICLR)*, 2025.

Perplexity. Shopping that puts you first. https://www.perplexity.ai/hub/blog/shopping-that-puts-you-first, November 2025.

Qwen Team. Qwen3.5: Towards native multimodal agents, February 2026. URL https://qwen.ai/blog?id=qwen3.5.

Subham Raj, Aman Vaibhav Jha, Mayank Anand, and Sriparna Saha. HARPO: Hierarchical agentic reasoning for user-aligned conversational recommendation. In *Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 35580–35599, 2026. doi: 10.18653/v1/2026.acl-long.1646. URL https://aclanthology.org/2026.acl-long.1646/.

Stephen Robertson and Hugo Zaragoza. The probabilistic relevance framework: BM25 and beyond. *Foundations and Trends in Information Retrieval*, 2009. doi: 10.1561/1500000019. URL https://doi.org/10.1561/1500000019.

Zhihong Shao, Peiyi Wang, Qihao Zhu, R. Xu, Jun-Mei Song, Mingchuan Zhang, Y. K. Li, Yu Wu, and Daya Guo. Deepseekmath: Pushing the Limits of Mathematical Reasoning in Open Language Models. *arXiv.org*, abs/2402.03300, 2024.

Leheng Sheng, An Zhang, Yi Zhang, Yuxin Chen, Xiang Wang, and Tat-Seng Chua. Language representations can be what recommenders need: Findings and potentials. In *International Conference on Learning Representations (ICLR)*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/e4bab1843c8d5a69f5abfd0824593493-Abstract-Conference.html.

Zhen Tan, Jun Yan, I-Hung Hsu, Rujun Han, Zifeng Wang, Long Le, Yiwen Song, Yanfei Chen, Hamid Palangi, George Lee, Anand Rajan Iyer, Tianlong Chen, Huan Liu, Chen-Yu Lee, and Tomas Pfister. In prospect and retrospect: Reflective memory management for long-term personalized dialogue agents. In *Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 8416–8439, 2025. doi: 10.18653/v1/2025.acl-long.413.

Huaixiao Tou, Ying Zeng, Yuemeng Li, Cong Ma, Muzhi Li, Minghao Li, Weijie Yuan, He Zhang, and Kai Jia. Shoppingcomp: Are llms really ready for your shopping cart? *arXiv preprint arXiv:2511.22978*, 2026.

Kuang Wang, Xianfei Li, Shenghao Yang, Li Zhou, Feng Jiang, and Haizhou Li. Know you first and be you better: Modeling human-like user simulators via implicit profiles. In *Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 21082–21107, 2025.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, brian ichter, Fei Xia, Ed H. Chi, Quoc V Le, and Denny Zhou. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

Di Wu, Hongwei Wang, Wenhao Yu, Yuwei Zhang, Kai-Wei Chang, and Dong Yu. Longmemeval: Benchmarking chat assistants on long-term interactive memory. In *International Conference on Learning Representations (ICLR)*, 2025. URL https://openreview.net/forum?id=pZiyCaVuti.

Jianfei Xiao, Xiang Yu, Chengbing Wang, Wuqiang Zheng, Xinyu Lin, Kaining Liu, Hongxun Ding, Yang Zhang, Wenjie Wang, Fuli Feng, and Xiangnan He. AlpsBench: An LLM personalization benchmark for real-dialogue memorization and preference alignment. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, pp. 3489–3497, 2026.

Wujiang Xu, Qitian Wu, Zujie Liang, Jiaojiao Han, Xuying Ning, Yunxiao Shi, Wenfang Lin, and Yongfeng Zhang. SLMRec: Distilling large language models into small for sequential recommendation. In *International Conference on Learning Representations (ICLR)*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/58ab3d5ad04589543889118f80654279-Abstract-Conference.html.

Haocheng Yu, Yaxiong Wu, Hao Wang, Wei Guo, Yong Liu, Yawen Li, Yuyang Ye, Junping Du, and Enhong Chen. Thought-augmented planning for LLM-powered interactive recommender agent. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 1821–1832, 2026. doi: 10.1145/3770854.3780286. URL https://doi.org/10.1145/3770854.3780286.

Jianyang Zhai, Zi-Feng Mai, Chang-Dong Wang, Feidiao Yang, Xiawu Zheng, Hui Li, and Yonghong Tian. Multimodal quantitative language for generative recommendation. In *International Conference on Learning Representations (ICLR)*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/b9121bbb3112975d33c527f046ae68f2-Abstract-Conference.html.

Weizhi Zhang, Xinyang Zhang, Chenwei Zhang, Liangwei Yang, Jingbo Shang, Zhepei Wei, Henry Peng Zou, Zijie Huang, Zhengyang Wang, Yifan Gao, et al. Personaagent: Bridging memory and action for personalized llm agents. In *Findings of the Association for Computational Linguistics (ACL)*, pp. 26421–26439, 2026a.

Yang Zhang, Wenxin Xu, Xiaoyan Zhao, Wenjie Wang, Fuli Feng, Xiangnan He, and Tat-Seng Chua. Reinforced latent reasoning for LLM-based recommendation. In *International Conference on Learning Representations (ICLR)*, 2026b. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/d0c6bc641a56bebee9d985b937307367-Abstract-Conference.html.

Yingyi Zhang, Junyi Li, Wenlin Zhang, Pengyue Jia, Xianneng Li, Yichao Wang, Derong Xu, Yi Wen, Huifeng Guo, Yong Liu, and Xiangyu Zhao. Evoking user memory: Personalizing LLM via recollection-familiarity adaptive retrieval. In *International Conference on Learning Representations (ICLR)*, 2026c. URL https://openreview.net/forum?id=f7p0F2X6XN.

Siyan Zhao, Mingyi Hong, Yang Liu, Devamanyu Hazarika, and Kaixiang Lin. Do LLMs recognize your preferences? evaluating personalized preference following in LLMs. In *International Conference on Learning Representations (ICLR)*, 2025a. URL https://openreview.net/forum?id=QWunLKbBGF.

Zheng Zhao, Clara Vania, Subhradeep Kayal, Naila Khan, Shay B Cohen, and Emine Yilmaz. Personalens: A benchmark for personalization evaluation in conversational ai assistants. In *Findings of the Association for Computational Linguistics (ACL)*, pp. 18023–18055, 2025b.

Yaochen Zhu, Harald Steck, Dawen Liang, Yinhan He, Vito Ostuni, Jundong Li, and Nathan Kallus. Rank-GRPO: Training LLM-based conversational recommender systems with reinforcement learning. In *International Conference on Learning Representations (ICLR)*, 2026. URL https://openreview.net/forum?id=Xgw2D9cALS.

Technical Appendix

TABLE OF CONTENTS

A Additional Benchmark Construction Details **16**

B In-Depth Evaluation of Frontier MLLMs on PersonaRec **16**

B.1 Evaluation of Query-Relevant Item Preference Identification 17

B.2 Evaluation of Recommendation Query Transformation 17

C Additional Statistics of PersonaRec **18**

D Human Study on Benchmark and Evaluation **19**

D.1 Human Audit & Review in PersonaRec Construction 20

D.2 Human Post-Evaluation on PersonaRec 20

D.3 Human Performance on PersonaRec 21

E Algorithm of Yo'Rec Training Pipeline **21**

F Implementation Details **21**

G Additional Empirical Results **23**

G.1 Evaluation with an Alternative Retriever 23

G.2 Evaluation of Frontier MLLM Variants 23

G.3 Fine-Grained Performance on Sub-Domains 24

H Limitations and Future Research Directions on PersonaRec **24**

I Additional Cases of Yo'Rec and Baselines **26**

J Additional Presentations of PersonaRec **26**

K Prompt Presentation **27**

K.1 Prompts Used for PersonaRec Construction 28

K.2 Prompts Designed in Yo'Rec 29

K.3 Prompts Used for Baselines 29

A Additional Benchmark Construction Details

We now provide further details on the construction of PersonaRec, focusing on four aspects and questions: (1) how session scenarios enable stable and natural disclosure of users' everyday item preferences, (2) how we curate the candidate item pool from existing recommendation datasets, (3) how distraction sessions are constructed and the roles they play in the benchmark, and (4) how we select the activated attributes of each target item for preference revelation.

For (1), we first collect a large set of seed scenarios from existing real-world human–LLM interaction benchmarks (Xiao et al., 2026; Jin et al., 2026). For each item attribute (e.g., "color" for Tops), we ask the LLM to retain scenarios where preferences over that attribute could be naturally revealed, while covering diverse attribute values (e.g., red and black). The LLM then adapts the selected scenarios to better accommodate natural preference expression (e.g., explicit, implicit), after which our human volunteers manually verify and refine them. Consequently, we construct 17,760 scenarios at fine-grained attribute-level. Notably, we introduce two constraints to avoid unreliable preference transfer. First, when an item appears in a scenario, we only treat the revealed attribute preference as valid supervision, rather than inferring preferences over all attributes of that item. This improves the reliability of preference annotations while naturally introducing necessary noise, as models cannot freely extrapolate preferences from unrelated item attributes. Second, we restrict preference transfer to cases where such transfer is reasonably stable across contexts. For example, a color preference expressed for Tops is not directly transferred to Bags, whereas more context-independent preferences, such as those related to sports, can be shared across some categories.

For (2), we randomly sample candidate items from the `Amazon-Reviews-2023` (Hou et al., 2026) dataset and the `MovieLens-32M` (Harper & Konstan, 2015) dataset. Based on their original metadata, we manually refine the item attributes and corresponding values to remove noisy or inconsistent annotations, resulting in a pool of 30,182 items with verified attribute labels. Notably, we do not retrospectively modify the item pool to introduce hard negatives, as doing so could distort the original item distribution. Instead, the pool is constructed in a purely forward manner.

For (3), we construct distraction sessions from three main categories: **(i)** interactions entirely unrelated to the three recommendation domains, such as casual conversations and general tool-use requests, which simulate the large amount of irrelevant content naturally accumulated in long-term daily interactions, **(ii)** interactions involving the three recommendation domains but revealing no user preferences, which reflect the fact that users may frequently discuss relevant products or topics without expressing any personal preference, and **(iii)** interactions involving these domains in which users express dislikes toward certain item attributes or values, capturing the naturally mixed positive and negative preference signals in daily interactions. Together, (ii) and (iii) also reduce shortcut learning by preventing models from simply treating domain-related or preference-like attribute mentions as valid positive preferences. All distraction sessions are constructed using the same scenario- and template-based process as the preference-revealing interactions.

For (4), we activate only a small subset of attributes for each target item rather than exposing all through the interaction history. This reflects everyday interactions, where users rarely express every preference relevant to an item in daily interactions and instead reveal only a few scattered preferences over time. We follow three principles when selecting the activated attributes. **First**, whenever possible, we prioritize attributes whose preferred values are supported by the persona profile, although such information is often sparse or absent. **Second**, the activated preferences associated with different target items of the same persona must remain mutually consistent, avoiding contradictory preference signals across the long-term history. **Third**, the selected attributes must be sufficiently discriminative to uniquely identify the target item within the candidate pool. We therefore retain only attribute subsets for which the target remains the unique item satisfying all activated preferences.

B In-Depth Evaluation of Frontier MLLMs on PersonaRec

We conduct in-depth evaluations of frontier MLLMs on PersonaRec, focusing on two core intermediate steps for LIP-Rec: item preference identification and recommendation query transformation.

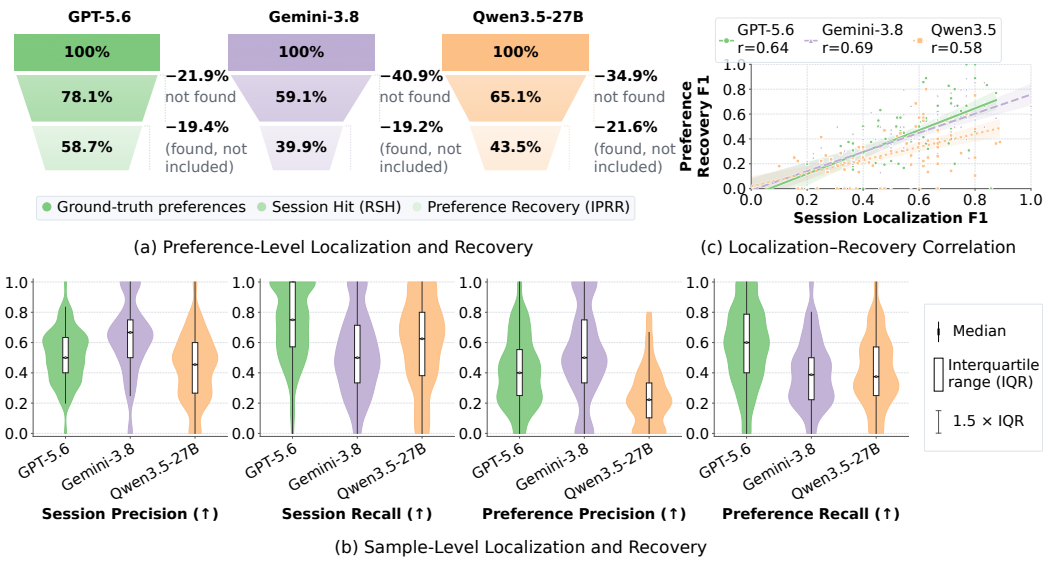


Figure 6: Fine-grained evaluation of frontier MLLMs on PersonaRec in item-preference location.

Table 4: Performance on PersonaRec when provided oracle item preferences of users.

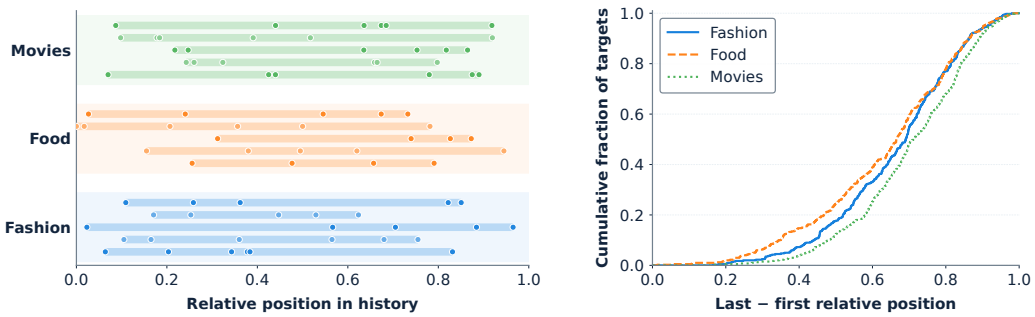
	Fashion			Food			Movie			AVG		
Method	R@1	R@10	N@10	R@1	R@10	N@10	R@1	R@10	N@10	R@1	R@10	N@10
GPT-5.6-Sol	46.69	84.54	64.69	31.78	67.44	49.50	8.25	29.25	17.77	28.91	60.41	43.99
Gemini-3.8-Flash	34.01	59.84	46.26	22.25	47.21	34.80	5.78	19.95	12.22	20.68	42.33	31.09
Qwen3.5-27B	47.95	85.80	65.73	31.78	67.44	48.87	8.00	29.00	17.43	29.24	60.75	44.01

B.1 Evaluation of Query-Relevant Item Preference Identification

Specifically, we evaluate item-preference identification at two levels. (1) **Preference-level analysis** traces each ground-truth preference from its supporting interaction sessions to its eventual recovery in the rewritten query. We report Relevant-Session Hit (**RSH**), which measures the proportion of preferences for which at least one supporting session is correctly localized, and Item-Preference Recovery Rate (**IPRR**), which measures the proportion of ground-truth preferences successfully recovered in the final query. (2) **Sample-level analysis** evaluates the overall identification quality for each recommendation instance. We report the precision and recall of relevant-session localization, measuring how accurately and completely the target-relevant sessions are identified, together with the precision and recall of item-preference recovery, measuring how accurately and completely the target-item preferences are reflected in the final query. We further analyze the **correlation** between session localization and preference recovery across samples, revealing how the quality of relevant sessions localization relates to later item-preference identification. As reported in Figure 6, we provide the results of the frontier MLLMs, including `GPT-5.6-Sol`, `Gemini-3.8-Flash`, and `Qwen3.5-27B`. As shown in Figure 6, frontier MLLMs lose substantial preference information throughout the identification process. A considerable fraction of ground-truth preferences are missed because their supporting sessions are not successfully localized, while additional preferences are lost even after the relevant evidence has been found. At the sample level, session-localization F1 exhibits a consistent positive correlation with preference-recovery F1 across all models.

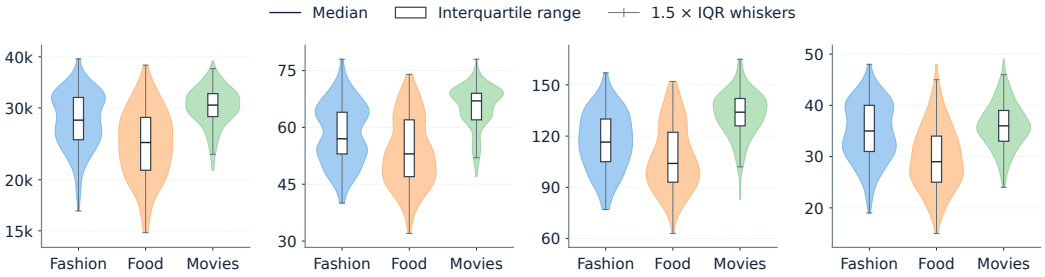
B.2 Evaluation of Recommendation Query Transformation

Beyond identifying item-relevant personalized preferences, LIP-Rec ultimately requires recommending the most appropriate items for individual users. We therefore conduct an additional evaluation of frontier MLLMs on recommendation query transformation, i.e., transferring the recovered user-specific preferences into effective queries for identifying the target item from the candidate pool. Specifically, to isolate this capability from preference identification, we directly provide each model with the **oracle item preferences of the user** together with the original incomplete recommendation query, and require it to formulate the final query and perform recommendation. Ideally, given access to oracle preferences, the recommendation success rate should approach the 100%.



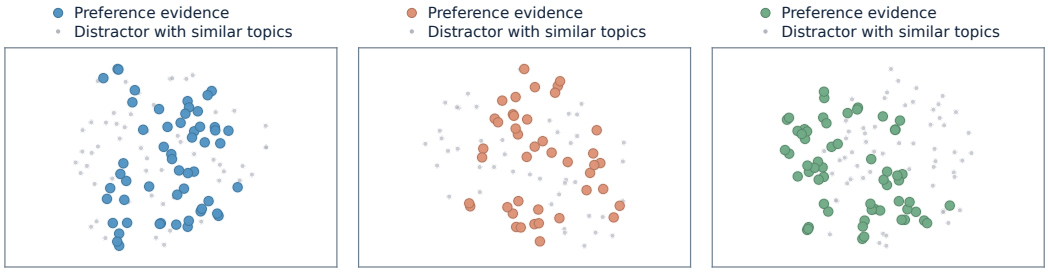
(a) Relevant session positions in user history (b) Relevant-session span distribution

Figure 7: Distribution of preference-revealing sessions across the whole long-term history.



(a) History tokens per persona (b) Sessions per persona (c) Turns per persona (d) Images per persona

Figure 8: Distribution of interaction-history statistics, including total history length, number of sessions, number of turns, and number of images across all personas.



(a) Fashion (b) Food (c) Movies

Figure 9: T-SNE visualization of the preference-revealing sessions and the distraction sessions.

Nevertheless, as reported in Table 4, these models still exhibit a substantial gap from the oracle performance on average, particularly in the Movie domain, highlighting the importance of effectively transforming recovered preferences into recommendation queries.

C Additional Statistics of PersonaRec

In this section, we provide additional statistics of our constructed PersonaRec. Specifically, we **first** examine where item-preference-revealing sessions occur within persona histories in Figure 7. In Figure 7(a), we present five example targets from each domain. Each row corresponds to one target, and each point marks the relative position of a session containing relevant preference evidence, with 0 denoting the beginning of the history and 1 the query boundary. The horizontal segment connects the earliest and latest relevant sessions. In Figure 7(b), we report the empirical cumulative distribution of this span across all target items in each domain (the x-axis value is the difference between the normalized positions of the latest and earliest relevant sessions). We observe that preference-revealing sessions are sparsely scattered throughout the interaction histories, mirroring real-world interactions in which item-relevant preferences surface only occasionally among abundant everyday conversations. We **then** report persona-level interaction-history statistics in Figure 8, including the total number of tokens, sessions, turns, and images. Each violin shows the distribution of one statistic within a domain, while the inner box and central line indicate the interquartile range and median,

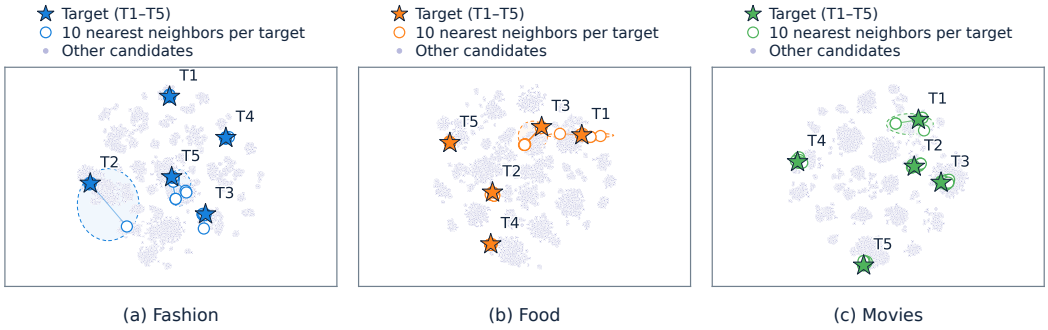


Figure 10: T-SNE visualization of five target items with neighbors and other items.

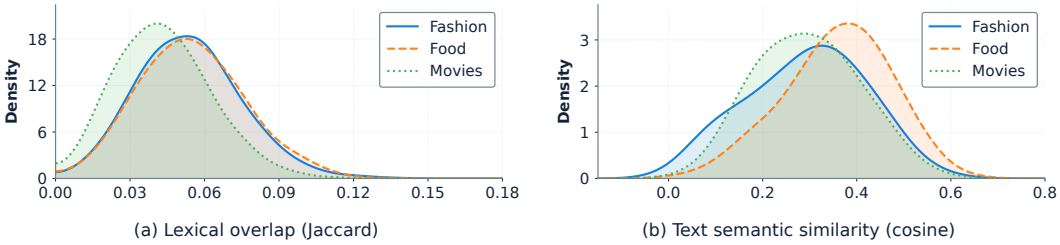


Figure 11: Overlap between user's personalized recommendation queries and preference sessions.

respectively. These distributions characterize the scale and variability of the histories that a model must process. History token counts are shown on a logarithmic axis.

Thereafter, we visualize 2D projections (T-SNE) of session representations separately for Fashion, Food, and Movies in Figure 9, using 5 persona's histories. Colored points denote sessions containing item-preference evidence relevant to the corresponding query, while gray points denote topically similar distractor sessions. This visualization examines whether preference-relevant evidence is readily separable from ordinary interactions in the representation space. We observe clear overlap between the two groups, indicating that preference-revealing sessions are not trivially distinguishable from semantically similar distractors. **Moreover**, we visualize the candidate-item space using five example targets per domain in Figure 10. Stars mark the targets, and outlined circles identify the ten nearest neighbors of each target, selected in the original feature space before projection. Each domain uses a shared projection for its candidate pool. We observe that the nearest-neighbor items are densely clustered around their corresponding targets, highlighting the fine-grained distinctions required to identify the intended item among the candidates. **Finally**, we examine lexical and semantic overlap between user's personalized yet incomplete recommendation queries and their relevant sessions in Figure 11. For each domain, we estimate the distribution of similarity scores over query–relevant-session pairs. The lexical view captures shared wording, whereas the semantic view captures similarity in meaning. Together, they characterize how directly a query resembles its supporting historical evidence. We observe that, at both the lexical and semantic levels, the original personalized recommendation queries exhibit only limited similarity to the sessions that reveal the corresponding user preferences. This is consistent with our experimental finding that semantic retrieval often fails to surface such preference-relevant interactions, as highly personalized and underspecified recommendation queries (e.g., "recommend a comedy film that fits me") provide only weak semantic cues for identifying the historical evidence needed to interpret them.

D Human Study on Benchmark and Evaluation

We detail the human studies conducted throughout the construction and evaluation of PersonaRec: Some human annotators participates in the construction stage to audit and refine the generated data toward more natural and realistic daily interactions, while others conducts post-hoc evaluations to validate the quality and reliability of the resulting benchmark. We further evaluate human performance on PersonaRec to provide a human reference for model performance.

Table 5: Human evaluation of the long-term human–MLLM interactions in PersonaRec.

Group 1					Group 2					Average				
N	**C**	**PC**	**PN**	**TI**	**N**	**C**	**PC**	**PN**	**TI**	**N**	**C**	**PC**	**PN**	**TI**
2.50	2.00	82.50	90.00	80.00	3.00	2.50	75.00	90.00	90.00	2.75	2.25	78.75	90.00	85.00

D.1 Human Audit & Review in PersonaRec Construction

To ensure the quality of PersonaRec throughout construction, we invite two independent groups of human volunteers, with several volunteers in each group, to audit the intermediate and final outputs at multiple stages. Each component is independently reviewed by both groups and is accepted only when both groups approve it. If one group identifies an issue, the corresponding component is revised and reviewed again. If both groups reject it, we conduct a substantial revision or reconstruction before another round of review. This procedure is applied throughout the following four stages. **Item Pool Audit.** We first ask the volunteers to audit the curated item pool. They review the refined attribute schema and the corresponding value spaces for each item category, checking whether the attributes are meaningful, discriminative, and correctly defined. Afterward, they further inspect individual items together with their textual and visual content to verify the correctness of the annotated attribute values and correct any remaining noisy or inconsistent annotations. This process ensures that both the attribute system and item-level annotations used for recommendation evaluation are reliable. **Persona–Target Item Audit.** For each target item assigned to a persona, the volunteers independently examine whether the item is compatible with the persona's basic information. The purpose of this stage is to ensure that the persona–item pairing is locally coherent and contains no obvious contradiction. Any target item judged to conflict with the persona information is replaced before subsequent interaction construction. **Scenario and Preference-Expression Audit.** We next audit the scenarios used to construct daily interactions. Both groups examine every scenario for its realism, clarity, and plausibility as an interaction that could naturally occur between a user and a personal MLLM assistant. For preference-revealing scenarios, the volunteers further verify that the intended preference can be naturally disclosed within the scenario, without requiring artificial or overly explicit disclosure. Scenarios that fail these criteria are revised until both groups reach agreement. **Long-Term Interaction Audit.** Finally, the volunteers review the complete interaction history constructed for each persona. Rather than inspecting individual sessions in isolation, they read through the full history to identify inappropriate, inconsistent, or erroneous content and to assess the overall coherence of the accumulated interactions. In particular, they check whether temporal dependencies, evolving preferences, recurring information, and causal relations remain consistent across the history. We do not require strict narrative continuity across all sessions, since real-world daily interactions accumulated over months or years are naturally fragmented and loosely connected. Instead, the audit focuses on preserving local plausibility, causal consistency, and the absence of contradictions that would undermine the intended user preferences or recommendation targets. Any detected issue is revised and re-reviewed under the same approval procedure.

D.2 Human Post-Evaluation on PersonaRec

To evaluate the quality of the resulting benchmark, we invite two groups of volunteers (each group has several volunteers). We follow prior human-evaluation practices in Li et al. (2026c); Zhao et al. (2025b) while adapting to the specific characteristics of PersonaRec. Accordingly, our post-evaluation on PersonaRec mainly focuses on the following three aspects.

Long-Term Human-MLLM Interaction Quality. We first evaluate the overall quality of the constructed human–MLLM interactions using two metrics, Naturalness (**N**) and Coherence (C). N measures whether the interactions resemble natural and realistic everyday communication between users and personal MLLM assistants, while C assesses whether each session is logically and contextually coherent. Each group is randomly assigned with 20 personas for the evaluation. Both metrics are rated on a **three-point Likert scale**, where a higher score indicates better interaction quality.

Item Preference Expression Quality. We next evaluate whether the target preferences are appropriately expressed in their annotated interaction sessions using two complementary metrics, Preference Naturalness (**PN**) and Preference Correctness (**PC**). For PN, volunteers make a binary `Yes/No`

Table 6: Performance on humans across three domains. AVG is the average across domains.

U-Query	Fashion	Food	Movie	AVG
Method	**Accuracy** (R@1)	**Accuracy** (R@1)	**Accuracy** (R@1)	**Accuracy** (R@1)
Group 1	75.00	75.00	37.50	62.50
Group 2	87.50	87.50	50.00	75.00
Average	81.25	81.25	43.75	68.75

judgment on whether the target preference is expressed naturally within the interaction rather than being artificially or explicitly inserted. For PC, volunteers are shown the corresponding interaction segment without its preference annotation and are asked to independently infer the user preference. Each group is randomly assigned with 100 sessions for the evaluation. Their inferred preference is then compared with the benchmark annotation, allowing us to verify whether the annotated preference can be correctly recovered from its designated evidence.

Target Item Identifiability. Finally, we conduct an end-to-end validation of whether the annotated item preferences are sufficient to identify the intended target item. Specifically, for each target item, we provide volunteers with its associated user preferences (oracle) together with a candidate set consisting of the target item and nine similar items, and ask them to identify the item that best matches the given preferences. Each group is randomly assigned with 100 target items for the evaluation. We report Target Identifiability (**TI**) as the resulting accuracy, which verifies that the designed preference set is sufficient to uniquely identify the target item.

As shown in Table 5, we report the results from both volunteer groups separately, together with their average performance across N, C, PC, PN, and TI. We observe consistently high scores on N, C, and PN, indicating that the long-term daily interactions in PersonaRec are natural and coherent, while the preference-revealing sessions express the intended user preferences in a plausible and non-contrived manner. Moreover, the high TI score shows that the annotated preference sets are sufficiently discriminative to identify the intended target item among highly similar candidates, further supporting the reliability of the target construction and preference annotations in PersonaRec.

D.3 Human Performance on PersonaRec

We evaluate human performance on PersonaRec to establish a human reference for interpreting model performance and to examine whether the benchmark remains solvable by humans despite its long and noisy interaction histories. We invite two groups of volunteers (each group has several volunteers) to complete a subset of the recommendation tasks in PersonaRec. Specifically, each volunteer is assigned the 10 randomly sampled personas (totally 60 target items) and performs the corresponding recommendation tasks based on the provided interaction histories and recommendation queries. Notably, we do not ask volunteers to search for the target item from the entire candidate pool, which is impractical for human evaluation. Instead, they select from a set of ten items consisting of the target item and nine similar candidates. We evaluate human performance using the same evaluation metrics as those used for the models. Table 6 reports the average human performance. We observe that human volunteers achieve higher accuracy on PersonaRec than the best-performing models. This comparison should be interpreted with caution, however, as humans select from only 10 highly similar candidate items rather than the full candidate pool used for model evaluation. Nevertheless, volunteers spend approximately 60 minutes per persona even under this reduced search space, underscoring that LIP-Rec requires substantial effort to locate, integrate, and reason over sparse personalized evidence accumulated throughout long-term interactions.

E Algorithm of Yo'Rec Training Pipeline

We provide the training algorithm of Yo'Rec in Algorithm 1, including the whole two stages.

F Implementation Details

In this section, we provide additional implementation details of our study, covering (1) the implementation of Yo'Rec, (2) the implementation of the baselines, and (3) the retriever configuration.

Algorithm 1 Training Yo'Rec with PRPL-CoT SFT and PRC-PO

Require: M personas with interaction histories $\boldsymbol{H}_U$ and target items, initial MLLM policy π_{θ_0}, group size G, clipping parameter ϵ, KL coefficient η, and retrieval cutoff K
Ensure: Trained MLLM policy π_θ

/* **Stage I: PRPL-CoT-Based Supervised Fine-Tuning** */
Initialize $\boldsymbol{D}_{\mathrm{SFT}} \leftarrow \varnothing$
for all personas U and target items I_U^* **do**
  Identify the target-item-relevant interaction subset $\boldsymbol{H}_U^* \subseteq \boldsymbol{H}_U$
  Extract the corresponding oracle item preferences $\boldsymbol{A}$ from $\boldsymbol{H}_U^*$
  Rewrite the original recommendation query Q_U into the preference-enriched query $\widetilde{Q}_U$ using $\boldsymbol{A}$
  Construct $\boldsymbol{x}_U = (Q_U, \boldsymbol{H}_U)$ and $\boldsymbol{y}_U = (\boldsymbol{H}_U^*, \boldsymbol{A}, \widetilde{Q}_U)$
  Add $(\boldsymbol{x}_U, \boldsymbol{y}_U)$ to $\boldsymbol{D}_{\mathrm{SFT}}$
end for
Optimize π_{θ_0} on $\boldsymbol{D}_{\mathrm{SFT}}$ with supervised fine-tuning and obtain $\pi_{\theta_{\mathrm{SFT}}}$
/* **Stage II: Personalized Recommendation-query Completion Policy Optimization** */
Initialize $\pi_\theta, \pi_{\theta_{\mathrm{old}}}, \pi_{\theta_{\mathrm{ref}}} \leftarrow \pi_{\theta_{\mathrm{SFT}}}$, with $\pi_{\theta_{\mathrm{ref}}}$ frozen
for step $= 1, \ldots, T$ **do**
  Sample a training batch $\mathcal{B} \subseteq \boldsymbol{D}_{\mathrm{SFT}}$
  for all $(\boldsymbol{x}_U, \boldsymbol{y}_U) \in \mathcal{B}$ **do**
    Sample G trajectories $\{o_i\}_{i=1}^G \sim \pi_{\theta_{\mathrm{old}}}(\cdot \mid \boldsymbol{x}_U)$
    for all $o_i \in \{o_1, \ldots, o_G\}$ **do**
      Recover $\widehat{\boldsymbol{H}}_{U,i}^*$, $\widehat{\boldsymbol{A}}_i$, and $\widehat{\widetilde{Q}}_{U,i}$ from o_i
      Compute the preference reward R_{pref}^i using fine-grained F1 between the recovered and oracle relevant interactions and item preferences
      Retrieve candidate items using $\widehat{\widetilde{Q}}_{U,i}$ and re-rank the Top-K items with the MLLM
      Compute the recommendation reward R_{recom}^i as NDCG@K on the re-ranked items with respect to I_U^*
      Compute the final reward R_i from R_{pref}^i and R_{recom}^i
    end for
    Compute group-relative advantages $A_{i,t}$ using Eq. (3)
  end for
  Update π_θ using the PRC-PO objective in Eq. (2)
  $\pi_{\theta_{\mathrm{old}}} \leftarrow \pi_\theta$
end for
return π_θ

Implementation of Yo'Rec. We build Yo'Rec on `Qwen3.5-9B` (Qwen Team, 2026) and train three separate models, each with its own set of LoRA weights shared across all three steps of PRPL-CoT. We apply LoRA with rank 32 and $\alpha = 64$ with dropout disabled, while keeping the vision encoder of `Qwen3.5-9B` frozen. For training, we additionally construct 10 personas that are strictly disjoint from the 200 personas used for evaluation. To prevent leakage from the evaluation domain, each model is trained only on the other two domains, excluding data from its evaluation domain during both SFT and PRC-PO. For example, the model evaluated on Food excludes the Food portion of all 10 training personas and is trained only on their Fashion and Movie data. The same exclusion protocol is applied to the models evaluated on Fashion and Movie. During the PRPL-CoT SFT stage, we train the model for 2 epochs with a learning rate of 1×10^{-4} using AdamW and a cosine learning-rate schedule with 3% warm-up. The resulting SFT model is then used to initialize the policy for PRC-PO. During PRC-PO, we sample $G = 8$ trajectories for each input and optimize the policy with a learning rate of 5×10^{-6} for 20 training rounds, with four target items processed in each round. The rollout policy is synchronized with the updated policy after every round. Group-relative advantages are normalized independently within trajectories corresponding to the same PRPL-CoT step. The retriever remains frozen throughout training and first retrieves the Top-20 candidate items, which are subsequently re-ranked by the policy itself based on the generated ideal profile. We use NDCG@K with $K = 100$ as the recommendation reward and assign equal weights to the preference-oriented

Table 7: Performance on PersonaRec across three domains using BM25 as retriever.

U-Query	Fashion			Food			Movie			AVG		
Method	R@1	R@10	N@10	R@1	R@10	N@10	R@1	R@10	N@10	R@1	R@10	N@10
GPT-5.6-Sol	4.57	19.09	11.29	15.12	41.47	26.05	0.62	3.12	1.65	6.77	21.23	13.00
Gemini-3.8-Flash	3.42	16.89	9.33	9.23	31.75	19.43	1.93	5.33	3.33	4.86	17.99	10.70
Qwen3.5-27B	2.21	11.36	6.20	5.81	23.64	13.74	0.50	2.00	1.08	2.84	12.33	7.01
PS-Query	Fashion			Food			Movie			AVG		
Method	R@1	R@10	N@10	R@1	R@10	N@10	R@1	R@10	N@10	R@1	R@10	N@10
GPT-5.6-Sol	7.10	25.08	14.68	16.28	48.45	30.47	1.62	5.00	3.02	8.33	26.18	16.06
Gemini-3.8-Flash	4.53	18.55	10.52	6.51	26.32	15.55	1.40	7.35	3.92	4.15	17.41	10.00
Qwen3.5-27B	3.31	12.46	7.01	6.20	25.19	14.09	0.25	3.50	1.59	3.25	13.72	7.56

and recommendation-oriented rewards. All training and inference experiments are conducted on two NVIDIA L40S GPUs with 48 GB of memory each.

Implementation of Baselines. For fair, all baselines follow the same recommendation pipeline, while differing in how historical interactions and retrieved candidates are utilized. **Query-Only** ignores the interaction history and directly retrieves items with the original Q_U. The frontier MLLMs and `Qwen3.5-9B` receive the complete interaction history $\boldsymbol{H}_U^T$ and formulate a personalized recommendation query accordingly. **RAG** retrieves the Top-5 interaction sessions most relevant to Q_U and provides them to the MLLM in chronological order. **Memory** constructs and retrieves segment-level memory units following SeCom (Pan et al., 2025). **Rerank** further asks the same MLLM to re-rank the Top-20 retrieved items, while preserving the original retrieval order for the remaining candidates. Notably, all frontier MLLMs are evaluated with re-ranking enabled by default.

Retriever Configuration. We experiment with both a dense and a lexical retriever, which are shared by all methods. **For the dense retriever**, which is adopted in the main paper, we leverage the multimodal embedding model `Qwen3-VL-Embedding-2B` (Li et al., 2026b). Items are ranked by the cosine similarity. **For the lexical retriever**, we adopt BM25 (Robertson & Zaragoza, 2009) over the same textual descriptions of items, with lowercased alphanumeric tokenization and no stemming or stop-word removal. When replacing the dense retriever with BM25, we reuse exactly the same recommendation queries formulated by each model, so that the retriever is the only varying factor. The corresponding results are reported in Table 7. Notably, BM25 only accesses the textual content of the items. The retriever is kept frozen throughout the training and evaluation of Yo'Rec.

G Additional Empirical Results

In this section, we provide additional empirical results on our PersonaRec benchmark.

G.1 Evaluation with an Alternative Retriever

In Table 7, we report the performance of frontier MLLMs using BM25 (Robertson & Zaragoza, 2009), a lexical retriever that differs from the semantic retriever adopted in the main paper. We observe that lexical retrieval underperforms semantic retrieval on PersonaRec, which is expected, as the model-generated ideal profile may only preserve the semantic content of the relevant preferences rather than their exact lexical form, while semantic retrieval is better suited for such content.

G.2 Evaluation of Frontier MLLM Variants

In this section, we report additional results for selected frontier MLLMs equipped with **RAG** and **Memory** variants, following the same implementations used for the 9B model in the main paper. From Table 8, we observe that equipping these frontier MLLMs with such variants does not consistently improve their performance. As discussed in the main paper, both RAG and Memory rely on semantic retrieval over the interaction history, which can fail to surface preference-relevant evidence when the recommendation query is highly personalized and underspecified. Memory further constructs dedicated retrieval units and, in the case of SeCom, uses compressed representations for retrieval, which does not fundamentally resolve this retrieval bottleneck.

Table 8: Performance of variants of frontier MLLMs on PersonaRec across three domains.

U-Query	**Fashion**			**Food**			**Movie**			**AVG**		
Method	**R@1**	**R@10**	**N@10**	**R@1**	**R@10**	**N@10**	**R@1**	**R@10**	**N@10**	**R@1**	**R@10**	**N@10**
GPT-5.6-Sol	**8.20**	**32.18**	**19.82**	8.53	**39.92**	**22.83**	**2.00**	**8.25**	**4.85**	6.24	**26.78**	**15.83**
+ RAG	7.10	24.92	15.02	**14.34**	28.68	20.76	0.75	3.75	2.00	**7.40**	19.12	12.59
+ Memory	2.52	13.88	7.38	8.53	30.23	17.88	0.75	3.62	1.87	3.93	15.91	9.04
Qwen3.5-27B	6.47	26.81	15.28	10.08	24.81	16.35	**2.00**	7.00	4.02	6.18	19.54	11.88
+ RAG	5.21	22.24	12.41	11.24	30.62	19.86	1.25	5.38	3.04	5.90	19.41	11.77
+ Memory	3.15	15.14	8.50	7.75	30.23	17.75	0.75	3.88	2.09	3.88	16.42	9.45
PS-Query	**Fashion**			**Food**			**Movie**			**AVG**		
Method	**R@1**	**R@10**	**N@10**	**R@1**	**R@10**	**N@10**	**R@1**	**R@10**	**N@10**	**R@1**	**R@10**	**N@10**
GPT-5.6-Sol	**12.15**	**38.64**	**24.03**	14.34	**41.47**	**26.74**	**3.88**	**12.75**	**7.76**	**10.12**	**30.95**	**19.51**
+ RAG	8.20	32.18	18.61	12.40	31.78	21.79	2.12	6.62	3.94	7.57	23.53	14.78
+ Memory	5.68	22.87	12.87	10.47	29.07	19.19	1.38	5.00	3.17	5.84	18.98	11.74
Qwen3.5-27B	9.31	30.44	18.48	7.75	23.26	14.68	2.00	10.25	5.73	6.35	21.32	12.96
+ RAG	9.62	31.39	19.30	**15.50**	34.50	24.34	1.25	6.50	3.73	8.79	24.13	15.79
+ Memory	7.57	23.66	14.45	12.02	29.07	19.43	0.88	5.50	2.94	6.82	19.41	12.27

Table 9: Performance on Fashion sub-domains with U-Query. Best performance is in **bold**.

U-Query	**Fashion**								
	Tops			**Bottoms**			**Shoes**		
Method	**R@1**	**R@10**	**N@10**	**R@1**	**R@10**	**N@10**	**R@1**	**R@10**	**N@10**
Query-Only	2.08	10.42	5.18	0.00	12.50	5.52	0.00	3.70	1.85
GPT-5.6-Sol	10.42	28.12	19.09	10.00	38.75	23.33	5.56	37.04	22.55
Gemini-3.8-Flash	5.10	25.52	14.57	8.75	34.13	19.31	2.59	11.67	7.60
Qwen3.5-27B	4.17	23.96	14.29	7.50	27.50	15.75	7.41	25.93	15.93
Qwen3.5-9B	4.17	13.54	8.68	3.75	18.75	10.06	0.00	20.37	8.46
+ RAG	5.21	11.46	8.26	1.25	22.50	11.41	**7.41**	11.11	8.74
+ Memory	2.08	11.46	5.72	6.25	21.25	12.40	1.85	12.96	6.77
+ Rerank	9.38	16.67	12.80	7.50	23.75	14.27	3.70	20.37	10.24
+ Yo'Rec	**11.07**	**22.14**	**16.29**	**17.19**	**30.47**	**23.39**	4.63	**20.83**	**12.08**
	Glasses			**Bags**			**Jewelry**		
Method	**R@1**	**R@10**	**N@10**	**R@1**	**R@10**	**N@10**	**R@1**	**R@10**	**N@10**
Query-Only	1.45	5.80	3.55	3.08	15.38	7.77	2.94	10.29	5.57
GPT-5.6-Sol	5.07	22.46	12.68	9.23	40.00	24.05	8.82	31.62	20.41
Gemini-3.8-Flash	8.11	21.81	14.31	10.77	34.46	21.23	7.20	27.80	16.09
Qwen3.5-27B	5.80	19.57	11.37	9.23	38.46	21.37	5.15	25.00	13.58
Qwen3.5-9B	5.07	18.84	10.99	6.92	26.15	14.13	6.62	19.12	12.31
+ RAG	5.07	18.84	10.81	12.31	**33.08**	20.85	6.62	18.38	11.47
+ Memory	2.17	10.14	5.96	5.38	23.85	13.53	3.68	10.29	6.79
+ Rerank	7.97	18.12	13.24	9.23	27.69	17.65	7.35	22.06	13.80
+ Yo'Rec	**11.78**	**26.27**	**18.86**	**16.83**	31.73	**24.71**	**14.71**	**28.95**	**21.53**

G.3 Fine-Grained Performance on Sub-Domains

In Tables 9 to 14, we report the fine-grained performance of all methods at the sub-domain level.

H Limitations and Future Research Directions on PersonaRec

Benchmark Scope. PersonaRec currently covers three representative domains that are closely tied to everyday recommendation needs, namely Fashion, Food, and Movies, together with multiple fine-grained sub-domains. However, real-world personalized recommendation spans a much broader range of content and products. Future extensions could include electronics, health-related services, and emerging content domains such as short-form video. Expanding PersonaRec along these directions would not only increase its coverage, but also test whether preference signals accumulated through daily interaction remain useful across domains with substantially different item structures, preference spaces, and recommendation mechanisms.

Table 10: Performance on Food sub-domains with U-Query. Best performance is in **bold**.

U-Query	Food								
	Beverages			Alcoholic Drinks			Snacks		
Method	R@1	R@10	N@10	R@1	R@10	N@10	R@1	R@10	N@10
Query-Only	0.00	7.14	2.06	7.14	14.29	10.71	0.00	3.57	1.13
GPT-5.6-Sol	7.14	50.00	25.88	10.71	39.29	25.35	8.93	41.07	25.59
Gemini-3.8-Flash	10.00	50.00	28.50	15.00	42.50	28.04	18.75	32.50	23.72
Qwen3.5-27B	7.14	32.14	19.78	35.71	50.00	41.83	7.14	23.21	13.69
Qwen3.5-9B	7.14	21.43	14.73	7.14	28.57	15.61	3.57	17.86	11.71
+ RAG	7.14	35.71	18.89	14.29	**35.71**	24.51	**16.07**	33.93	23.16
+ Memory	0.00	25.00	10.11	**25.00**	**35.71**	**29.63**	5.36	26.79	16.16
+ Rerank	7.14	28.57	16.44	14.29	28.57	21.87	5.36	16.07	9.80
+ Yo'Rec	**26.79**	**40.18**	**34.65**	22.32	26.79	25.14	15.63	**35.71**	**26.26**
	Meals			Baked Goods			Confectionery		
Method	R@1	R@10	N@10	R@1	R@10	N@10	R@1	R@10	N@10
Query-Only	0.00	4.35	1.87	0.00	7.41	2.67	0.00	4.35	1.55
GPT-5.6-Sol	6.52	30.43	15.94	7.41	25.93	15.65	10.87	58.70	31.42
Gemini-3.8-Flash	1.52	13.70	7.11	7.78	19.45	13.10	13.70	22.83	17.35
Qwen3.5-27B	8.70	19.57	12.95	3.70	12.96	7.81	8.70	26.09	15.40
Qwen3.5-9B	6.52	8.70	7.15	3.70	12.96	7.89	6.52	23.91	13.17
+ RAG	4.35	10.87	7.65	11.11	20.37	14.48	4.35	23.91	11.98
+ Memory	0.00	19.57	8.68	7.41	20.37	12.85	6.52	19.57	11.90
+ Rerank	6.52	10.87	8.23	11.11	14.81	12.83	10.87	30.43	19.16
+ Yo'Rec	**19.02**	**32.61**	**25.26**	**20.83**	**26.62**	**24.02**	**17.66**	**32.61**	**23.83**

Realism and Controllability of Long-Term Interaction Data. Ideally, LIP-Rec would be studied using naturally accumulated long-term interactions between real users and personal MLLM assistants. In practice, this setting raises two major challenges. First, such histories inevitably contain highly private personal information, making large-scale collection and public release difficult without compromising user privacy. Second, even when privacy concerns can be addressed, densely annotating months or years of real interaction histories remains extremely costly, especially when the goal is to trace sparse preference evidence, capture preference evolution, and establish precise links between historical evidence and target items. To **balance realism, privacy, and controllability**, PersonaRec combines several sources of real-world grounding with carefully controlled construction. Candidate items are collected from real recommendation datasets, including `Amazon-Reviews-2023` (Hou et al., 2026) and `MovieLens-32M` (Harper & Konstan, 2015). Personas are derived from `PersonaHub` (Ge et al., 2024), which provides diverse user profiles grounded in web-scale data. Interaction scenarios are seeded from existing human–LLM interaction datasets so that generated sessions inherit realistic patterns of everyday assistant use. We further introduce extensive human auditing in construction and independent post-evaluation after construction to improve the naturalness, consistency, and reliability of the resulting data. While these choices do not make PersonaRec equivalent to naturally collected long-term user histories, they enable controlled evaluation of the core capabilities required by LIP-Rec while preserving privacy. They also provide fine-grained provenance that is rarely available in real-world logs, including the locations of preference evidence, the associated item attributes and values, and the links between those preferences and target items. This enables systematic diagnosis of each intermediate stage of LIP-Rec.

Future Research Directions. As an initial study establishing LIP-Rec, our primary goal is to define the problem, build a systematic benchmark, and reveal the key limitations of current MLLMs, rather than to exhaustively optimize the best possible solution. Accordingly, Yo'Rec should be viewed as a simple baseline that targets the major failure modes identified in our analysis. Considerable room remains for future methods. **One direction** is to improve fine-grained localization and recovery of item preferences from long interaction histories. **Another** is to design stronger learning objectives or reward signals that better transform recovered preferences into effective recommendation queries. **More efficient** use of long-term history is also an important direction. Our evaluation suggests that existing memory mechanisms can struggle when preconstructed representations or retrieval procedures fail to preserve or surface fine-grained evidence that only becomes relevant after a future recommendation request is observed. This does not diminish the potential of memory-based approaches. Instead, it motivates memory mechanisms that retain detailed personalized evidence and

Table 11: Performance on Movie sub-domains with U-Query. Best performance is in **bold**.

U-Query	Movie								
	Action			Comedy			Romance		
Method	R@1	R@10	N@10	R@1	R@10	N@10	R@1	R@10	N@10
Query-Only	2.27	2.27	2.27	0.00	0.00	0.00	0.00	4.44	2.36
GPT-5.6-Sol	1.14	9.09	4.78	4.44	10.00	6.81	4.44	10.00	7.44
Gemini-3.8-Flash	3.19	11.14	7.60	6.22	11.67	8.86	3.11	17.11	9.81
Qwen3.5-27B	0.00	5.68	2.19	2.22	2.22	2.22	4.44	13.33	8.63
Qwen3.5-9B	0.00	2.27	0.72	2.22	2.22	2.22	4.44	11.11	7.12
+ RAG	0.00	4.55	1.97	0.00	1.11	0.37	4.44	11.11	7.32
+ Memory	1.14	4.55	3.29	0.00	4.44	2.28	0.00	8.89	3.65
+ Rerank	0.00	0.00	0.00	2.22	3.33	2.62	5.56	10.00	7.48
+ Yo'Rec	**5.12**	**12.79**	**9.45**	**10.00**	**15.83**	**12.66**	**13.34**	**19.17**	**15.80**
	Crime			**Horror**			**Suspense**		
Method	**R@1**	**R@10**	**N@10**	**R@1**	**R@10**	**N@10**	**R@1**	**R@10**	**N@10**
Query-Only	2.27	9.09	4.72	0.00	6.82	2.41	0.00	4.44	1.70
GPT-5.6-Sol	1.14	2.27	1.70	0.00	10.23	5.06	0.00	8.89	4.22
Gemini-3.8-Flash	0.00	3.98	1.34	0.00	10.34	4.49	3.89	11.67	7.39
Qwen3.5-27B	0.00	2.27	0.75	5.68	11.36	8.27	2.22	10.00	5.70
Qwen3.5-9B	**2.27**	4.55	3.20	0.00	9.09	3.91	2.22	4.44	3.62
+ RAG	0.00	0.00	0.00	0.00	2.27	1.16	0.00	3.33	1.58
+ Memory	1.14	2.27	1.49	0.00	3.41	1.22	0.00	2.22	0.85
+ Rerank	1.14	5.68	3.25	2.27	5.68	3.80	0.00	4.44	2.80
+ Yo'Rec	1.70	**8.52**	**5.23**	**2.56**	**9.38**	**5.63**	**10.00**	**15.00**	**12.18**
	Animation			**Sci-Fi**			**Documentary**		
Method	**R@1**	**R@10**	**N@10**	**R@1**	**R@10**	**N@10**	**R@1**	**R@10**	**N@10**
Query-Only	0.00	2.22	0.96	2.27	6.82	3.84	0.00	2.27	1.43
GPT-5.6-Sol	3.33	11.11	6.91	3.41	4.55	3.79	0.00	7.95	2.79
Gemini-3.8-Flash	1.55	12.45	6.22	1.59	11.14	6.07	0.80	5.57	2.60
Qwen3.5-27B	0.00	8.89	3.15	3.41	4.55	3.77	0.00	4.55	1.46
Qwen3.5-9B	0.00	2.22	0.64	2.27	7.95	4.63	1.14	1.14	1.14
+ RAG	1.11	4.44	2.46	1.14	2.27	1.70	0.00	0.00	0.00
+ Memory	0.00	1.11	0.70	0.00	3.41	1.35	1.14	2.27	1.70
+ Rerank	2.22	6.67	3.58	4.55	5.68	4.95	0.00	1.14	0.72
+ Yo'Rec	**10.00**	**16.67**	**12.92**	**6.82**	**9.38**	**7.69**	**7.67**	**11.93**	**9.72**

make it accessible on demand for recommendation. **Agentic approaches** also provide a promising direction because LIP-Rec does not require every step to be solved by the personal MLLM assistant itself. Future systems may coordinate retrieval modules, memory systems, recommender models, and other external tools to support personalized recommendation. Finally, although PersonaRec has training data to facilitate method development, real-world deployments may not have access to task-specific supervision for every user population or recommendation environment. Developing **training-free**, in-context, or otherwise generalizable approaches for leveraging long-term human–MLLM interactions therefore remains an important practical challenge.

I Additional Cases of Yo'Rec and Baselines

We provide additional cases of Yo'Rec on PersonaRec in Figure 12.

J Additional Presentations of PersonaRec

In this section, we provide additional presentations from PersonaRec, including some session templates, interaction scenarios, and long-term interaction histories between users and their personal MLLM assistants. Figure 13 jointly presents three preference-expression templates (one randomly sampled per type) and five preference-revealing scenarios spanning the three domains. Each example includes its metadata in the header and its content below. [X] denotes the user and [Y] an attribute

Table 12: Performance on Fashion sub-domains with PS-Query. Best performance is in **bold**.

PS-Query	Fashion								
	Tops			Bottoms			Shoes		
Method	R@1	R@10	N@10	R@1	R@10	N@10	R@1	R@10	N@10
Query-Only	3.12	17.71	9.33	3.75	21.25	11.37	3.70	18.52	8.75
GPT-5.6-Sol	5.21	37.50	19.95	17.50	60.00	34.88	7.41	35.19	19.94
Gemini-3.8-Flash	9.48	27.71	18.31	15.75	35.88	23.65	5.19	25.93	15.25
Qwen3.5-27B	8.33	31.25	18.64	6.25	28.75	16.40	7.41	31.48	17.27
Qwen3.5-9B	3.12	21.88	11.44	10.00	**41.25**	22.61	3.70	**27.78**	13.45
+ RAG	5.21	18.75	10.54	10.00	40.00	24.77	**11.11**	24.07	**16.98**
+ Memory	2.08	15.62	8.31	10.00	22.50	15.93	3.70	20.37	11.22
+ Rerank	6.25	**28.12**	15.40	**22.50**	36.25	**29.84**	9.26	24.07	16.63
+ **Yo'Rec**	**13.02**	27.34	**20.32**	17.97	33.59	26.39	9.26	19.68	14.21
	Glasses			Bags			Jewelry		
Method	R@1	R@10	N@10	R@1	R@10	N@10	R@1	R@10	N@10
Query-Only	3.62	14.49	7.84	4.62	23.85	13.48	3.68	19.12	10.61
GPT-5.6-Sol	10.14	34.78	21.63	16.92	37.69	26.24	13.24	33.09	22.45
Gemini-3.8-Flash	7.10	29.42	17.77	9.70	35.54	21.45	10.30	26.77	17.58
Qwen3.5-27B	13.04	25.36	17.81	12.31	42.31	25.06	5.88	24.26	14.46
Qwen3.5-9B	7.25	25.36	15.48	5.38	31.54	16.67	7.35	22.79	14.95
+ RAG	7.97	23.91	15.59	14.62	38.46	24.78	7.35	19.12	12.21
+ Memory	2.17	17.39	8.87	10.00	31.54	19.58	7.35	20.59	12.79
+ Rerank	13.77	**26.81**	**20.52**	12.31	38.46	24.81	11.03	**29.41**	19.98
+ **Yo'Rec**	**15.85**	23.10	20.17	**22.11**	**38.94**	**31.64**	**12.87**	28.95	**20.78**

Table 13: Performance on Food sub-domains with PS-Query. Best performance is in **bold**.

PS-Query	Food								
	Beverages			Alcoholic Drinks			Snacks		
Method	R@1	R@10	N@10	R@1	R@10	N@10	R@1	R@10	N@10
Query-Only	3.57	17.86	10.84	0.00	21.43	12.65	7.14	21.43	12.58
GPT-5.6-Sol	21.43	57.14	38.46	17.86	57.14	37.21	16.07	46.43	30.01
Gemini-3.8-Flash	5.00	47.50	26.22	7.50	40.00	23.93	18.75	43.75	29.91
Qwen3.5-27B	0.00	25.00	14.12	10.71	46.43	28.02	10.71	30.36	18.46
Qwen3.5-9B	0.00	17.86	9.06	17.86	42.86	27.50	17.86	32.14	25.24
+ RAG	14.29	**35.71**	24.61	**39.29**	42.86	**40.48**	14.29	35.71	24.16
+ Memory	7.14	25.00	17.35	17.86	**57.14**	35.69	14.29	**41.07**	25.57
+ Rerank	0.00	25.00	13.44	32.14	42.86	35.99	12.50	33.93	23.58
+ **Yo'Rec**	**26.79**	**35.71**	**31.84**	24.56	31.25	28.49	**18.98**	34.60	**27.19**
	Meals			Baked Goods			Confectionery		
Method	R@1	R@10	N@10	R@1	R@10	N@10	R@1	R@10	N@10
Query-Only	0.00	2.17	0.94	3.70	5.56	4.42	2.17	8.70	4.73
GPT-5.6-Sol	8.70	41.30	22.60	11.11	20.37	14.94	15.22	41.30	27.27
Gemini-3.8-Flash	4.56	31.96	16.21	10.37	22.04	15.79	9.13	22.83	16.16
Qwen3.5-27B	4.35	10.87	7.09	7.41	14.81	9.96	10.87	21.74	15.42
Qwen3.5-9B	4.35	15.22	9.58	0.00	16.67	7.87	4.35	19.57	10.40
+ RAG	4.35	21.74	11.87	3.70	20.37	11.66	10.87	21.74	16.25
+ Memory	0.00	17.39	8.28	1.85	16.67	8.00	8.70	**26.09**	14.14
+ Rerank	8.70	17.39	13.31	12.96	16.67	15.30	4.35	17.39	11.17
+ **Yo'Rec**	**14.94**	**27.17**	**20.98**	**15.04**	**27.78**	**21.06**	**17.66**	25.81	**20.36**

value. These examples come from the LIP-Rec design library and exclude distraction templates and scenarios. Long-term interaction histories are presented in Figures 14 to 16.

K Prompt Presentation

In this section, we present the prompts used throughout our study, including the prompts for PersonaRec construction (Appendix K.1), the prompts designed in Yo'Rec (Appendix K.2), and the

Table 14: Performance on Movie sub-domains with PS-Query. Best performance is in **bold**.

PS-Query	Movie								
	Action			Comedy			Romance		
Method	R@1	R@10	N@10	R@1	R@10	N@10	R@1	R@10	N@10
Query-Only	2.27	5.68	3.58	0.00	4.44	2.07	3.33	10.00	6.28
GPT-5.6-Sol	2.27	6.82	4.99	3.33	13.33	8.29	6.67	20.00	12.31
Gemini-3.8-Flash	6.36	13.52	9.86	4.67	11.67	8.31	9.33	17.11	13.32
Qwen3.5-27B	3.41	10.23	6.74	0.00	6.67	2.92	5.56	21.11	11.99
Qwen3.5-9B	1.14	2.27	1.70	0.00	5.56	2.53	4.44	12.22	8.46
+ RAG	2.27	6.82	4.25	0.00	3.33	1.51	2.22	12.22	6.54
+ Memory	1.14	5.68	3.27	0.00	3.33	1.30	3.33	10.00	7.01
+ Rerank	1.14	4.55	2.68	0.00	5.56	1.99	5.56	15.56	10.11
+ Yo'Rec	**12.79**	**15.34**	**14.29**	**13.34**	**15.00**	**14.17**	**13.34**	**20.00**	**16.88**
	Crime			Horror			Suspense		
Method	R@1	R@10	N@10	R@1	R@10	N@10	R@1	R@10	N@10
Query-Only	1.14	4.55	2.37	1.14	4.55	2.52	2.22	6.67	4.04
GPT-5.6-Sol	0.00	4.55	1.59	4.55	13.64	7.54	4.44	17.78	10.67
Gemini-3.8-Flash	3.19	7.95	5.17	2.39	14.32	8.17	1.55	12.45	7.32
Qwen3.5-27B	0.00	9.09	4.23	2.27	10.23	5.99	4.44	12.22	8.44
Qwen3.5-9B	0.00	2.27	0.81	0.00	10.23	4.14	0.00	6.67	2.78
+ RAG	0.00	2.27	0.80	1.14	6.82	3.23	0.00	5.56	2.64
+ Memory	0.00	4.55	1.82	0.00	5.68	3.13	0.00	7.78	3.47
+ Rerank	0.00	6.82	2.39	5.68	**15.91**	10.43	1.11	10.00	5.20
+ Yo'Rec	**2.56**	**9.38**	**6.15**	**10.23**	15.34	**13.05**	**8.33**	**10.83**	**9.69**
	Animation			Sci-Fi			Documentary		
Method	R@1	R@10	N@10	R@1	R@10	N@10	R@1	R@10	N@10
Query-Only	0.00	7.78	3.36	3.41	7.95	5.07	2.27	4.55	3.05
GPT-5.6-Sol	5.56	14.44	9.91	3.41	9.09	5.65	4.55	14.77	8.65
Gemini-3.8-Flash	1.55	12.45	6.20	1.59	7.16	3.89	3.98	9.55	6.37
Qwen3.5-27B	1.11	6.67	3.93	1.14	13.64	6.26	0.00	2.27	0.95
Qwen3.5-9B	0.00	4.44	2.50	1.14	6.82	3.54	1.14	4.55	2.27
+ RAG	1.11	2.22	1.43	2.27	6.82	4.55	0.00	2.27	0.90
+ Memory	0.00	5.56	2.58	5.68	9.09	7.31	0.00	3.41	1.08
+ Rerank	4.44	8.89	6.10	4.55	**13.64**	8.75	3.41	4.55	3.90
+ Yo'Rec	**12.50**	**15.83**	**14.21**	**6.82**	**13.64**	**10.22**	**12.79**	**14.49**	**13.86**

prompts used for the baselines (Appendix K.3). In all prompts, the text in angle brackets (e.g., <persona>) denotes a slot that is filled for each instance.

K.1 Prompts Used for PersonaRec Construction

We present the core prompts for the persona–target item alignment, the long-term human–MLLM assistant interactions, and the personalized recommendation queries of PersonaRec. For the alignment, the LLM only estimates the potential item preferences of a persona over the attribute values that exist in the item pool (Prompt 1), and the target item is then selected from the item pool based on these lightweight preferences. Each interaction session is constructed in a *select-outline-expand* manner: the LLM first selects the session scenario that best fits the persona (Prompt 2), then selects a session template and outlines the session (Prompt 3), and finally expands the outline into the complete session in a role-play manner (Prompt 4). Notably, the target item I_U^* is never exposed in any of these prompts, and each prompt for constructing preference-revealing sessions (Prompts 2–4) specifies a single attribute A to be revealed. The candidate distraction sessions are vetted against the persona before being placed into the history (Prompt 5), and the recommendation query Q_U is finally written in the voice of each persona (Prompt 6). Since the original prompts are lengthy and carry format specifications that only serve our construction code (e.g., input and output schemas), we present their abridged versions, which retain all the instructions on the content. The complete prompts will be released with our code.

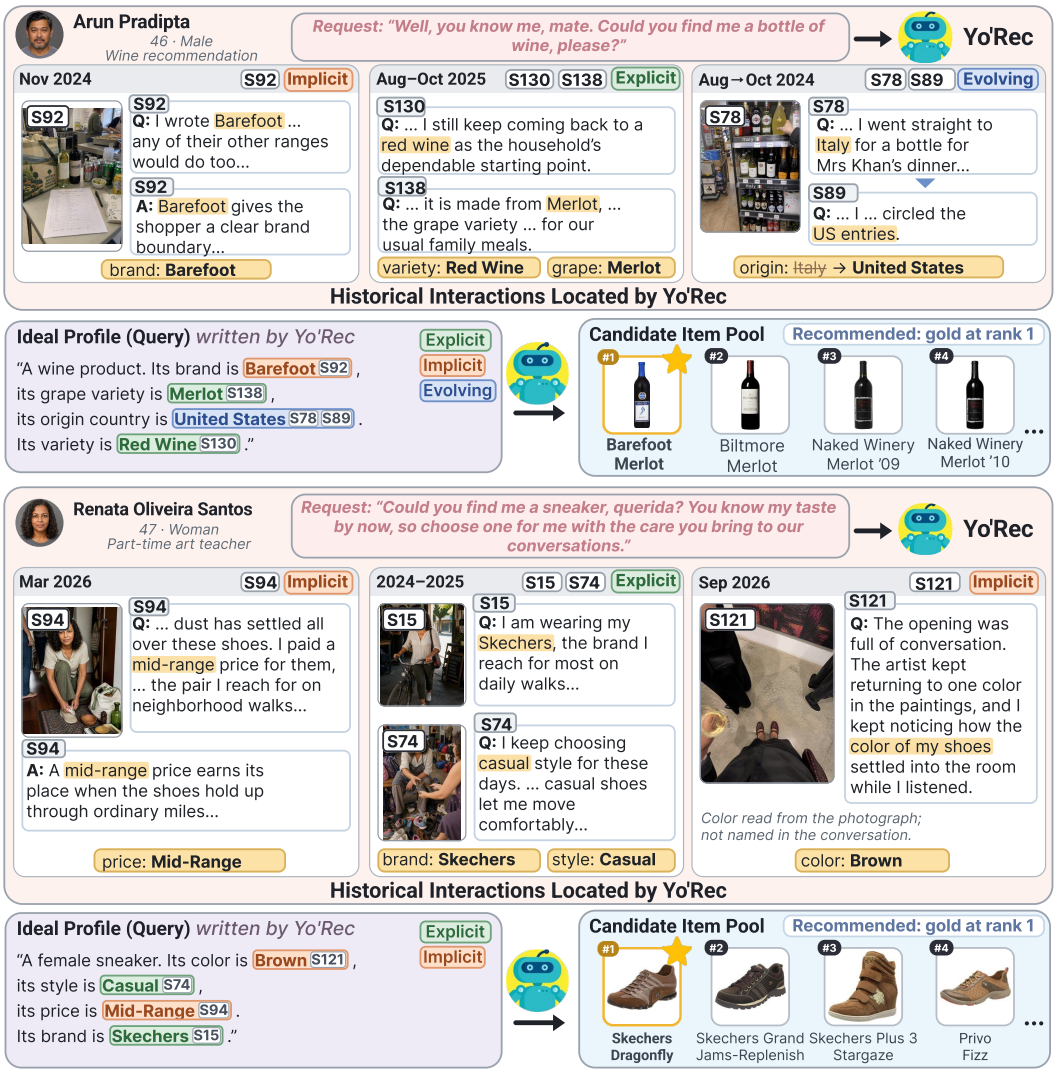


Figure 12: Additional cases of Yo'Rec on PersonaRec benchmark.

K.2 Prompts Designed in Yo'Rec

We present the prompts of Yo'Rec, which correspond to the three steps of PRPL-CoT (Prompts 7–9) and the final re-ranking (Prompt 10). The same prompts are used in the SFT stage, the PRC-PO stage, and inference. In Prompt 7 and Prompt 8, the interaction sessions are fed as multi-turn dialogues with their images, where the first user message of each session starts with a tag line of its session index and date. The wording of Prompt 9 and Prompt 10 is slightly adapted to each domain (e.g., the examples of the catalog style), and we present the Fashion version here.

K.3 Prompts Used for Baselines

We present the prompts shared by the frontier MLLMs and all the baselines built on `Qwen3.5-9B`: the prompt for formulating the final recommendation query from the (complete or retrieved) interaction history (Prompt 11), and the prompt for the **Rerank** baseline (Prompt 12). Only the content after "`PROFILE:`" in the response to Prompt 11 is used as the final recommendation query.

Type: Explicit; Sessions: 1; Modality: Text only
The user talks about the real-life experience defined by the current Scenario and clearly expresses how the target attribute value relates to the user's current or stable life. The dialogue follows the Scenario and naturally and explicitly mentions both the target attribute and its specific value.
Type: Implicit; Sessions: 1; Modality: Text + image
The user talks about the real-life experience and image in the current Scenario. The conversation must provide clear and sufficient positive evidence for a definite judgment that the user favors the target attribute value. The user does not directly state the preference conclusion, for example, 'I like waterproof shoes.' Instead, the preference is expressed through the real-life experience defined by the Scenario; the user's response, feeling, or evaluation is only an example, and the specific form follows the Scenario. The target product may appear naturally in the image so its category can be understood visually; the image may naturally present the real-life experience defined by the Scenario, and the specific attribute value is expressed naturally in the dialogue text. Requirement: (1) The dialogue text must naturally and explicitly mention and discuss both the target attribute and its specific value. (2) The dialogue text must not directly name the product category or target product.
Type: Evolving; Sessions: 2 Modality: Text only or text + image, per session
Across two sessions, the user shows an update for the same target attribute: session 1 occurs in the past and reflects an orientation toward value A, while session 2 occurs now or more recently and reflects an orientation toward value B. A and B must be different values of the same attribute. Each session selects a single-session template from allowed_session_template_ids whose dimensions match the current Scenario, and both sessions must belong to the same persona's continuous life.
Domain: Fashion; Category: Tops; Attribute: Sleeve length Preference types: Explicit / implicit
A friend gives [X] a top with [Y] sleeves, remembering their usual sleeve length and dislike of repeatedly rolling sleeves up.
Domain: Fashion; Category: Bags; Attribute: Capacity Preference types: Explicit / implicit
[X] repacks daily essentials into a [Y]-capacity bag. Their partner notes that this familiar size fits everything without accumulating clutter; [X] agrees.
Domain: Food; Category: Juice; Attribute: Pulp level Preference types: Explicit / implicit
A familiar restaurant remembers [X]'s juice with [Y] pulp content. Having this texture preference remembered keeps [X] coming back.
Domain: Food; Category: Confectionery; Attribute: Sugar status Preference types: Explicit / implicit
A colleague places [Y] candy beside [X]'s usual seat, knowing from many shared shifts that [X] always chooses this version.
Domain: Movie; Category: Action; Attribute: Action type Preference types: Explicit / implicit
Watching a [Y] action film at a martial arts school, [X] adds movement sketches to a well-used film notebook and watches until the credits.

Figure 13: Examples of templates and scenarios in PersonaRec.



Miguel

Albuquerque · Family life · Poetry & film

Fashion / Persona 25

170 sessions · 102 images

Sorting shirts before a poetry night

2025-11-16 / Session 003 — Preference-revealing

USER Exactly. **Casual tops** are what I reach for on accounting days and when I'm heading out to hear poems. This one has survived enough busy weeks to earn the keep pile.

ASSISTANT Then it has a practical place in your rotation. Fold it with the work-and-outing shirts and clear the rest of the cabinet.

Deciding whether to keep a shirt

2025-12-19 / Session 007 — Preference-revealing

USER I'm trying this shirt around the house before the return window closes. ...

ASSISTANT The neckline stays steady through those movements, sitting comfortably at the base of your neck. You shouldn't need to keep adjusting it.

USER ... The **neck style stayed put** and I haven't tugged at it once, so I'll **keep the shirt**.



Keeping a well-used pair of shoes

2026-06-15 / Session 030 — Preference-revealing

USER That park shortcut left these covered in mud. I've learned the color matters when I'm wiping them down; some shades show every streak, so I keep a cloth and little brush by the door. I don't mind the extra pass; this pair is **worth the routine**.

ASSISTANT A damp cloth first usually saves the finish before the mud dries; use the brush once it's dry, and keep those supplies by the door.



Solving a small painting mix-up

2026-11-11 / Session 049 — Non-preference interaction

USER I found two paint spots on the T-shirt after finishing the room, and I am trying to work out whether we opened the wrong can. The work photos show one came from the old primer, so, por favor, nobody needs to reopen the finish paint.

ASSISTANT The photos make the sequence clear: one spot is from the old primer, so the unopened finish paint can stay sealed. ...



Finding the right jacket at the theater

2027-01-25 / Session 060 — Non-preference interaction

USER At the theater coat check, my claim tag was hanging on the neighboring hanger, so another patron and I compared it with the entrance photo. The numbers finally lined up, and we each retrieved the right jacket. A small logistical mystery, solved.

ASSISTANT That sounds like a backstage accounting error, just with coats. Good thing the entrance photo preserved the original order; matching the claim numbers should settle it cleanly.

Choosing a patch at the tailor

2029-02-12 / Session 164 — Preference-revealing

USER I keep coming back to the cotton patch. It feels close to the shirt in my hands and bends with me; the shade is a little off, but that matters less than keeping this familiar movement.

ASSISTANT Cotton is the practical choice here. If it keeps the sleeve's hand feel and bends comfortably, a slight shade difference is a small tradeoff for a repair you can live in every day.



Helping with a cardboard portrait

2029-03-30 / Session 169 — Non-preference interaction

USER ... Now she has removed part of the neckline too, so I am holding it beside the cutout to explain which edge is the garment and which area is exposed neck.

ASSISTANT That sounds like a careful little anatomy lesson. You can trace the garment edge with your finger, then point to the open area above it so she can see how the portrait and clothing meet.

Figure 14: Presentation of persona Miguel's long-term interactions with personal MLLM assistant.



Marisol

San Antonio · Community life · Cycling

Food / Persona 6

172 sessions · 96 images

Listening to a family story

2025-03-16 / Session 001 — Non-preference interaction

USER I'm sitting with Tía Elena after lunch, and she's telling me about leaving home alone for the first time. She remembers the bus stations, the little rented rooms, and how frightened she felt carrying everything herself. …

ASSISTANT The stations and temporary homes must have made the journey feel very real, especially with that mixture of fear and determination.

Choosing a crust for the neighborhood potluck

2025-09-21 / Session 027 — Preference-revealing

USER I'm in the kitchen getting the pies ready for Monday's neighborhood potluck. I **keep coming back** to this **crust type** because the slices held together when I portioned the last batch, and the pieces sat neatly on the plates for volunteers.



ASSISTANT Let the pies cool fully before cutting, then set a small offset spatula beside the plates so volunteers can lift each slice cleanly.

Restoring order on the dining trolley

2026-01-20 / Session 044 — Non-preference interaction

USER A curve sent a wine bottle sliding from one trolley bay into the neighboring order section. I'm helping the attendant check seat numbers so the handoff stays orderly for everyone. I've attached the scene for our notes.



ASSISTANT Good catch. Let's match each order card to its seat number, then put the bottle back in its assigned bay and keep the service sequence clear.

Reading a bottle label with family

2026-01-31 / Session 046 — Non-preference interaction

USER My family brought out a bottle a friend left behind when moving, and we were trying to read the label properly. The front says it was made in Germany, but the back has an import notice, so I wasn't sure which side listed production and bottling.



ASSISTANT Check the original label first, then use the import notice as the added information. The production address should belong to the original entry, while the bottling address may appear in the later notice.

A favorite filling at a family tasting

2026-11-28 / Session 061 — Preference-revealing

USER I just finished my slice, crust and all. The **apple filling** was **warm and familiar**, exactly the center flavor I keep coming back to, even with that flaky crust. The other pies were fun to try, but this one disappeared first.

ASSISTANT The apple pie earned its place as your reliable favorite. Changing the crust is a nice experiment, but the apple filling gives you the familiar finish you want.

Fine-tuning pastries for a fundraiser

2028-02-04 / Session 120 — Preference-revealing

USER I pressed and tasted one of the fundraiser batch, and the texture is close to what I always enjoy: a **flaky outside** with a **moist center**. The crumb is holding together nicely.

ASSISTANT The flaky outside and moist center make it feel finished without being heavy.

Following a community cooking class

2028-04-06 / Session 129 — Non-preference interaction

USER I've brought the screen back, but the flour is everywhere. I'm looking at the rim now; should I press it down with a fork or smooth it by hand before we move on?

ASSISTANT Smooth it by hand first so the edge sits evenly, then add the fork marks if that is the finish being demonstrated. Work around the rim lightly and keep the shape consistent.

Figure 15: Presentation of persona Marisol's long-term interactions with personal MLLM assistant.



Amina

London · Community screenings · Film discussions

Movies / Persona 11

200 sessions · 110 images

Rewatching a romance at home

2025-12-20 / Session 074 — Preference-revealing

USER I noticed the same pattern in several romances I have rated highly recently: **just two people** carrying the **central relationship** from beginning to end. Perhaps that is why these films stay with me after the credits.

ASSISTANT It sounds as though the concentrated relationship arc makes the emotional stakes feel deeper for you.

Reflecting after an art-therapy film

2026-01-19 / Session 078 — Preference-revealing

USER After today's art-therapy film session, I kept thinking about its content intensity. The difficult themes were present, but held at a **low to moderate** level; there was enough emotional distance to consider them without being continually overwhelmed. That **balance matters** to me, particularly when the subject touches lived trauma.

ASSISTANT A low-to-moderate content intensity can protect reflective space without making painful themes feel trivial.

Keeping an alumni screening respectful

2026-08-15 / Session 109 — Non-preference interaction

USER I am helping with tonight's alumni screening of the second-chance romance short. A few attendees are whispering that the characters resemble classmates who are here, and I worry the speculation may become uncomfortable.

ASSISTANT Remind the room that the script is composite and resemblance does not identify anyone. A brief, calm note may protect the classmates' privacy.

Checking a film's continuity notes

2027-08-29 / Session 159 — Non-preference interaction



USER ... The third incident carries one date in the dialogue and another on a newspaper prop; the script-supervisor records finally establish when the character actually reads it. I have attached the working papers so the continuity note is precise.

ASSISTANT The continuity records should control the sequence: use the date when the character reads the newspaper, and note the conflicting dialogue date separately so the discrepancy remains traceable.

Choosing a family crime film

2028-01-29 / Session 179 — Preference-revealing



USER After dinner, my child and I watched a family crime film with only a brief chase; the disagreement was resolved through conversation. I find a film's violence level matters: **None or Low** feels humane, while **Moderate** is the highest I can comfortably accept.

ASSISTANT Dialogue-led tension can remain compelling in a crime film without making physical attacks the engine of the story.

Testing a projector before a screening

2028-04-11 / Session 186 — Preference-revealing



USER I am testing the borrowed projector for tonight's human-rights screening. On this ordinary hall setup, HD keeps the speakers' faces and subtitles clear, which is the practical video quality we need.

ASSISTANT If HD preserves clear faces and readable subtitles on the existing equipment, there is little practical reason to pursue a more elaborate upgrade.

Verifying a screening record

2028-07-12 / Session 198 — Non-preference interaction

USER I am checking the screening record in post-production now. The relationship configuration is listed as Love Triangle, and my colleague is verifying the spelling of the title, version, and screening date. It is meticulous work, but the record needs to be exact.

ASSISTANT That sounds like careful final verification. Once the spelling, version, and screening date are confirmed, the record should be ready for the next production step.

Figure 16: Presentation of persona Amina's long-term interactions with personal MLLM assistant.

PROMPT 1 **Persona–Item Preference Estimation**

Input:

– the basic profile of the persona *<persona>*;
– the kind of item of the assigned sub-domain *<item kind>*;
– for each attribute of this kind of item, the candidate values that exist in the item pool *<attribute candidates>*.

Instruction:

Estimate what THIS persona would prefer in this kind of item, not in the sub-domain at large.

– For every attribute, give the preferred value(s), a confidence between **0.0 and 1.0**, and a rationale grounded in the persona.
– Each preferred value must be **copied exactly** from the candidate values of its attribute.
– If the persona gives no basis for an attribute, answer "unknown" rather than guessing, and prefer "unknown" over a value that is only realistic for an incompatible audience.

Output:

```
{"dimensions": {"<attribute>": {
  "preferred": ["..."], "confidence": 0.0-1.0,
  "rationale": "..."}}}
```

Prompt 1: Persona–Item Preference Estimation Prompt.

PROMPT 2 **Session Scenario Selection**

Input:

– the basic profile of the persona *<persona>*;
– the attribute to be revealed *<attribute>* with the value of the target item *<value>*;
– the candidate scenarios *<scenarios>*, all of which can naturally reveal this value.

Instruction:

Pick THE scenario that best fits this persona. Read each candidate scenario and choose the ONE whose life situation this persona would most plausibly live, and inside which the value would surface most naturally through the situation logic of the scenario (need, feeling, outcome, judgment).

Judge persona fit only:

– Does the situation belong in this persona's actual life (career, routines, living situation, hobbies, life stage)? A scenario that contradicts a persona fact is out.
– Would the preference read as an incidental, believable by-product of that situation for this persona?
– Prefer mundane, recurring situations over exotic one-offs when both fit equally well.
– Do not prefer a scenario because its scene happens to be about the kind of item that will be requested later. A history in which every session is about the same kind of item reads as a list, not as a life.

Do not consider session templates, modality, or session and turn counts, which are decided later. **Do not rewrite or improve** any scenario, and **do not invent scenarios**. Write a one-line rationale that is entirely about persona fit and cites the persona facts that ground the pick.

Output:

```
{"chosen_scenario_id": "...",
  "scenario_rationale": "..."}
```

Prompt 2: Session Scenario Selection Prompt.

PROMPT 3 **Session Template Selection and Session Outline**

Input:

- *<persona>*;
- *<attribute>* and *<value>*;
- the chosen scenario *<scenario>*;
- the candidate session templates *<templates>*, already restricted to the designated preference type (explicit, implicit, or evolving);
- the range of turns per session *<turn budget>*;
- for an evolving preference, the candidate old values with their own scenarios *<old values>*.

Instruction:

A session is one independent chat at one point in time, and a turn is one round trip of a user message and an assistant reply inside it. Sessions of the same preference are separate chats placed at different points in the history, with unrelated conversations in between.

Step 1. Pick ONE template that most naturally lets THIS persona reveal the value inside the chosen scenario as an ordinary daily chat. The template sets the preference type, the number of sessions, and whether the session is text-only or multimodal. With a cumulative template (2-3 sessions), the same preference must show up in every session through different everyday occasions: repetition across time is the evidence. For an evolving preference, there are exactly two sessions: pick the OLD value whose shift to the current value reads most naturally for this persona, together with one of its scenarios for the first session, which is set in the past.

Step 2. Decide how many turns each session runs from its content, not from the range: a remark made in passing is one turn; an occasion that moves (the user does the thing, then reports how it went) is two; something with a real middle is three.

Step 3. Instantiate the scenario into THIS persona's life: concrete, mundane, theirs. Do not invent evidence that the scenario does not contain.

Step 4. Write a summary of 2-3 short sentences for each session. The chat must have its own mundane reason to exist (a trip, a gift, a repair, a weekend plan, ...), and the preference surfaces incidentally inside it:

- **explicit:** the user states the preference naturally in passing, grounded in a concrete past use or event, NOT as a survey answer.
- **implicit:** the preference is shown through behavior, a shared life photo, or repetition, and is never stated as a conclusion; the user does not ask for a recommendation. Pack at least two incidental cues.
- **evolving:** an old value (session 1, in the past) and then the current value (session 2, now). Each session must stand on its own as an ordinary chat: a reader seeing only session 2 must not be told "I used to prefer X".

Hard rules:

- Never change the value or substitute a synonym.
- When the template lets the image carry the value, the value is carried ONLY by the image: the image prompt must actually show it, and the dialogue must NOT state it, though the dialogue must still naturally discuss the attribute itself. Otherwise, the image only shows the event as background and the dialogue carries the value.
- When the template hides the item category, the dialogue never names it.
- No survey or interview phrasing ("what do you prefer", "rate ...").
- No leakage: do not phrase a finished retrieval query, and do not name any candidate item or its title.

PROMPT 3 **Session Template Selection and Session Outline** (continued)

– When the value means that something is absent (e.g., a solid pattern, no motif), frame the absence as the quality, and never describe the absent thing as a visible or appealing feature.
– Stay in the voice and register of the persona.

Output:

the chosen template, the number of sessions and turns, the summary and the beat of each session, and one image prompt per multimodal session.

Prompt 3: Session Template Selection and Session Outline Prompt.

PROMPT 4 **Session Construction**

Input:

– *<persona>* with the declared speaking style;
– the scenario and the summary of THIS session *<summary>*;
– the preference type;
– *<attribute>* and the value of this session *<value>*;
– the number of turns *<turns>*;
– whether the value is carried by the image or by the dialogue *<value visibility>*.

Instruction:

Write ONE everyday, casual conversation, i.e., one session, between the user and their personal assistant, and nothing else. Write exactly the given number of turns. Events happen in the user turns: the assistant was not there, so it never narrates what happened.

Voice. The declared speaking style of the persona is authoritative. Render at least two observable markers of it in every session (contractions and register, sentence rhythm, punctuation habits, their own vocabulary, opening habit), and never invent a marker that the description does not license. The assistant does not take the persona's voice: it stays neutral, warm, and brief.

How the preference surfaces.

– **explicit:** the user states the preference naturally in passing, grounded in the scenario or a concrete past use, never as a survey answer.
– **implicit:** the preference is shown through behavior, a shared life photo, or repetition, and is never stated as a conclusion; the user does not ask for a recommendation. Include at least two cues.
– **evolving:** write only the value of THIS session. The first session must read as a genuine, contemporaneous preference for the old value with no hint that it will change; the second must read as the current state with no reference back. The contrast is visible only to someone reading both.
– **distraction:** an ordinary unrelated chat grounded in its scenario; keep the core event of the scene and weave in the persona's own life details where they naturally fit.

Where the value lives. If the value is carried by the image, the photo must actually show it, and the words never state it, yet the attribute is plainly the thing being talked about. This binds both speakers: if the user has not said the value, the assistant cannot see it. If the value is carried by the dialogue (the image only sets the scene, or the session is text-only), state it naturally or paraphrase it so that a reader could name the value afterwards.

Style. User turns have **at most 60 words** and assistant turns **at most 80 words**, with at most one short paragraph per turn: real casual chats are short. The assistant answers the

PROMPT 4 **Session Construction** (continued)

actual question directly and stops. No survey or interview phrasing. No leakage: never phrase a finished retrieval query, and never name a candidate item or its title.

Output:

the messages of the session, where an image is marked by a placeholder together with the prompt for generating it.

Prompt 4: Session Construction Prompt.

PROMPT 5 **Distraction Session Vetting**

Input:

– *<persona>*;
– the candidate scenes for distraction sessions *<scenes>*.

For a scene in which the user dislikes or merely mentions an attribute value, the value *<value>* is also given.

Instruction:

These scenes are noise on purpose: a model has to read past them to find the sessions that matter. Your ONLY job is to veto the scenes that this persona's life CANNOT contain. A scene is in conflict when:

– it presumes a person or a thing they do not have (a child, a partner, a pet, a car, ...);
– it presumes work they do not do, or a life shape they do not live;
– it presumes a capability or a constraint they do not have;
– (dislike only) the profile states or clearly implies that they like, choose, or need this very value;
– the scene with the value filled in does not make sense.

Unremarkable is not a conflict. Untypical is not a conflict: people do things that are not in their profile. Silence is not a conflict: if the profile does not say whether they have something, they may have it. Do not judge taste fit, interest, or quality. When in doubt, do not veto.

Output:

for each scene, `{"conflicts": true / false, "reason": "..."}`

Prompt 5: Distraction Session Vetting Prompt.

PROMPT 6 **Personalized Recommendation Query (U-Query)**

Input:

– *<persona>* with the declared speaking style;
– the noun phrase of the requested kind of item *<noun>* (e.g., "shoulder bag", "jacket").

Instruction:

The query is the last user message after months of chat history: the user now asks the assistant to find them an item of the given kind, trusting the assistant to remember their taste. Write ONE query (1-2 short sentences) that:

– Names the noun and nothing else. The exact noun phrase must appear. Do NOT make it more specific or more general, and do NOT name ANY other property of the item.
– Sounds like THIS persona. Use their register, rhythm, and typical phrasing. Two personas asking for the same noun should sound clearly different.

PROMPT 6 **Personalized Recommendation Query (U-Query)** (continued)

– Stays natural. An everyday chat message, not a form. Leaning on the shared history is welcome ("you know my taste", "pick for me").
– Exposes nothing. Absolutely NO color, material, style, pattern, fit, size, price, brand, occasion, season, or feature words, not even ones the persona might plausibly say. No life events ("for my trip"), no constraints ("something light"). The chat history is the only evidence allowed to narrow the choice.

Output:

```
{"query": "..."}
```

Prompt 6: Personalized Recommendation Query (U-Query) Prompt.

PROMPT 7 **PRPL-CoT Step 1: Relevant Session Localization**

System:

You are reviewing a user's long-term chat history with an assistant. The history is a series of sessions in date order, and each session opens with a tag line like [S07 · 2025-03-14]. The user now asks for a certain kind of item. Your job is to find the sessions that say something about what this user wants in that kind of item: a wish stated outright, a habit, a complaint, or a photo that shows what they use. Sessions about everyday life with no bearing on the item do not count, and neither do sessions about a different kind of item. Reply with JSON only.

History:

<all sessions of the interaction history>

User:

New request: "*<recommendation query>*"

Which sessions carry evidence about what this user wants in this kind of item? Reply with JSON only, in the form `{"selected": ["S01", "S04"]}`.

Prompt 7: PRPL-CoT Step 1: Relevant Session Localization Prompt.

PROMPT 8 **PRPL-CoT Step 2: Item Preference Extraction**

System:

You are reading the sessions of a user's chat history that concern the kind of item the user is now asking for, and writing down what the user wants in that kind of item as attribute-value pairs, using the product catalog's attribute names and values. Read the photos as evidence too: what the user owns, wears or carries in them, and who is in frame. Write down only what the sessions support and do not invent a preference. If sessions disagree, **the later one wins**. Reply with JSON only.

History:

<the sessions located in Step 1>

User:

New request: "*<recommendation query>*"

List what this user wants in this kind of item, one attribute per entry, as JSON in the form `{"preferences": [{"dimension": "...", "value": "...", "from": ["S01"]}]}`, where `"from"` names the sessions that show it.

Prompt 8: PRPL-CoT Step 2: Item Preference Extraction Prompt.

PROMPT 9 **PRPL-CoT Step 3: Recommendation Query Rewriting**

System:

You write search queries for a fashion product catalog. Every item in the catalog is described in one short paragraph, and your query is a paragraph in exactly that form: a lead sentence naming who the item is for and what it is - "A female jacket." - then clauses of the form "Its color is Gray, its material is Canvas", three clauses per sentence. Given a request and the user's preferences, describe the catalogued item to look for. Cover every attribute the preferences name and no others, write each one the way the catalog lists it rather than the way the user put it, and give a multi-valued attribute all of its values ("its season is Spring and Fall"). Reply with the paragraph and nothing else.

User:

Request: "*<recommendation query>*"

Preferences:

<the item preferences extracted in Step 2, one "attribute = value" per line>

Describe the item to search for, in one paragraph, in the catalog's own style.

Prompt 9: PRPL-CoT Step 3: Recommendation Query Rewriting Prompt.

PROMPT 10 **Re-ranking of the Retrieved Items**

System:

You are re-ranking the shortlist a shopping assistant retrieved for one specific user. You are given the user's request, the ideal item profile generated in Step 3, and the candidates that were retrieved. Judge each candidate against that profile and put the best match first.

User:

The user's request:

<recommendation query>

What this user wants, as read from their history:

<the ideal item profile generated in Step 3>

Candidates (*<N>* of them, in retrieval order):

<the retrieved items, each with its index and textual attributes>

Rank all *<N>* candidates from best match to worst match for THIS user. A candidate that matches every preference belongs at the top; one that contradicts a preference belongs near the bottom even if it matches everything else. Weigh the preferences over generic quality or popularity.

Think it through in two or three sentences, then end your reply with a single line listing every candidate number exactly once, best first:

```
RANKING: 4, 1, 9, ...
```

Prompt 10: Re-ranking of the Retrieved Items Prompt.

PROMPT 11 **Recommendation Query Formulation**

System:

You are a personalization engine for a shopping assistant. You are given a user's long-term chat history with the assistant (text and photos they have shared over months) and their new request. The history contains scattered evidence about what this specific user likes - stated

PROMPT 11 **Recommendation Query Formulation** (continued)

outright, revealed through repeated behavior, or shown in a photo. Your job is to reconstruct the profile of the single item that is right for THIS user.

History:

<the interaction sessions provided to the method>

User:

The user's new request:

<recommendation query>

Preserve any preferences explicitly stated in the request, and use the history to infer additional preferences.

Work through it in three steps, using exactly these section headers:

`ASPECTS:` name the 3-5 things that decide whether an item of this kind suits this particular user. Derive them from the user's request and history - do not work down a standard checklist.

`EVIDENCE:` for each aspect, state what the request or history shows and **cite its source**. The photos the user shared are evidence too: say what you see in them and what it implies. If neither the request nor the history provides evidence for an aspect, write "no evidence" - **do not invent a preference**.

`PROFILE:` one short paragraph describing the single item that is right for this user, the way you would describe a product to someone. Be concrete about what it looks like, what it is made of, and what it is for - describe the ITEM, not the user's tastes ("dark brown pebbled leather with no visible branding", not "prefers warm, understated tones"). **Write nothing after this paragraph.**

Prompt 11: Recommendation Query Formulation Prompt.

PROMPT 12 **Re-ranking in the Rerank Baseline**

System:

You are re-ranking the shortlist a shopping assistant retrieved for one specific user. You are given the profile of the item that is right for this user - reconstructed from months of their chat history - and the candidates that were retrieved. Judge each candidate against that profile and put the best match first.

User:

The user's request:

<recommendation query>

The item that is right for this user, as reconstructed from their history:

<the formulated final recommendation query>

Candidates (*<N>* of them, currently in retrieval order):

<the retrieved items, each with its index and textual attributes>

Rank all *<N>* candidates from best match to worst match for THIS user. Weigh the specific things the profile commits to - what the item is made of, how it looks, what it is for - over generic quality or popularity. A candidate that contradicts the profile on one of those belongs near the bottom even if it matches everything else.

Think it through in two or three sentences, then end your reply with a single line listing every candidate number exactly once, best first:

```
RANKING: 4, 1, 9, ...
```

Prompt 12: Re-ranking in the Rerank Baseline Prompt.