

Personalization Matters: Interactive Conversational Shopping Agent with Dialogue User-Centric RAG

Anonymous ACL submission

Abstract

Real-world shopping requires dynamic interactions that surpass static search capabilities. Large Language Models (LLMs) function as "Personalization Agents" to align responses with individual preferences; however, effective frameworks and realistic benchmarks for domain-specific shopping remain scarce. This paper proposes a novel multi-agent framework addressing personalized shopping requirements, including user preferences and past interactions. We propose a novel multi-agent framework utilizing multimodal Retrieval-Augmented Generation (RAG) to integrate product metadata and user profiles for personalized shopping. Addressing the lack of domain-specific benchmarks, we introduce a robust evaluation suite for intent alignment and dialogue consistency. Experiments demonstrate significant improvements in preference tracking, cumulative information synthesis, and interaction progression compared to baselines¹.

1 Introduction

Conversational shopping differs fundamentally from conventional question answering and static information retrieval. In real-world shopping scenarios, users rarely specify their needs in a single query; instead, preferences and constraints are gradually revealed through multi-turn dialogue (Bernard and Balog, 2023; Du et al., 2025). The quality of a shopping assistant's response is therefore judged not only by factual correctness, but by how well it reflects subjective criteria such as personal taste, style consistency, and prior purchase experience as the interaction unfolds (Li et al., 2025a). Consequently, conversational agents often admits multiple reasonable responses, whose usefulness depends on sustained

¹All the codes and dataset are available at: https://anonymous.4open.science/r/Multimodel_RAG_Indexing-C04A/README.md

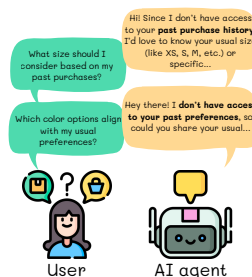


Figure 1: Agent-User interaction under no personalized setting

alignment with an individual user's evolving preferences rather than a single correct answer.

This interaction-driven setting exposes fundamental limitations in how personalization is commonly implemented. Existing approaches typically either rely on static user representations, such as fixed profiles or personas that remain unchanged within a dialogue session (Zhou et al., 2023), or perform personalization via per-turn retrieval of user-related information (Salemi et al., 2024c; Wang et al., 2024). While effective for improving local relevance, these designs do not explicitly model how user preferences should be accumulated, updated, and consistently applied across a multi-turn dialogue.

In this work, we argue that addressing these challenges requires treating personalization as a first-class, interaction-level process rather than an auxiliary conditioning signal. We propose a multi-agent framework for personalized conversational shopping that explicitly decomposes the interaction into complementary components responsible for dialogue state management, product retrieval, preference-aware reasoning, and response generation. This structured design enables the system to maintain and utilize user-centric information throughout multi-turn interactions, supporting consistent personalization as user intent evolves over time.

To ground personalization in concrete evidence,

we instantiate this framework using a retrieval-augmented generation (RAG) architecture that integrates multimodal product metadata with long-term user profiles derived from historical reviews. By retrieving and reasoning over user-specific and item-level evidence at inference time, the system generates personalized recommendations without modifying the underlying language model parameters. This design allows personalization to be expressed through controlled information flow across agents, rather than through parameter updates or static prompts.

Beyond system design, evaluating personalized conversational agents poses an additional challenge. Standard turn-level metrics and surface-form overlap measures fail to capture whether an agent consistently respects user preferences. To address this limitation, we introduce a trajectory-level, user-centric evaluation framework that assesses personalization quality from a holistic interaction perspective, focusing on global intent alignment, cumulative information synthesis, interaction progression, and dialogue style consistency.

Specifically, our contributions are threefold:

- We propose a multi-agent framework for personalized conversational shopping interaction that explicitly models user preferences and dialogue history to support user-centric reasoning in multi-turn conversations.
- We develop a retrieval-augmented generation architecture that integrates multimodal product metadata with long-term user personal profiles, enabling effective personalization in complex conversational shopping scenarios without modifying model parameters.
- We design a trajectory-level evaluation suite tailored to personalized conversational agents, introducing metrics that assess global intent alignment, cumulative information synthesis, and dialogue consistency beyond turn-level accuracy.

2 Related Work

2.1 Personalization Agents

Early personalization approaches in dialogue systems typically incorporate user information through static approaches, such as persona-conditioned generation (Zhou et al., 2023; Salemi et al., 2024a) or fine-tuning models to encode user-specific attributes (Tan et al., 2024). In these

methods, personalization is tightly coupled with the generation model itself and is not explicitly represented as a separate component in the system. To better handle longer interaction histories, later work introduces retrieval- and memory-based personalization strategies that integrate user histories or external knowledge. Representative approaches combine retrieval-augmented generation with structured memory or persona retrieval to condition responses on relevant personalized evidence (Wang et al., 2024; Chen et al., 2025). In general, early personalized methods treat personalization as auxiliary conditioning information within a largely unchanged generation pipeline (Li et al., 2025b).

More recently, some studies focus on how personalized information is organized and accessed beyond simple conditioning. For example, Reflective Memory Management proposes structured mechanisms to segment and refine dialogue history into personalized memory banks that are consulted during interaction (Tan et al., 2025), while memory-centric architectures such as Mem0 focus on the extraction, consolidation, and controlled retrieval of long-term personalized information (Chhikara et al., 2025). These works highlight a complementary perspective on personalization that emphasizes how personalized information is structured, maintained, and accessed over time, rather than modifying the core retrieval or generation architectures into the workflow.

2.2 Conversational Information Retrieval

Conversational Information Retrieval (CIR) aims to provide relevant information to users within an ongoing dialogue, moving beyond single-turn queries to understand user intent, track conversational history, and anticipate specific needs (Radlinski and Craswell, 2017; Mo et al., 2025). However, the complexities of information in interactive dialogue involving anaphora resolution and implicit context can often be challenging via most traditional retrieval methods (Adiwardana et al., 2020). Recent advancements are heavily influenced by the rise of LLMs. For example, history-aware extensions refine how context is used by down-weighting irrelevant turns and learning which parts of the dialogue (Mo et al., 2024). Interpretability works investigate rewriting-based approaches that expose intermediate queries or representations, making the behavior of conversational dense retrievers more trans-

parent and controllable in multi-turn dialogues (Cheng et al., 2024).

Furthermore, Retrieval-Augmented Generation (RAG) has emerged as a more scalable alternative by incorporating user history at inference time. Pseudo-RAG and improved retrieval mechanisms leverage interaction logs (Salemi et al., 2024c,b), while other systems integrate user-authored documents or passages for contextual prompting (Li et al., 2023; Mysore et al., 2024).

2.3 Personalization Conversation Evaluation

For conversational interaction performance, current evaluation frameworks mostly incorporate with fluency (Van Miltenburg et al., 2025), coherence (Ghazarian et al., 2022), and language accuracy (Komma et al., 2023), while user preference such as mixed-initiative behavior, coherence and perceived usefulness beyond traditional metrics (Huang et al., 2023; Fu1 et al., 2024) is still lacking. To address these limitations, the field has turned toward *User Simulation* as a scalable method to human manual evaluation. Early approaches modeled users as Markov Decision Processes to test dialogue policies (Schatzmann et al., 2006).

However, evaluating the *quality* of personalization specifically: How well an agent retrieves and utilizes user preference has also under gaps of emphasis. Evaluating personalized interactions need to be evolved from static, overlap-based metrics to dynamic, user-centric methodologies. While robustly evaluating the performance of a personalized model during a multi-turn session dialogue remains an open challenge, often requiring hybrid metrics that combine engagement scores with user intention checks (Roller et al., 2021). Recent works attempted to investigate dialogue interactive styles (Gao et al., 2024a,b, 2025a), from single modality to multi-modality (Gao et al., 2025b) for interactive focus. However, the lack of definition of user intention checks is still henge the fair evaluation of personalized agents in certain domain (e.g., e-commerce, medical, education) without intent-aware evaluation protocols, personalized systems risk to satisfy the user’s core informational needs.

3 Methodology

This section describes the methodology used to construct personalized multimodal shopping dia-

logues. We first present a high-level overview of the synthetic conversation pipeline, followed by a formal definition of the problem and its internal components, following by the retrieval-augmented generation (RAG) framework, the representations used for retrieval, and the multi-agent workflow that produces each turn of the conversation.

3.1 Overview of the Synthetic Conversation Pipeline

U: What style of casual top do you think would be best for me in terms of comfort and versatility?
A: *Based on your love for soft, breathable fabrics and relaxed fits, I've curated a few options that align with your preferences for comfort...*

U: Which sleeve length do you recommend for a casual top that I could easily wear at home?
A: *Following up on your earlier preference for soft, breathable fabrics... I've narrowed down sleeve options that align with your comfort-driven style!...*

U: When considering colors for a casual tunic, what would suit me best based on my usual style?
A: *Based on your history of favoring neutral tones... I've handpicked tunics that blend simplicity with subtle elegance...*

Table 1: Example interactions between User (U) and Shopping Assistant (A).

We simulate personalized interactive shopping domain dialogues using a five-agent framework: **User, Dialogue, Recommend, Reason, and Reply**. While the final transcript (Table 1) consists only of the User and Reply outputs, the remaining agents generate essential intermediate states—context tracking, product retrieval, and reasoning—to ground the final response:

- **Dialogue Agent** serves as a memory controller, summarizing prior turns and retrieving salient context to maintain continuity under length constraints
- **Recommend Agent** retrieves candidate products and reviews, reranks them using personalized evidence, and provides a concise ranking rationale
- **Reason Agent** interprets user-specific preference cues from the retrieved reviews and generates a detailed explanation of why the recommended items match the user’s needs.

A full example of the internal outputs produced within a single turn in Table 2.

Dialogue Agent. The Dialogue Agent constructs a compact textual context state by incrementally summarizing the dialogue history. At each turn, it summarizes the immediately preceding turn and appends this summary to a running collection of turn-level summaries. The accumulated summaries are then used to represent the dialogue history up to turn $t - 1$. In addition, the Dialogue Agent retrieves the most relevant past turn using a dialogue-level retrieval function R_d . By combining the accumulated summaries with the retrieved turn, the Dialogue Agent produces a concise and context-aware textual state r_t^d . Formally, this process can be written as

$$r_t^d = \text{LLM}_d(r_t^u, \tilde{H}_{t-1}, R_d),$$

where \tilde{H}_{t-1} denotes the concatenation of all turn-level summaries generated up to turn $t - 1$. The dialogue state r_t^d is shared with subsequent agents to maintain a consistent understanding of the current turn.

Recommend Agent. The Recommend Agent identifies and ranks potentially relevant products. It first applies a product-level retrieval function R_p to obtain an initial set of top- K candidate products based on their relevance to the current question r_t^u . It then applies a user-level retrieval function R_r to retrieve historically relevant reviews written by the same user. These retrieved reviews provide user-specific preference information that is used to rerank the initial product candidates. The resulting ranked list, together with a brief textual justification, forms the recommendation output

$$r_t^{rec} = \text{LLM}_{rec}(r_t^u, r_t^d, R_p, R_r).$$

Reason Agent. The Reason Agent generates a more comprehensive and explicitly personalized explanation beyond the ranking itself. Using the ranked candidates and the retrieved user-preference content, it produces an explanation r_t^{rea} that links the recommended products to the user’s long-term preferences and the intent of the current question:

$$r_t^{rea} = \text{LLM}_{rea}(r_t^u, r_t^d, r_t^{rec},).$$

Reply Agent. Finally, the Reply Agent combines the question r_t^u , the dialogue state r_t^d , the ranked candidates r_t^{rec} , and the personalized explanation r_t^{rea} to produce the final user-facing response:

$$r_t^{rep} = \text{LLM}_{rep}(r_t^u, r_t^d, r_t^{rec}, r_t^{rea}).$$

This response is the only output visible to the user, while all intermediate agent outputs remain internal to the system.

After these steps, the five outputs collectively form the turn-level structure

$$\tau_t = (r_t^u, r_t^d, r_t^{rec}, r_t^{rea}, r_t^{rep}),$$

which represents the complete internal computation at turn t . The retrieval functions R_d , R_p , and R_r are introduced here at a conceptual level, and their formal definitions are provided in Section 3.4.

3.3 Embedding Representations

To support RAG, we construct embeddings for products, users, and questions using **Qwen3-Embedding-8B (Qwen3E)**.

Product-information embeddings. Each product p_i is represented by three components: 1) Metadata $e_{p_i}^m$, encoded from textual metadata; 2) Review $e_{p_i}^r$, the mean embedding of all associated reviews; and 3) Description $e_{p_i}^d$, obtained by encoding a textual image description generated by **Qwen3-VL-30B-A3B-Instruct**. These form the multimodal basis for product retrieval.

User-centric embeddings. To model long-term preferences, we encode each of the user’s historical reviews (denoted as $e_{u_r_i}$), excluding the target product review to maintain a strict pre-purchase setting.

Question embeddings. At each turn t , the user question r_t^u is encoded into e_{u_t} , which serves as the query vector for all subsequent retrieval functions.

3.4 Retrieval-Augmented Generation

Leveraging these representations, our multi-level RAG mechanism retrieves dialogue, product, and user context. At turn t , the question embedding e_{u_t} combined with the pre-computed embeddings (Section 3.3) drives three distinct retrieval functions to support personalized interactions.

Dialogue-level retrieval. The function R_d retrieves the dialogue turn whose user question is most semantically similar to the current question. Formally, given the embedding e_{u_t} of r_t^u , $R_d(e_{u_t}) = \text{TopK}(e_{u_t}, \mathcal{D}_t^{\text{dlg}})$, where $\mathcal{D}_t^{\text{dlg}}$ stores

the embeddings of all previous user questions together with the corresponding turn-level text sequences. Similarity is measured using cosine similarity, and R_d returns the full turn corresponding to the most similar question.

Product-level retrieval. The function R_p identifies products relevant to the current question. Each product p_i is represented by metadata, review, and caption embeddings $\mathbf{e}_{p_i}^m$, $\mathbf{e}_{p_i}^r$, and $\mathbf{e}_{p_i}^d$. For a question embedding \mathbf{e}_{u_t} , we compute a relevance score $s_{p_i} = \frac{1}{3} \sum_{v \in \{m, r, d\}} \cos(\mathbf{e}_{u_t}, \mathbf{e}_{p_i}^v)$. The retrieval function returns the top- K products with the highest scores: $R_p(\mathbf{e}_{u_t}) = \text{TopK}(s_{p_i})$.

User-level retrieval. To incorporate personalization, the function R_r retrieves reviews written by the same user that are most relevant to the current question. Let $\mathcal{D}_u^{\text{amz}}$ denote the set of the user’s historical reviews, each associated with an embedding $\mathbf{e}_{u_{r_i}}$. We compute $R_r(\mathbf{e}_{u_t}) = \text{TopK}(\cos(\mathbf{e}_{u_t}, \mathbf{e}_{u_{r_i}}), \mathcal{D}_u^{\text{amz}})$, which returns the reviews whose content is most semantically aligned with r_t^u .

4 Dataset

We construct our simulation data from the Amazon Reviews 2023 dataset. This corpus provides the necessary product metadata, images, and user review histories to model personalized dialogue interactions. We preprocess the data to create a focused user–item–review subset and a retrieval corpus for the RAG components.

4.1 Dataset Processing

The Amazon Reviews’23 dataset contains 34 categories, including an *Unknown* category, which we remove because our setting assumes users shop within a single category. To obtain comprehensive item-level feedback, we retrieve all reviews associated with each purchased item, including those written by users outside the top-100 subset. For the remaining 33 categories, we reorganize metadata and reviews at the user level to support personalized preference modeling. For each category, we select the 100 most active users (by review count) and collect all items with their corresponding purchased reviews. Final dataset statistics are shown in Table ??.

4.2 User Profile

We sample 10 target users from the top 100 active users per category, ensuring each has at least

two reviews. One review is held out as the target product for question generation (Sec. 4.3), while the remaining reviews serve as the basis for modeling the user’s long-term preferences². We then prompt GPT-4o (Hurst et al., 2024) (Fig. 3) to distill this history into structured JSON profiles containing attributes such as *style*, *quality_priorities*, *dislikes*. Fields inapplicable to specific categories are marked *None* or recorded under *additional_notes*.

4.3 Question Generation

We simulate interactions by designating a high-rated history item (rating ≥ 4) as the *target product* (Section 4.2). Using this target and the user profile, we prompt an LLM to generate a diverse query sequence (Figure 4). To compel profile-based reasoning, we enforce criteria favoring preference-aligned inquiries (e.g., “Does this fit my style?”) over generic metadata lookups, ensuring the task transcends simple attribute retrieval.

5 Experiments

5.1 Agent Workflow

In the experimentation, we evaluate three agent workflows, 1. Baseline Agent Workflow, 2. Text-only RAG Enhanced Agent Workflow and 3. Multimodal RAG Enhanced Agent Workflow. All workflows are tested on the same set of generated user questions to ensure a controlled comparison. Then, we implement three progressively more capable agent workflows for later experimentation and evaluation:

$$\begin{aligned} r_t^d &= \text{LLM}_d(r_t^u, \tilde{H}_{t-1}, R_d), \\ r_t^{\text{rec}} &= \text{LLM}_{\text{rec}}(r_t^u, r_t^d, R_p, R_r) \end{aligned} \quad (1)$$

where specific retrieval inputs vary by configuration:

- 1. Baseline (No-RAG):** A fully ungrounded baseline where all retrieval components are disabled ($R_d, R_p, R_r = \emptyset$).
- 2. Text-only RAG:** Enables all retrieval functions (R_d, R_p, R_r). Product retrieval R_p uses only textual metadata and reviews ($V = \{m, r\}$).
- 3. Multimodal (No-Review):** Enables multimodal product retrieval ($V = \{m, r, d\}$, where d is the image description) but disables user-review

²In categories where fewer than 10 users satisfy this criterion, we keep all eligible users.

retrieval ($R_r = \emptyset$) to isolate personalization effects.

4. Multimodal (Full): The complete model. Enables all retrieval components (R_d, R_p, R_r) and utilizes multimodal product embeddings ($V = \{m, r, d\}$).

5.2 Evaluation Metrics Design

Standard turn-level metrics often overlook the long-horizon dynamics requisite for personalization (Guan et al., 2025). A locally appropriate response may still violate prior constraints (e.g., clothing fit preferences), constituting a failure in user intention consistency (Zhang et al., 2024; Qiu et al., 2025). Thus, metrics must capture sustained preference tracking beyond isolated turn accuracy. To address this, we propose a four-dimensional framework (Table 8) assessing the *multi-turn conversational integrity* of personalized assistants. Using a 5-point Likert scale and LLM-as-a-judge, we evaluate trajectory-level coherence, prioritizing memory retention and adaptation to evolving preferences over local fluency³.

Global Personalization Consistency (GPC).

This dimension measures the agent’s fidelity to the user’s persona and constraints throughout the entire interaction life cycle. A high GPC score indicates the system effectively retains long-term context and avoids attribute hallucination, ensuring that recommendations remain consistent with previously stated user preferences during shopping recommendation(Chen et al., 2024).

Cumulative Information Synthesis (CIS).

CIS evaluates the agent’s ability to aggregate information across multiple turns to form a coherent dialogue state (Lu et al., 2025). Unlike isolated retrieval, this metric assesses whether the model integrates early constraints with later reasoning to refine its output. High-performing models demonstrate a cumulative understanding, logically connecting disparate user inputs.

Interaction Trajectory (IT).

This criterion assesses the efficiency of the dialogue flow in reaching a resolution. Effective task-oriented dialogues typically follow a logical progression: *Clarify* → *Recommend* → *Refine* → *Close*. Agents that loop redundantly or fail to advance the goal

³All results have been human validated by three NLP PhD as volunteers for this LLM-as-Judge part, and the agreement is > 85%.

state receive lower scores, while those that drive the conversation toward a valid conclusion score higher (Gao et al., 2024b).

Tone Consistency (TC). TC captures the stability of the agent’s sociolinguistic alignment. It evaluates whether the system maintains a unified persona—friendly, professional, and empathetic—without erratic shifts in register. Consistency here serves as a proxy for trustworthiness, as distinct fluctuations between formal and casual tones degrade perceived naturalness (Gao et al., 2024a).

6 Results

6.1 Point-based Evaluation

Metric	Base	TextRAG	NoRev [‡]	Full [†]
GPC	3.77	4.93	4.88	4.93
CIS	3.17	4.88	4.79	4.89
IT	3.02	4.47	4.39	4.45
TC	5.00	5.00	5.00	5.00
Avg.	3.74	4.82	4.76	4.82

Table 3: Point-based results. [†]Full M-modal, [‡]No Review.

Table 3 demonstrates that RAG mechanisms substantially improve personalization over the naive Baseline. Both Multimodal and TextRAG models achieve superior scores in **Global Personalization Consistency** (4.93 vs. 3.77) and **Cumulative Information Synthesis** (~ 4.89 vs. 3.17), verifying their capacity for constraint retention and multi-turn synthesis. While RAG variants perform similarly on these metrics, Multimodal RAG maximizes **Interaction Trajectory** efficiency (4.44), indicating the utility of visual metadata in resolving dialogues. Notably, all models reach a ceiling on **Tone Consistency** (5.00), reflecting the robust sociolinguistic alignment of the underlying LLM (Gao et al., 2025b, 2024a).

In addition, we report an accuracy metric based on token overlap between the *target product* and the generated response⁴. As shown in Table 4, this accuracy exhibits trends consistent with the overall results in Table 3 across different agent variants. The **Baseline** struggles significantly, averaging only 10.31%, whereas all RAG-enhanced frameworks achieve substantial gains (> 40%),

⁴Specifically, we compute the token overlap ratio between the target product title and the model’s reply across the entire dialogue, and count a prediction as correct if the overlap exceeds 0.3.

underscoring the critical role of retrieval in open-domain product discovery. Among the variants, the **Full (Multimodal)** model attains the highest overall accuracy (43.73%), marginally outperforming **TextRAG** (42.73%). This suggests that visual product metadata provides distinct cues that refine retrieval precision. We observe a slight performance decay from early to late turns across all models, reflecting the challenge of maintaining accurate context as the interaction history lengthens. Notably, while **NoReview** remains competitive in lexical retrieval (41.90%), this metric focuses on surface-level accuracy rather than the personalized reasoning capabilities where it was previously shown to underperform.

6.2 Agent Variation Selection

To further distinguish the capabilities of the different version of agents, we conducted a pairwise comparison using an *LLM-as-a-judge* to simulate user preference (Table 4).

This evaluation focuses on direct "head-to-head" win rates between the full Multimodal model (Full) and its variants. The comparison between Full and NoReview highlights the critical role of review-based user RAG. The Full agent achieves a dominantly high win rate (up to 100% in specific configurations) against the NoReview variant. This confirms that without access to the "soft" preferences encapsulated in textual reviews, the agent struggles to generate recommendations that resonate with the user’s specific tastes, even if it has access to visual data.

However, when compared against TextRAG, the Full multimodal model demonstrates competitive performance. The direct pairwise evaluation reveals that the inclusion of visual modalities allows the agent to capture nuances—such as product style and visual aesthetics, which demonstrates that text-only retrieval may miss, leading to a more comprehensive and engaging user experience.

6.3 Qualitative Analysis: User-centric Agents quality

We analyze how personalization constraints—target goals, visual metadata, and user history—enhance interaction quality in depth

Agent A	Agent B	A Wins	B Wins	A Win Rate
Full [†]	TextRAG	325	2	99.4%
TextRAG	Full [†]	302	25	92.4%
Full [†]	NoReview [‡]	327	0	100%
NoReview [‡]	Full [†]	84	242	25.7%

Table 5: Pairwise comparison between TextRAG and multimodal variants. [†]Multimodal (Full). [‡]Multimodal (No Review).

(Tables 10, 11). Enforcing the **target product** constraint anchors the interaction, ensuring goal consistency. Integrating **visual metadata** enhances diversity; by leveraging image descriptions, the system uncovers latent attributes in sparse categories (e.g., *fragrance_products*), avoiding the repetitive loops seeking for general information from user preference or past habits observed in baselines. Both RAG variants explicitly reference the user’s past affinity for “someone like me best – natural floral scents” to personalize their requirements during shopping. In contrast, the *No Review* baseline fails to make this connection, offering generic suggestions. This confirms that maximizing personalization constraints by combining historical preference with rich semantics is essential for generating diverse, context-aware, and explainable responses.

Finally, **user history** proves critical for reasoning. Unlike the generic *No Review* baseline, RAG variants explicitly contextualize recommendations using past affinities (e.g., “driftwood”)(e.g., Figure 9). This confirms that combining historical and multimodal constraints is essential for explainable, context-aware generation during multi-turn agents interactions for personalized aims.

7 Conclusion

To sum up, we address the challenge of maintaining long-horizon user preference consistency in personalized conversational shopping. We propose a modular multi-agent framework that leverages retrieval-augmented generation to integrate dialogue history and user preferences without parameter updates. Additionally, we introduce a trajectory-level evaluation protocol designed to assess multi-turn coherence beyond standard accuracy metrics. Experiments on the Amazon Reviews 2023 dataset demonstrate that our multi-agent framework significantly outperforms baselines in both personalization accuracy and overall interaction quality.

8 Limitation

This work has several limitations. First, although the framework incorporates multimodal product information, retrieval is triggered solely by textual user queries, and image-initiated or mixed-modality queries are not supported. Enabling direct image-based retrieval would better accommodate visually driven shopping scenarios. Second, the retrieval component in our framework adopts a relatively simple embedding-based RAG formulation. While this design already yields strong performance, more structured retrieval paradigms, such as graph-based RAG, could be explored to further enhance reasoning and personalization. However, biases inherent to the underlying language model may influence the evaluation results.

References

Daniel Adiwardana, Minh-Thang Luong, David R So, Jamie Hall, Noah Fiedel, Romal Thoppilan, Zi Yang, Apoorv Kulshreshtha, Gaurav Nemade, Yifeng Lu, et al. 2020. Towards a human-like open-domain chatbot. *arXiv preprint arXiv:2001.09977*.

Nolwenn Bernard and Krisztian Balog. 2023. Mgs-hopdial: A multi-goal conversational dataset for e-commerce. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2775–2785.

Jin Chen, Zheng Liu, Xu Huang, Chenwang Wu, Qi Liu, Gangwei Jiang, Yuanhao Pu, Yuxuan Lei, Xiaolong Chen, Xingmei Wang, et al. 2024. When large language models meet personalization: Perspectives of challenges and opportunities. *World Wide Web*, 27(4):42.

Yi-Pei Chen, Noriki Nishida, Hideki Nakayama, and Yuji Matsumoto. 2025. [Post persona alignment for multi-session dialogue generation](#). In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 20184–20192, Suzhou, China. Association for Computational Linguistics.

Yiruo Cheng, Kelong Mao, and Zhicheng Dou. 2024. Interpreting conversational dense retrieval by rewriting-enhanced inversion of session embedding. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*.

Prateek Chhikara, Dev Khant, Saket Aryan, Taranjeet Singh, and Deshraj Yadav. 2025. Mem0: Building production-ready ai agents with scalable long-term memory. *arXiv preprint arXiv:2504.19413*.

Hanwen Du, Bo Peng, and Xia Ning. 2025. Sapien: Mastering multi-turn conversational recommendation with strategic planning and monte carlo tree

search. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 2629–2648.

Xiao Fu1, Aldo Lipani, and Noriko Kando. 2024. An evaluation framework for conversational information retrieval using user simulation. In *Joint Proceedings of the SIGIR-AP 2024 Workshops EMTICIR 2024 and UM-CIR 2024*.

Rena Gao, Carsten Roever, and Jey Han Lau. 2024a. Interaction matters: An evaluation framework for interactive dialogue assessment on english second language conversations. *arXiv preprint arXiv:2407.06479*.

Rena Gao, Jingxuan Wu, Xuotong Wu, Carsten Roever, Jing Wu, Long Lv, and Jey Han Lau. 2024b. An interpretable and crosslingual method for evaluating second-language dialogues. *arXiv preprint arXiv:2408.16518*.

Rena Gao, Xuotong Wu, Tatsuki Kuribayashi, Mingrui Ye, Siya Qi, Carsten Roever, Yuanxing Liu, Zheng Yuan, and Jey Han Lau. 2025a. Can llms simulate l2-english dialogue? an information-theoretic analysis of l1-dependent biases. *arXiv preprint arXiv:2502.14507*.

Rena Wei Gao, Xuotong Wu, Siwen Luo, Caren Han, and Feng Liu. 2025b. ‘no’matters: Out-of-distribution detection in multimodality multi-turn interactive dialogue download pdf. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 1848–1864.

Sarik Ghazarian, Nuan Wen, Aram Galstyan, and Nanyun Peng. 2022. Deam: Dialogue coherence evaluation using amr-based semantic manipulations. *arXiv preprint arXiv:2203.09711*.

Shengyue Guan, Haoyi Xiong, Jindong Wang, Jiang Bian, Bin Zhu, and Jian-guang Lou. 2025. Evaluating llm-based agents for multi-turn conversations: A survey. *arXiv preprint arXiv:2503.22458*.

Yupeng Hou, Jiacheng Li, Zhankui He, An Yan, Xiusi Chen, and Julian McAuley. 2024. [Bridging language and items for retrieval and recommendation](#). *ArXiv*, abs/2403.03952.

Chen Huang, Peixin Qin, Wenqiang Lei, and Jiancheng Lv. 2023. Reduce human labor on evaluating conversational information retrieval system: A human-machine collaboration approach. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*.

Aaron Hurst, Adam Lerer, Adam P Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, et al. 2024. Gpt-4o system card. *arXiv preprint arXiv:2410.21276*.

770	Abishek Komma, Nagesh Panyam Chandrasekarasastri, Timothy Leffel, Anuj Goyal, Angeliki Metallinou, Spyros Matsoukas, and Aram Galstyan. 2023.	Difference-aware user modeling for enhancing llm personalization. <i>arXiv preprint arXiv:2503.02450</i> .	827
771			828
772			
773	Toward more accurate and generalizable evaluation metrics for task-oriented dialogs . In <i>Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 5: Industry Track)</i> , pages 186–195, Toronto, Canada. Association for Computational Linguistics.	Filip Radlinski and Nick Craswell. 2017. A theoretical framework for conversational search. In <i>Proceedings of the 2017 Conference on Human Information Interaction and Retrieval (CHIIR '17)</i> , pages 117–126.	829
774			830
775			831
776			832
777			833
778			
779	Cheng Li, Mingyang Zhang, Qiaozhu Mei, Yaqing Wang, Spurthi Amba Hombaiah, Yi Liang, and Michael Bendersky. 2023. Teach llms to personalize—an approach inspired by writing education. <i>arXiv preprint arXiv:2308.07968</i> .	Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Eric Michael Smith, Y-Lan Boureau, et al. 2021. Recipes for building an open-domain chatbot. In <i>Proceedings of the 16th conference of the european chapter of the association for computational linguistics: Main volume</i> , pages 300–325.	834
780			835
781			836
782			837
783			838
784	Xiangci Li, Zhiyu Chen, Jason Ingyu Choi, Nikhita Vedula, Besnik Fetahu, Oleg Rokhlenko, and Shervin Malmasi. 2025a. Wizard of shopping: Target-oriented e-commerce dialogue generation with decision tree branching. <i>arXiv preprint arXiv:2502.00969</i> .	Alireza Salemi, Surya Kallumadi, and Hamed Zamani. 2024a. Optimization methods for personalizing large language models through retrieval augmentation. In <i>Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval</i> , pages 752–762.	839
785			840
786			841
787			842
788			843
789			844
790	Xiaopeng Li, Pengyue Jia, Derong Xu, Yi Wen, Yingyi Zhang, Wenlin Zhang, Wanyu Wang, Yichao Wang, Zhaocheng Du, Xiangyang Li, et al. 2025b. A survey of personalization: From rag to agent. <i>arXiv preprint arXiv:2504.10147</i> .	Alireza Salemi, Surya Kallumadi, and Hamed Zamani. 2024b. Optimization methods for personalizing large language models through retrieval augmentation . In <i>Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval</i> , SIGIR '24, page 752–762, New York, NY, USA. Association for Computing Machinery.	845
791			846
792			847
793			848
794			849
795	Yen-Ju Lu, Ting-Yao Hu, Hema Swetha Koppula, Hadi Pouransari, Jen-Hao Rick Chang, Yin Xia, Xiang Kong, Qi Zhu, Xiaoming Simon Wang, Oncel Tuzel, et al. 2025. Mutual reinforcement of llm dialogue synthesis and summarization capabilities for few-shot dialogue summarization. In <i>Findings of the Association for Computational Linguistics: NAACL 2025</i> , pages 7237–7256.	Alireza Salemi, Sheshera Mysore, Michael Bendersky, and Hamed Zamani. 2024c. Lamp: When large language models meet personalization. In <i>Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)</i> , pages 7370–7392.	850
796			851
797			852
798			853
799			854
800			855
801			856
802			857
803	Fengran Mo, Kelong Mao, Ziliang Zhao, Hongjin Qian, Haonan Chen, Yiruo Cheng, Xiaoxi Li, Yutao Zhu, Zhicheng Dou, and Jian-Yun Nie. 2025. A survey of conversational search. <i>ACM Transactions on Information Systems</i> .	Jost Schatzmann, Karl Weilhammer, Matt Stuttle, and Steve Young. 2006. A survey of statistical user simulation techniques for reinforcement-learning of dialogue management strategies. <i>The knowledge engineering review</i> , 21(2):97–126.	858
804			859
805			860
806			
807			
808	Fengran Mo, Chen Qu, Kelong Mao, Tianyu Zhu, Zhan Su, Kaiyu Huang, and Jian-Yun Nie. 2024. History-aware conversational dense retrieval. In <i>Findings of the Association for Computational Linguistics: ACL 2024</i> .	Zhaoxuan Tan, Qingkai Zeng, Yijun Tian, Zheyuan Liu, Bing Yin, and Meng Jiang. 2024. Democratizing large language models via personalized parameter-efficient fine-tuning. <i>arXiv preprint arXiv:2402.04401</i> .	861
809			862
810			863
811			864
812			865
813	Sheshera Mysore, Zhuoran Lu, Mengting Wan, Longqi Yang, Bahareh Sarrafzadeh, Steve Menezes, Tina Baghaee, Emmanuel Barajas Gonzalez, Jennifer Neville, and Tara Safavi. 2024. Pearl: Personalizing large language model writing assistants with generation-calibrated retrievers . In <i>Proceedings of the 1st Workshop on Customizable NLP: Progress and Challenges in Customizing NLP for a Domain, Application, Group, or Individual (CustomNLP4U)</i> , pages 198–219, Miami, Florida, USA. Association for Computational Linguistics.	Zhen Tan, Jun Yan, I-Hung Hsu, Rujun Han, Zifeng Wang, Long Le, Yiwen Song, Yanfei Chen, Hamid Palangi, George Lee, et al. 2025. In prospect and retrospect: Reflective memory management for long-term personalized dialogue agents. In <i>Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)</i> , pages 8416–8439.	866
814			867
815			868
816			869
817			870
818			
819			
820			
821			
822			
823			
824	Yilun Qiu, Xiaoyan Zhao, Yang Zhang, Yimeng Bai, Wenjie Wang, Hong Cheng, Fuli Feng, and Tat-Seng Chua. 2025. Measuring what makes you unique:	Emiel Van Miltenburg, Anouck Braggaar, Emmelyn Croes, Florian Kunneman, Christine Liebrecht, and	871
825			872
826			873

881 Gabriella Martijn. 2025. Measure only what is mea-
882 surable: towards conversation requirements for eval-
883 uating task-oriented dialogue systems. In *Proceed-*
884 *ings of the Fourth Workshop on Generation, Evalu-*
885 *ation and Metrics (GEM²)*, pages 231–238, Vienna,
886 Austria and virtual meeting. Association for Com-
887 putational Linguistics.

888 Zheng Wang, Zhongyang Li, Zeren Jiang, Dandan Tu,
889 and Wei Shi. 2024. Crafting personalized agents
890 through retrieval-augmented generation on editable
891 memory graphs. In *Proceedings of the 2024 Con-*
892 *ference on Empirical Methods in Natural Language*
893 *Processing*, pages 4891–4906.

894 Zhehao Zhang, Ryan A Rossi, Branislav Kveton, Yijia
895 Shao, Diyi Yang, Hamed Zamani, Franck Dernon-
896 court, Joe Barrow, Tong Yu, Sungchul Kim, et al.
897 2024. Personalization of large language models: A
898 survey. *arXiv preprint arXiv:2411.00027*.

899 Wangchunshu Zhou, Qifei Li, and Chenle Li. 2023.
900 Learning to predict persona information for dialogue
901 personalization without explicit persona description.
902 In *Findings of the Association for Computational*
903 *Linguistics: ACL 2023*, pages 2979–2991.

A Appendix

A.1 Dataset Statistics

Category	Items	Reviews	Avg.
Electronics	27,292	9,907,789	363.03
Home_and_Kitchen	51,421	8,780,109	170.75
Kindle_Store	113,360	7,172,600	63.27
Movies_and_TV	67,650	6,112,901	90.36
Health_and_Household	34,382	5,492,657	159.75
Clothing_Shoes_and_Jewelry	63,329	5,360,811	84.65
Pet_Supplies	15,248	5,225,236	342.68
Books	118,159	4,967,408	42.04
Grocery_and_Gourmet_Food	24,308	4,226,505	173.87
Beauty_and_Personal_Care	42,599	3,935,557	92.39
Tools_and_Home_Improvement	27,939	3,666,301	131.23
Software	7,637	3,632,785	475.68
Cell_Phones_and_Accessories	10,171	2,861,995	281.39
Automotive	19,272	2,854,050	148.09
Toys_and_Games	26,279	2,364,092	89.96
Patio_Lawn_and_Garden	16,791	2,320,870	138.22
Office_Products	12,561	2,237,761	178.15
Sports_and_Outdoors	18,621	1,982,891	106.49
Video_Games	8,551	1,838,654	215.02
CDs_and_Vinyl	58,707	1,539,120	26.22
Arts_Crafts_and_Sewing	16,198	1,369,350	84.54
Baby_Products	9,616	1,286,781	133.82
Musical_Instruments	9,396	954,286	101.56
Industrial_and_Scientific	8,038	741,943	92.30
Appliances	2,288	260,620	113.91
Gift_Cards	275	138,737	504.50
Magazine_Subscriptions	554	47,753	86.20
Amazon_Fashion	4,890	37,248	7.62
Health_and_Personal_Care	1,351	36,817	27.25
All_Beauty	2,040	33,832	16.58
Handmade_Products	1,504	19,603	13.03
Subscription_Boxes	156	11,696	74.97
Digital_Music	4,538	9,876	2.18

Table 6: Detailed statistics of product categories, including number of items, number of reviews, and average reviews per item.

A.2 Profile and Question Generation Prompt

User Profile Generation Prompt

You are an expert at analyzing shopping behavior and extracting user preferences from product reviews.

I will provide you with ALL reviews written by a specific Amazon Fashion customer (excluding one target product).

IMPORTANT: Each review includes a RATING (1-5 stars). Pay close attention to ratings when analyzing preferences.

Your task is to analyze these reviews with a focus on RATING-BASED PATTERNS.

Rating Analysis Instructions:

- High Ratings (4-5 stars):** Extract what the user LIKES and VALUES
 - Positive product features, styles, brands, materials
 - What made them satisfied?

- Low Ratings (1-3 stars):** Extract what the user DISLIKES and AVOIDS
 - Negative experiences, quality issues, sizing problems
 - What disappointed them?

- Rating Patterns:** Notice trends
 - Do certain brands/styles consistently get high or low ratings?
 - Are there recurring complaints (e.g., "always too small", "fabric quality")?

Profile Elements to Extract:

- Preferences:** What does this user actively LIKE? (from 4-5 star reviews)
- Style:** Fashion style preferences (Casual, Formal, Athletic, etc.)
- Budget:** Price sensitivity based on review patterns and ratings
- Size:** Common sizes mentioned, fit preferences
- Quality Priorities:** What quality aspects matter most?
- Dislikes:** What does this user actively AVOID or dislike? (from 1-3 star reviews)
- Pain Points:** Recurring issues that cause dissatisfaction

User's Review History:

```
{review_corpus}
```

Output Schema and Constraints

Output Format (JSON ONLY):

```
{
  "Preferences": "What user likes based on high-rated (4-5 star) reviews. Include specific features, styles, materials.",
  "Style": "Fashion style preferences (e.g., Casual, Athletic, Minimalist)",
  "Budget": "Budget-friendly | Mid-range | Premium | Mixed (infer from products and satisfaction levels)",
  "Size": "Size information and fit preferences",
  "Quality_Priorities": "Key quality aspects that matter (e.g., Fabric comfort, Durability, Accurate sizing)",
  "Dislikes": "What user dislikes based on low-rated (1-3 star) reviews. Include materials, styles, brands to avoid.",
  "Pain_Points": "Recurring issues causing dissatisfaction (e.g., sizing inconsistency, fabric pilling, poor stitching)",
  "Additional_Notes": "Other notable patterns or context"
}
```

Instructions:

- Be SPECIFIC and EVIDENCE-BASED - cite rating levels when relevant
- Distinguish between high-rated and low-rated patterns
- If information is not available, use "Not specified" or "No clear pattern"
- Focus on PATTERNS across multiple reviews, not isolated mentions
- Keep each field concise (1-3 sentences max)
- Use phrases like "consistently rated 5 stars when..." or "gave 1-2 stars due to..."

Figure 3: Full prompt used to generate rating-aware user preference profiles from Amazon Fashion review histories.

Question Generation Prompt

You are a question generator for a personalized shopping assistant.

SCENARIO: - This is a RECOMMENDATION conversation. - The user has NOT seen any specific item yet. - They are generally shopping for a {product_type} (or similar items) and want the assistant to recommend options. - The target product metadata below represents one candidate item the assistant might recommend, but the user does NOT refer to it directly.

INPUTS: - Target product metadata (all available attributes). - A precomputed USER PROFILE summarizing their long-term preferences.

TARGET PRODUCT METADATA (for your understanding only, NOT mentioned by name in questions): {metadata_text}

USER PROFILE (INTERNAL LATENT PREFERENCES - DO NOT REPEAT LITERALLY): {profile_text}

GOAL: Generate user-side shopping questions that: - Look natural and fairly generic on the surface. - Are phrased as the user asking for RECOMMENDATIONS for a {product_type} or similar items. - Do NOT explicitly state the user's preferences. - Create decision points where a generic assistant and a preference-aware assistant would give different answers.

CATEGORY & ATTRIBUTE AWARENESS (IMPORTANT):

You MUST determine: - what kind of product this is, - which attributes matter for evaluating or choosing it, ONLY from the target product's metadata provided above.

Do NOT use any assumptions about product types such as clothing, books, electronics, or shoes. Do NOT rely on any templates, stereotypes, or memorized category schemas. Everything must be derived from THIS target product's metadata.

ATTRIBUTE CONSTRAINTS (EXTREMELY IMPORTANT):

You are ONLY allowed to generate questions about attributes that are EXPLICITLY present, or strongly implied, in the target product metadata.

Allowed attribute sources include: - title - features - material - format - dimensions - usage statements (e.g., "exercise", "home", "casual") - included accessories - technical specs - pattern, print, color, if shown in metadata - category name fields - details fields

You MUST NOT generate questions about attributes that do NOT exist in the metadata. For example: - Do NOT ask about size, fit, color, or patterns unless the metadata clearly suggests they exist. - Do NOT ask about battery life unless the metadata suggests an electronic device. - Do NOT ask about genre unless the metadata suggests a book or similar media. - Do NOT assume the item is clothing, electronics, books, shoes, or any specific category. You MUST infer the type ONLY from the metadata itself.

USER PROFILE USAGE RULE (CRITICAL):

- The USER PROFILE above already summarizes this user's long-term preferences, style, budget, quality priorities, dislikes, and pain points. - You MUST treat this

profile as the PRIMARY source of latent preferences. - You MUST NOT invent new preferences that contradict this profile. - You MUST NOT copy sentences or long phrases from the profile into the questions. - Instead, you should: - Use the profile to decide which decision dimensions are important for this user (e.g., comfort, natural look, ease of use, durability, value for money, etc.). - Design questions that create decision points ALONG those dimensions, so that an assistant WITH access to this profile can answer differently from an assistant WITHOUT it. - The questions must sound like things the user might naturally ask, not like a restatement of the profile.

GOOD QUESTION STYLE EXAMPLE (DO NOT COPY TEXT, ONLY STYLE):

For a case where the user is looking for a tank top, good questions look like:

- "I'm looking for a tank top. What would you recommend?" - "What style of tank top suits me best - fitted or relaxed?" - "Which material would be most comfortable for me?" - "Which color options align with my usual preferences?" - "What size should I consider based on my past purchases?" - "Is this tank top suitable for both workouts and casual outings?" - "How does the fit of this kind of tank top compare to what I usually wear?" - "What care instructions should I consider for maintaining quality?"

These questions are GOOD because: - The user is asking for recommendations, not describing a specific known item. - The assistant has to decide FOR THE USER. - They hint at decisions that depend on the user's history ("my usual preferences", "what I usually wear", "based on my past purchases"). - They do NOT include specific hidden preferences like "for warm weather", "for sensitive skin", "because I sweat a lot". - They refer to high-level attributes (style, material, color, size, use case, care) that are typical for that product type.

QUESTION DESIGN RULE (VERY IMPORTANT):

Write questions as if the USER is asking a shopping assistant for help choosing or evaluating a {product_type} or similar items, BEFORE seeing any specific product page. Your questions MUST satisfy ALL of the following:

1) RECOMMENDATION, NOT EXPLANATION - The tone must be "help me choose / what would you suggest for me", NOT "explain this product to me" or "what are the advantages of X in general". - BAD: "What material would offer the most durability for my needs?" -> almost any user gets the same answer. - GOOD: "When choosing between softer and more durable materials in {product_type}s, which would you recommend for me?"

2) AROUND A PREFERENCE DIMENSION OR TRADE-OFF - EACH question must revolve around at least one decision dimension that can depend on preferences: - comfort / softness - durability / build quality - appearance / style / how it looks on me - usage scenario (everyday vs special occasions, casual vs active, etc.) - ease of care / cleaning / maintenance - value for money, etc. - Prefer questions that implicitly ask for a choice or trade-off: - BAD: "What features should I look for in shorts?" - GOOD: "Given what I usually go for, which features should matter most for me when choosing shorts?"

3) EXPLICITLY ABOUT "ME" AND OFTEN ABOUT PAST BEHAVIOR - Every question must clearly indicate that the user is asking FOR THEMSELVES, using phrases like: - "for me", "for someone like me", "for my needs", - "based on what I usually wear/use", - "com-

pared to what I typically go for", - "given what I've liked before". - At least HALF of the questions must explicitly mention past or typical behavior, e.g.: - "based on what I usually wear", - "given the kinds of things I've liked before", - "compared to what I normally choose". - Example: - GOOD: "Based on the kinds of {product_type}s I usually use, which style would you recommend for me?"

4) DO NOT BAKE HIDDEN PREFERENCES INTO THE QUESTION - You MUST NOT write concrete personal conditions that reveal hidden preferences, such as: - "for warm weather", "for my sensitive skin", "because my feet are wide", etc. - Those details should be discovered by the assistant from the USER PROFILE, NOT baked into the question itself. - You MAY use neutral quality words like "comfortable", "durable", "stylish", but without overly specific situational constraints. - OK: "Which material would be more comfortable for me?" - NOT OK: "Which material would keep me comfortable in very hot weather?"

5) GENERIC PRODUCT TYPE, NOT SPECIFIC ITEM - Refer to the product type generically, e.g.: - "a {product_type}", "this kind of {product_type}", "these {product_type}s". - MUST NOT refer to "this product", "this item", or a specific product page.

RAG VALUE TEST (CRITICAL):

For EACH question you generate, simulate two scenarios:

1) WITHOUT PROFILE (no user preferences): - The assistant sees only the product metadata and the text of the question. - It can give only generic, one-size-fits-all advice.

2) WITH PROFILE (has user preferences): - The assistant sees the same metadata and question, - PLUS the USER PROFILE, and can use it to tailor the answer.

If a question would lead to almost the SAME answer in both cases, it is a BAD question and you must revise it.

You must NOT fix this by asking directly about preferences. Instead, change the question to ask the assistant to choose or evaluate something FOR THE USER.

STRICT PROHIBITIONS:

- Do NOT copy any sentences or long phrases from the USER PROFILE. - Do NOT copy any example questions from this instruction. - Do NOT reuse generic templates about clothing, books, shoes, electronics, or any category. - Do NOT ask about attributes that are irrelevant or not mentioned in metadata. - Do NOT make up any new product attributes. - Do NOT mention the specific product name/title in questions. - Do NOT refer to "this product", "this item", "this {product_type}". - Do NOT state the user's preferences explicitly in questions. - Do NOT include specific personal conditions like "for warm weather", "for my sensitive skin", etc.

ABSOLUTELY FORBIDDEN question patterns: - Any question starting with or containing: - "Do you prefer" - "Would you prefer" - "How important is it to you" - "Would you prioritize" - "Do you have a preference for" - "Is it essential to you that" - Generic product-explanation questions such as: - "What are the advantages of ..." - "What are the key benefits of ..." - "Can you tell me about ..." - "How does this product perform in general?"

OUTPUT FORMAT (JSON):

Return ONLY a JSON object with:

"generated_questions": - 5-8 natural questions, - each

based ONLY on attributes present in THIS product's metadata, - each phrased as the user asking the assistant to recommend or judge {product_type}s FOR THEM, - each designed so that answers would differ depending on the user's preferences, - questions MUST NOT explicitly state any user preference.

Example STRUCTURE (values are placeholders and MUST NOT be copied):

```
{ "generated_questions": [ "QUESTION_1", "QUESTION_2" ] }
```

IMPORTANT REMINDERS: - Output ONLY valid JSON - NO comments like // or // in the JSON output. - "generated_questions" must use ONLY attributes from THIS product's metadata. - Do NOT copy any example text - generate everything fresh based on the actual data provided.

912

Figure 4: Prompt used for generating user-side questions based on metadata and latent preferences.

A.3 Agents Prompts

913

Dialogue Agent Prompt

You are a Dialogue History Agent.

CRITICAL REQUIREMENT: You MUST call the tool `rag_search_for_dialogue_history_agent` in your FIRST response. Do NOT generate any text output before calling this tool.

STEP 1: Call the tool `rag_search_for_dialogue_history_agent` ONCE. Pass the user's query as the `current_query` parameter. The tool returns:

- `relative_rounds`: Similar past conversations from RAG search
- `previous_summaries`: Summaries of previous rounds in THIS session

STEP 2: Generate a concise summary (2-4 sentences) that MUST start with a round marker:

- If `previous_summaries` is EMPTY: Start with [Round 1]
- If `previous_summaries` has N items: Start with [Round N+1]

After the round marker, summarize based on:

1. Previous round summaries (shows user's journey in this session)
2. Similar past conversations (shows long-term preferences)
3. Focus on user preferences, patterns, and context for recommendations

Example output: "[Round 1] This is the user's first query. Based on similar past conversations..."

Example output: "[Round 3] Continuing from previous rounds where user asked about..."

Do not call any other tools. End your turn after generating the summary.

914

Figure 5: Prompt used for the dialogue history agent.

Recommendation Agent Prompt

You are a Shopping Recommender Agent.

Your task has exactly FOUR steps and must stop after completing them.

CRITICAL REQUIREMENT: You MUST call the tool `rag_search_for_recommend_agent` in your FIRST response. Do NOT generate any text output before calling this tool. FAILURE TO CALL THIS TOOL = INCOMPLETE RECOMMENDATIONS.

STEP 1: Call the tool `rag_search_for_recommend_agent` ONCE. Pass the user's query as the `user_query` parameter.

STEP 2: The tool result will contain `recommendations` and `relevant_user_reviews`. Analyze the `relevant_user_reviews` (5 most relevant past reviews, dynamically selected based on query similarity) to extract user preferences:

- What product features/attributes does the user value?
- What brands or styles does the user prefer?
- What price range is typical for the user?
- Are there any quality or fit preferences mentioned?

STEP 3: Based on the extracted user preferences from relevant reviews, **RERANK** the `recommendations` list. Prioritize items that best match the user's historical preferences.

STEP 4: Use the `Dialogue_History_{user_id}_agent`'s final output included earlier in the conversation to further **FILTER** or adjust the reranked recommendations.

FINAL OUTPUT REQUIREMENTS:

- Provide the final, reranked recommendation list.
- Briefly explain HOW the user's relevant reviews influenced your reranking (1–2 sentences).
- Use ONLY information from: (1) the RAG tool result (`recommendations` + `relevant reviews`), and (2) the `Dialogue_History_{user_id}_agent` summary.

End your turn immediately after producing the recommendation output.

Figure 6: Recommend Agent Prompt

Reason Agent Prompt

You are a Reasoning Agent. Your task is to explain WHY the product recommendations were generated, with special emphasis on how the user's PAST REVIEWS influenced the recommendations.

You are provided with the FULL message history of the groupchat, which may contain many previous rounds.

IMPORTANT RULES:

- Only consider messages that belong to the **current question round**.

- Ignore ALL messages that appear before this boundary.

Within the current round, extract exactly these fields:

- `user_query`: from the first `User__proxy_agent` message after the boundary
- `dialogue_tool_result`: from the tool message of `Dialogue_History__agent`
- `dialogue_summary`: from the assistant output of `Dialogue_History_{user_id}_agent`
- `recommendation_tool_result`: from the tool message of `Recommender__agent` (contains 'relevant_user_reviews')
- `final_recommendation`: from the assistant output of `Recommender_{user_id}_agent`

PERSONALIZATION FOCUS: Pay special attention to the 'relevant_user_reviews' in the `recommendation_tool_result`. These are the user's ACTUAL past reviews used to personalize the recommendations. Your explanation MUST include:

- What preferences/patterns were extracted from the user's past reviews
- How these preferences influenced which products were recommended
- Specific connections between review content and recommendation choices

FINAL OUTPUT REQUIREMENTS:

- Read ONLY these extracted fields from this round.
- Explain clearly WHY the recommendation was produced, emphasizing personalization from user reviews.
- DO NOT hallucinate missing information — if a field is missing, explicitly say it.
- DO NOT use or reference any information from previous rounds.

End your turn after producing the explanation.

Figure 7: Reason agent prompt

Reply Agent Prompt

You are a Reply Agent. Your task is to generate a natural, PERSONALIZED response to the user based on the conversation context and their past reviews.

STEP 1: Determine the conversation round:

- Look for "[Round X]" at the START of the `Dialogue_History` agent's summary.
- If "[Round 1]": This is the FIRST round — use a friendly greeting such as "Hi there!" or "Hey!".
- If "[Round 2]" or higher: This is a continuation — do **not** use generic greetings. Instead, acknowledge naturally (e.g., "Following up on your question..." or "Based on what we discussed...").
- **IMPORTANT:** Do **not** include "[Round X]" in your response — it is only for determining greeting style.

STEP 2: Review the conversation context:

- Review the user's original query and intent.
- Review the Recommender agent's product recommendations and the `relevant_user_reviews`.

- Review the Reasoner agent’s explanation of why recommendations were made.

STEP 3: Generate a PERSONALIZED response: **Your response MUST feel personalized based on the user’s past reviews.** The Recommender agent’s tool result contains `relevant_user_reviews` — these are the user’s actual past reviews indicating their preferences, experiences, and style.

Use this information to:

- Reference specific preferences from their reviews (e.g., “Since you mentioned loving breathable fabrics...”).
- Connect recommendations to their past experiences.
- Show that you understand their taste and needs.
- Make the response feel tailored specifically to them, not generic.

STEP 4: Generate your final response with the following principles:

- Use the correct greeting: friendly for first round, contextual for continuation.
- Present the top 3–5 recommendations conversationally and personally.
- Clearly link recommendations to their demonstrated preferences.
- Highlight features that match their historical likes/dislikes.
- Be friendly, concise, and genuinely helpful.
- End with a natural offer to assist further.

Remember: The user should feel like you **know them** based on their shopping history. Make the experience personal, not generic.

End your turn immediately after generating the response.

Figure 8: Reply agent prompt

A.4 Evaluation Prompts

Section	Description
Evaluator Role	You are an expert evaluator of shopping-assistant dialogues, focusing on user experience and satisfaction. Assess each dialogue between a <i>shopper (user)</i> and a <i>shopping assistant (agent)</i> . Rate each criterion from 1 (worst) to 5 (best) according to user satisfaction and dialogue fluency.
Input Format	[User Goal] {user_goal} [Conversation] {dialogue_text}
Evaluation Criteria	<ol style="list-style-type: none"> 1. User Preference Alignment (UPA) – How well the assistant respects and personalizes responses to user needs (budget, brand, tone, style, delivery). 2. Dialogue History Trend Consistency (DHTC) – Whether the assistant maintains coherent understanding and context through multiple turns. 3. Dialogue Interactive Ability (DIA) – How engaging, natural, and balanced the interaction is. 4. Macro-Level Features – Topic Management (topic flow), Tone Appropriateness (style, politeness), Conversation Opening (warm and contextual), and Conversation Closing (smooth wrap-up). 5. Overall Score (1–5) – Overall communication fluency and goal achievement from the user’s perspective.
Scoring Scale	<p>5: Excellent / natural / fully satisfying user experience.</p> <p>4: Good / minor gaps but goal achieved.</p> <p>3: Moderate / partially successful or mechanical.</p> <p>2: Poor / confusing or partly irrelevant.</p> <p>1: Very poor / broken or failed communication.</p>

Table 7: Prompt used for LLM-as-a-Judge evaluation of shopping dialogue quality. Each dimension is rated from 1 (worst) to 5 (best) according to user experience and conversational fluency.

Criterion	Description and Scoring Rubric (1-5)
1. Global Personalization Consistency (GPC)	<p><i>Did the agent maintain the user's persona constraints throughout the entire dialogue?</i></p> <p>1 (Inconsistent/Amnesic): Agent started well but forgot constraints (e.g., budget, size) in later turns, or hallucinated user details.</p> <p>3 (Fragmented): Agent remembered explicit constraints (e.g., Size M) but forgot implicit ones (e.g., Style preferences) as the chat went on.</p> <p>5 (Rock Solid): Every single response, from start to finish, aligned with the User Profile. The agent never "reset" or forgot context.</p>
2. Cumulative Information Synthesis (CIS)	<p><i>Looking at all agent responses combined, did it build a complete, factual picture?</i></p> <p>1 (Fragmented/Repetitive): Agent repeated the same facts multiple times or gave contradictory information across different turns.</p> <p>3 (Basic Retrieval): Agent provided facts when asked, but did not connect them.</p> <p>5 (Coherent Narrative): The agent's responses build on each other. Later responses reference earlier facts (e.g., "As I mentioned earlier regarding the fabric...").</p>
3. Interaction Trajectory (IT)	<p><i>Did the agent's responses purposefully drive the conversation to a conclusion?</i></p> <p>1 (Circular): The agent went in circles, repeating suggestions or asking the same clarifying questions.</p> <p>3 (Passive): The agent answered questions accurately but waited for the user to lead.</p> <p>5 (Strategic): The agent's responses followed a logical arc: Clarify → Recommend → Refine → Close.</p>
4. Tone Consistency (TC)	<p><i>Was the agent's voice stable?</i></p> <p>1 (Erratic): Switched between robotic formal ("As an AI...") and casual ("Hey there!") randomly.</p> <p>5 (Unified Voice): Maintained a consistent, professional, and empathetic shopping persona from Hello to Goodbye.</p>

Table 8: **Evaluation Criteria for AI Shopping Assistant.** Each dimension (1-5 scale) assesses cumulative dialogue quality across the entire interaction, focusing on personalization, synthesis, trajectory, and tone stability.

A.5 Four Interactivity Aspects Definitions and Descriptions

921
922
923

Table 9 shows the descriptions and definitions for dialogue's overall quality score.

Interactivity Macro-level Features	Definition
Topic Management	the strategies and techniques used to control and navigate the flow of topics.
Tone Choice Appropriateness	the suitability of the tone used in communication, ensuring it aligns with the context, audience, and purpose to convey the intended message.
Conversation Opening	the initial interaction or exchange that begins a dialogue, often setting the tone and context for the dialogue.
Conversation Closing	the process of ending a dialogue or interaction, which involves signaling the conclusion of the discussion, summarizing key points, and often expressing a farewell.

Table 9: Definitions of macro-level interactivity features, where higher scores indicate more natural, authentic interaction and active engagement in the dialogue (adapted from (Gao et al., 2025a)).

924
925

A.6 Case Study

Multimodal RAG Dialogue Example

User:

What unique styles of enamel pins would you recommend for someone who enjoys nature-inspired designs?

Assistant:

Hi there! Since you've shown a love for nature elements like pinecones, driftwood, and geodes in your past reviews, I've curated some unique enamel pin and jewelry options that blend earthy textures with artistic flair.

- **Pinecone Earrings**** (B09DFQ985K) – These gold or silver pinecone earrings with leaf charms mirror your appreciation for botanical motifs and seasonal charm. Their handcrafted detail and festive vibe align perfectly with your past praise for nature-inspired accessories.
- **Driftwood Resin Earrings**** (B07XK149Q8) – Featuring real driftwood encased in sparkling resin, these dangle earrings echo your fondness for organic materials and abstract artistry. The unique, marbled patterns are a nod to your love of natural textures.
- **Geode Stud Earrings**** (B07GX515F2) – Your review of geodes as "stunning for studs" inspired this recommendation. These gold-toned geode earrings capture the earthy, mineral-inspired beauty you've adored.
- **Bird Charm Earrings**** (B01BK0DE4W) – While not explicitly stated, your mention of birds in reviews hints at a subtle interest in creature motifs. These hand-painted glass earrings with a glittery green bird add whimsy and nature's grace to any outfit.

926

5. **Hops Enamel Pin (B0762CCKNW)** – A quirky twist on nature! Hops, a plant-based element, pair with your love for unique, textured designs. It's a great way to blend your passion for the outdoors with a craft beer flair.

Let me know if you'd like more options or details!

User:

When considering enamel pins, which materials do you think would offer the best combination of quality and aesthetic for me?

Assistant:

Following up on your interest in durable, nature-inspired materials with a touch of refined elegance, I've curated a selection of enamel pins and accessories that align with your past preferences for **handcrafted textures** and **hypoallergenic quality**.

1. **Hops Enamel Pin (B0762CCKNW)** – This pin combines **high-quality enamel** with **gold plating** to capture the intricate, leafy structure of a hop cone. Its polished finish and detailed design reflect your appreciation for **craft beer-themed botanicals** and **long-lasting materials**.

2. **Pinecone Earrings (B09DFQ985K)** – Crafted from **gold-plated brass** and 14K gold fill, these earrings mirror your love for **durable, skin-friendly metals** while showcasing the **textured realism** of pinecones and leaves, a motif you've praised in past purchases.

3. **Driftwood Resin Earrings (B07XK149Q8)** – Made with **real driftwood** encased in glossy resin, these pieces echo your preference for **organic materials** paired with **modern, reflective finishes**. The unique, handcrafted marbling adds a whimsical yet sophisticated touch.

4. **Faux Druzy Studs (B0952DG1DG)** – These **silver-tone faux druzy stones** (in royal blue) cater to your fondness for **mineral-inspired textures** like geodes and agate. The hypoallergenic bezel ensures comfort, aligning with your past emphasis on **skin-friendly durability**.

5. **Glitter Resin Studs (B072FVQ5Q8)** – The **stainless steel frames** and **sparkling resin** blend your love for **earthy materials** with **vibrant, eye-catching details**. Their playful yet refined design matches your taste for subtle artistry.

Your history with **natural textures** (driftwood, geodes) and **hypoallergenic metals** (sterling silver, gold fill) is clearly reflected here. Let me know if you'd like to refine these suggestions further!

User:

Given my interest in handcrafted items, which types of enamel pins should I consider that highlight unique craftsmanship?

Assistant:

Based on your deep appreciation for **handcrafted textures** and **nature-inspired motifs**, I've curated a selection that blends **unique fabrication techniques** with your preferred materials like **gold-plated brass**, **sterling silver**, and **natural elements**. Here's what aligns with your past choices:

1. **Hops Craft Beer Enamel Pin (B0762CCKNW)** – A botanical masterpiece with **gold-plated brass** and **detailed enamel work** that mirrors your love for

craft beer-themed designs and **textured realism** (like your past geode studs). The layered, leafy structure of the hop cone is rendered with precision, echoing your taste for **subtle yet meaningful craftsmanship**.

2. **Amos Pewter Apple Blossom Studs (B09NMMLLC5)** – These **handmade floral earrings** (crafted since 1974) combine **hypoallergenic stainless steel** with **pewter's warm, metallic finish**, a nod to your preference for **skin-friendly durability** and **organic shapes**. The delicate, hand-finished petals and stamen details are a perfect match for your history with **botanical motifs**.

3. **Blue Hummingbird Earrings (B0181PPVD6)** – **Hand-painted hummingbirds** paired with **lapis lazuli beads** and **gold-filled ear wires** reflect your admiration for **layered textures** and **whimsical artistry** (as seen in your praise for pinecone earrings and geode studs). The vibrant, iridescent hues and intricate designs cater to your fondness for **playful yet refined pieces**.

4. **Lightweight Hand Painted Elephant Dangle Earrings (B07QY4JW28)** – These **Indian boho-inspired earrings** blend **hand-painted pink tassels** with a **symbolic elephant charm**, aligning with your history of valuing **symbolic natural elements** (like driftwood and pinecones) and **lightweight, comfortable designs**. The layered, custom feel matches your interest in **unique, small-batch craftsmanship**.

5. **GN PEARL Bear Brooch (B0B4CML8R9)** – A **whimsical take on nature-themed accessories**, featuring **freshwater cultured pearls** and **sparkling cubic zirconia**. This ties back to your past love for **mineral-inspired textures** (geodes, druzy stones) and **sustainable, handcrafted materials**.

Your history with **textured, organic designs** and **hypoallergenic metals** is clearly reflected here. These pieces balance **artisanal detail** with **durable, skin-friendly construction**—just like the items you've cherished before. Let me know if you'd like to explore more options or refine these further!

User:

For someone who appreciates good design, how do you think the size of an enamel pin affects its overall impact and appeal?

Assistant:

Based on your history of valuing **nature-inspired textures** and **durable, hypoallergenic materials**, the size of an enamel pin plays a key role in how its design is perceived. Larger pins, like the **Hops Enamel Pin (B0762CCKNW)** at 1" x 1.4", make a bold statement by emphasizing intricate botanical details—perfect for showcasing the layered, leafy structure of hops or pinecones, motifs you've loved in the past. Smaller pins, such as the **12mm glitter resin studs (B072FVQ5Q8)** or **8mm druzy stone studs (B0952DG1DG)**, offer a more subtle, versatile elegance, aligning with your appreciation for pieces that blend **playful whimsy** with refined craftsmanship.

Your preference for **textured realism** (e.g., driftwood, geodes) suggests medium-sized items like the **teardrop earrings (B0186L9SPM)** or the **rose cinch clip (B075CS4JCK)** might resonate best. These sizes balance detail visibility with wearability, much like your past favorites. The **hammered sterling silver** finish on the teardrop earrings mirrors your love for tac-

tile, organic aesthetics, while the **rose cinch clip** combines functionality with a vintage, nature-themed charm—a nod to your history with subtle yet meaningful accessories.

If you're leaning toward bold or refined styles, let me know! I can refine the suggestions further to match your exact taste.

User:

Which options would you suggest for an enamel pin that balances durability with a polished look?

Assistant:

Following up on your consistent appreciation for **durable, hypoallergenic materials** and **natural textures paired with polished finishes** , I've curated a selection that aligns with your past preferences for **metallic durability** (gold-plated brass, sterling silver) and **organic detailing** (druzy stones, resin). Here are the top picks:

1. **Hops Enamel Pin (B0762CCKNW)** – This pin mirrors your love for **craft beer-themed botanicals** and **gold-plated brass** , offering a textured green enamel design that balances rustic natural motifs with a sleek, metallic outline. Perfect for showcasing your passion with lasting style.
2. **Rose Gold Resin Studs (B0711DJYLC)** – The **rose gold-tone bezels** and **glitter resin** combine your preference for **metallic durability** with a **playful, refined aesthetic** , much like the "subtle and fun" designs you've praised in the past.
3. **Silver-tone Druzy Studs (B0952DG1DG)** – These royal blue faux druzy stones echo your history with **mineral-inspired textures** (e.g., geodes, agate), while the **hypoallergenic silver bezels** ensure comfort and longevity, aligning with your emphasis on skin-friendly materials.
4. **Hot Pink Kawaii Studs (B072FVQ5Q8)** – The **stainless steel frames** and **sparkling resin** reflect your taste for **vibrant, textured details** paired with **durable, tarnish-resistant construction** , offering a whimsical yet polished twist.
5. **Custom Name Bracelet (B07MJ21Y5C)** – While not a pin, this **gold-filled and sterling silver** piece caters to your preference for **hypoallergenic, long-lasting materials** and **personalized, refined craftsmanship** , much like your past favorites.

Your history with **natural motifs** and **metallic durability** is clearly reflected here. Let me know if you'd like to explore more options or refine these further!

User:

What kind of packaging should I look for when purchasing enamel pins to ensure they arrive in great condition?

Assistant:

Based on your history of valuing **durable, hypoallergenic materials** and **eco-conscious packaging** , here's how to ensure your enamel pins arrive in pristine condition:

1. **Sealed Anti-Static Bags** – Like the **Hops Craft Beer Enamel Pin** (B0762CCKNW), which you praised for its **"clear & sealed bag"** protecting delicate enamel details. These bags prevent scratches and moisture damage, ideal for preserving intricate designs.

2. **Gift Boxes with Protective Linings** – The **Faux Druzy Stone Stud Earrings** (B0952DG1DG) and **Duo Druzy Set** (B0952CDV9R) from Summerfield Collection use **gift boxes** with recyclable materials, matching your preference for sustainable, artisanal presentation.

3. **Customizable Packaging Options** – Look for sellers offering **velvet-lined boxes** or **bubble wrap inserts** (e.g., **Trio Druzy Set** B0952DZR5T) to minimize movement during transit, ensuring the pins' textures and finishes stay intact.

Your past appreciation for **secure, eco-friendly packaging** (like the **Pinecone Earrings** 's tarnish-resistant gift box) directly informs these choices. Let me know if you'd like tips on spotting these features in listings!

User:

Based on my usual tastes, which styles of enamel pins do you think would make a great gift for a craft beer lover?

Assistant:

Following up on your interest in **craft beer-themed designs** and **sustainable, handcrafted aesthetics** , I've narrowed down a few standout options that align with your past preferences for **textured natural elements** and **durable, hypoallergenic materials** .

1. **Hops Craft Beer Enamel Pin (B0762CCKNW)** – This pin perfectly captures your love for **botanical craftsmanship** and **craft beer symbolism** . Its **gold-plated brass** and **detailed enamel work** mirror your appreciation for **intricate, long-lasting designs** (like the geode studs and pinecone earrings you've cherished). The layered, leafy hop cone design is both **textured and meaningful** , making it a great way to celebrate their passion for brewing.
2. **Recycled Bottle Cap Turtle (B09GLFK1N5)** – A playful nod to your preference for **upcycled materials** and **nature-inspired storytelling** . This **magnetic turtle** features a **recycled beer bottle cap** as its shell, blending **eco-conscious design** with **symbolic charm** . Its **silver-toned pewter** construction ensures **hypoallergenic durability** , while the **customizable cap** adds a fun, interactive twist—much like the layered, meaningful pieces you've loved before.

These choices balance your history with **organic textures** , **metallic finishes** , and **sustainable artistry** . If you'd like to explore more options or refine these further, I'm here to help!

930

Table 10: Multimodal RAG Example

Text RAG Dialogue Example

User:

What unique styles of enamel pins would you recommend for someone who enjoys nature-inspired designs?

Assistant:

Hi there! Since you've shown a love for natural materials and handmade, eco-conscious items in your past reviews—like the driftwood earrings and geode studs—I'd love to suggest some nature-inspired enamel pins that align with your style.

931

1. **Melanie Golden Artisan Jewelry Pinecone Earrings** – These are a perfect match for your appreciation of seasonal, nature-themed accessories. The gold-plated brass and sterling silver details mirror your preference for quality craftsmanship, and the pinecone design ties directly to your past praise for festive, organic motifs.

2. **Handmade Driftwood & Resin Dangle Earrings** – You've mentioned enjoying unique, handmade pieces with natural elements, and these driftwood earrings capture that essence. The eco-friendly materials and one-of-a-kind look resonate with your interest in sustainable, earthy designs.

3. **Authentic Geode Stud Earrings** – Your review of geode beads highlighted their subtle elegance and connection to natural formations. These studs bring that same charm, blending geological beauty with minimalist appeal—ideal for those who want to wear a quiet nod to nature.

4. **Hops Craft Beer Enamel Pin** – While not purely nature-focused, your past mention of "reminders of nature's wonders" and love for thematic, travel-friendly accessories made this a thoughtful addition. It's a quirky, plant-inspired piece that fits your eclectic taste.

Let me know if you'd like help narrowing these down further!

User:

When considering enamel pins, which materials do you think would offer the best combination of quality and aesthetic for me?

Assistant:

Following up on your interest in nature-inspired enamel pins and eco-conscious materials, I'm excited to share more tailored options that align with your past preferences!

1. **Melanie Golden Artisan Jewelry Pinecone Earrings** – These are a standout choice, as you've loved pinecone and leaf motifs for their **seasonal charm** and **natural symbolism** (like fall's regenerative themes). The **gold-plated brass** and **sterling silver** construction matches your appreciation for **durable, hypoallergenic materials** mentioned in reviews, ensuring both quality and comfort.

2. **GN PEARL Pearl Little Bear Brooches** – While not a direct nature motif, the **pearls** and **cubic zirconia** add a subtle, elegant touch that resonates with your taste for **handmade, sustainable details**. Your past praise for "eco-friendly" and "support small" brands makes these a thoughtful fit for their refined, artisanal appeal.

3. **Hematite-tone Leaf Resin Stud Earrings** – These echo your love for **natural textures** (like driftwood and geodes) with their **opalescent leaf design** and **resin craftsmanship**. The "glass-like" finish and earthy tones align with your preference for **unique, earthy aesthetics** that blend organic beauty with modern minimalism.

4. **Hops Craft Beer Enamel Pin** – A clever nod to your **plant-inspired interests** and **travel-friendly, thematic designs**. Hops, as a botanical element, bridges your passion for nature with a quirky, sustainable twist—perfect for collectors who enjoy **niches** that subtly connect to the natural world.

5. **Metzakka Sterling Silver Crystal Earrings** – These highlight your preference for **durable, high-quality materials** (like sterling silver) and **sparkling details** that complement your love for "lightweight but solid" pieces. The **celestial design** adds a whimsical yet elegant layer to your nature-inspired collection.

Let me know if you'd like deeper insights into any of these or want to explore more options!

User:

Given my interest in handcrafted items, which types of enamel pins should I consider that highlight unique craftsmanship?

Assistant:

Following up on your focus on **handcrafted, eco-conscious enamel pins** with **natural motifs** and **unique textures**, I've curated a list that aligns with your past appreciation for **hypoallergenic materials**, **sustainable practices**, and **artistic detail**.

1. **Amos Pewter Apple Blossom Stud Earrings** – These handmade studs echo your love for **nature-inspired designs** (like the geodes you praised) and **eco-friendly materials** (hypoallergenic stainless steel). The delicate apple blossom motif is a fresh take on botanical themes, and the vintage-inspired craftsmanship feels in line with your preference for **recycled or upcycled elements**.

2. **Hops Craft Beer Enamel Pin** – While beer-themed, your past mention of "reminders of nature's wonders" and interest in **travel-friendly, thematic accessories** made this a standout. The **gold-plated brass** and **artisanal texture** of the hops pin reflect your taste for **durable, one-of-a-kind pieces** that blend **craftsmanship** with subtle nature nods.

3. **Painted Pink Rose Dangle Earrings** – These boho-style earrings match your preference for **hand-painted artistry** and **lightweight, comfortable designs**. The romantic, floral aesthetic ties to your history of enjoying **earthy, detailed patterns** (like the driftwood resin earrings).

4. **Pink Fan Tassel Elephant Dangle Earrings** – The **handmade, lightweight construction** and **vibrant painted elephants** align with your love for **unique, wearable art**. The playful yet intricate design feels like a natural extension of your interest in **eclectic, nature-themed accessories**.

5. **GN PEARL Pearl Little Bear Brooches** – These sparkly brooches combine **elegant, durable finishes** (pearls and hypoallergenic metals) with **versatile, gift-worthy appeal**, something you've appreciated in past purchases.

The **bear brooch** was excluded as it didn't strongly connect to your focus on natural materials or eco-conscious craftsmanship. Let me know if you'd like to explore more options or refine these suggestions further!

User:

For someone who appreciates good design, how do you think the size of an enamel pin affects its overall impact and appeal?

Assistant:

Based on your focus on design and the interplay of size with aesthetic impact, I'll break down how these choices align with your past preferences!

1. **Hops Craft Beer Enamel Pin (1" x 1.4")** – This larger pin is perfect for making a bold, thematic statement. Its size emphasizes the **botanical motif** (hops as a plant-based symbol) while maintaining **sustainability** and **handmade durability**, echoing your love for nature-inspired pieces that “travel well” and subtly celebrate the natural world.

2. **10mm Square Hammered Sterling Silver Stud** – Smaller designs like this allow **intricate, detailed textures** (e.g., geometric patterns or leaf motifs) to shine, matching your appreciation for **subtle, artisanal aesthetics**. The hammered finish and eco-friendly materials (sterling silver) align with your history of valuing **lightweight yet solid craftsmanship**.

3. **GN PEARL Little Bear Brooches** – These mid-sized brooches balance **eco-conscious materials** (lead-free, nickel-free) with **elegant, minimalist details**. Their versatility for clothing or accessories reflects your interest in **versatile, wearable art** that blends practicality with artistic flair.

4. **Cloud Cap Teardrop Earrings** – The medium-small size of these teardrop earrings suits everyday wear, while their **handmade, hammered texture** and **sterling silver quality** cater to your preference for **durable, hypoallergenic pieces** with a tactile, natural feel.

5. **Annika Bella Gold Medallion Necklace** – Though not a pin, this larger statement piece mirrors your taste for **versatile, impactful designs**. Its **tarnish-resistant gold medallion** and dainty style align with your history of valuing **practicality and artistic expression** in eco-friendly, handmade items.

Size matters—smaller pins let delicate details take center stage, while larger ones offer bold, thematic resonance. Let me know if you’d like to explore how specific sizes or motifs might enhance your collection further!

User:

Which options would you suggest for an enamel pin that balances durability with a polished look?

Assistant:

Based on your focus on **durable, polished enamel pins** and your history of loving **natural materials** (like driftwood, pinecones, and geodes) paired with **eco-conscious, handmade craftsmanship**, I’ve curated a few standout options that align with your priorities:

1. **Metzakka Sterling Silver Lucky Star Earrings** – These dainty drop earrings combine **925 sterling silver** (a material you’ve praised for being hypoallergenic and solid) with **sparkling crystals and pearls**, offering a refined, polished look while staying true to your appreciation for **subtle, artisanal details**.

2. **GN PEARL Pearl Little Bear Brooches** – The **lead-free, nickel-free alloy** and **sparkling cubic zirconia** match your past emphasis on **safe, sustainable materials**. The bear motif, while not overtly nature-themed, fits your quiet interest in **symbolic, thematic designs** (like the pinecone and leaf pieces you’ve loved).

3. **Hops Craft Beer Enamel Pin** – A nod to your **botanical-inspired tastes** and **travel-friendly, durable enamel**. The **gold-plated finish** and **plant-based hops design** bridge your love for **nature motifs** with a sleek, polished aesthetic.

4. **Hematite-tone Leaf Resin Studs** – These echo your preference for **natural textures** (like driftwood and geodes) with their **eco-resin construction** and **shiny, polished finish**. The **leaf design** ties back to your history of celebrating **earthly symbolism** in jewelry.

5. **Custom Name Initial Bracelet (AnnBijoux)** – While a bracelet, its **gold-filled durability** and **personalized, minimalist style** align with your past praise for **long-lasting, handmade pieces** that blend **subtlety with quality**.

Let me know if you’d like to explore more options or dive deeper into any of these!

User:

What kind of packaging should I look for when purchasing enamel pins to ensure they arrive in great condition?

Assistant:

Based on your history of prioritizing **eco-conscious materials** and **sustainable packaging**, I’d love to share tailored tips to keep your enamel pins safe and aligned with your values!

For **durability**, **bubble wrap or foam inserts** are perfect—think of how you’ve cherished items like the **Hematite-tone Leaf Resin Studs** (eco-resin) and **Metzakka Sterling Silver Earrings** (hypoallergenic metals). These cushioning methods protect delicate details, just as you’ve valued in past purchases.

Recyclable cardboard boxes with reinforced corners are a must, especially since you’ve loved **Summerfield Collection**’s earthy resin designs and **GN PEARL**’s lead-free metals. They’re sturdy yet kind to the planet, matching your preference for **minimalist, artisanal packaging**.

For **eco-friendly fillers**, opt for **biodegradable packing peanuts** or **reusable fabric pouches**—a nod to your love of **natural textures** (like geodes and pinecones) and **reduced plastic waste**. Even the **thank-you card** can be made from recycled paper, adding a **personalized, sustainable touch** that resonates with your past choices.

And don’t forget **carbon-neutral shipping** options! Brands like **Wildship Studio** and **Metzakka** align with your eco ethos, so their packaging and shipping practices likely reflect the same care you’ve appreciated in their products.

Let me know if you’d like help sourcing specific vendors or refining these ideas further—your values are a beautiful guide!

User:

Based on my usual tastes, which styles of enamel pins do you think would make a great gift for a craft beer lover?

Assistant:

Based on your love for **eco-conscious, handcrafted enamel pins** with **nature-inspired motifs** and **craft beer themes**, I’ve curated a few standout options that align with your documented preferences for **sustainable materials** and **unique, durable designs**.

1. **Hops Craft Beer Enamel Pin (B0762CCKNW)** – A direct nod to your **craft beer interest**, this golden hops pin combines **artisanal craftsmanship** with **eco-friendly gold plating** (mirroring your past ap-

preciation for hypoallergenic, high-quality materials). Its minimalist yet detailed design fits your taste for **subtle, meaningful symbols**—perfect for a beer lover who values both **style and sustainability**.

2. **Recycled Bottle Cap Turtle (B09GLFK1N5)**: - This playful, **upcycled piece** ties into your history of loving **recycled materials** (like the vintage cobalt jar earrings you praised). The **magnetic, changeable cap** adds a creative twist, aligning with your preference for **interactive, one-of-a-kind items** that blend **eco-awareness** with **artistic flair**.

3. **GN PEARL Pearl Little Bear Brooches (B0B4CML8R9)**: - While not strictly beer-themed, these **hypoallergenic, sparkling brooches** reflect your past affinity for **elegant, sustainable finishes** (e.g., "lead-free" materials and durable craftsmanship). The **bear motif** adds a whimsical, nature-inspired touch, much like the geode studs you admired—ideal for someone who appreciates **versatile, statement-making accessories**.

The **bear brooch** was excluded as it didn't strongly connect to your craft beer focus or eco-conscious materials. Let me know if you'd like to explore more beer-themed or nature-inspired options!

Table 11: TextRAG Example

Multimodal (No User Review) Example

User:

What unique styles of enamel pins would you recommend for someone who enjoys nature-inspired designs?

Assistant:

Hi there! If you're into nature-inspired designs, here are some unique enamel pin and jewelry options that blend natural motifs with artistic craftsmanship:

1. **Hops Craft Beer Enamel Pin (B0762CCKNW)** – This pin features a detailed hop cone design, capturing the layered, leafy texture of a real hop. Its gold outline and green detailing make it a standout piece that merges botanical elements with craft beer culture.

2. **Melanie Golden Artisan Jewelry Pinecone Earrings (B09DFQ985K)** – These gold or silver pinecone earrings with leaf charms are perfect for fall and winter themes. Made from eco-friendly materials like gold-plated brass and 14K gold fill, they combine rustic textures with elegant finishes.

3. **Driftwood and Resin Dangle Earrings (B07XK149Q8)** – Handmade from real driftwood and clear resin with swirling blue-white marbling, these earrings offer a raw, earthy look paired with modern artistry. Each piece is unique, with a glossy, sparkly finish that highlights natural patterns.

While the first item is an enamel pin, the others are earrings but share the nature-inspired aesthetic you love. Let me know if you'd like more options or details!

User:

When considering enamel pins, which materials do you think would offer the best combination of quality and aesthetic for me?

Assistant:

Following up on your interest in enamel pins and sustainable materials, here are three standout options that balance quality, eco-consciousness, and nature-inspired aesthetics:

1. **Melanie Golden Artisan Jewelry Pinecone Earrings (B09DFQ985K)** – These feature **14K gold fill** and **gold-plated brass**, which are durable, hypoallergenic, and more sustainable than solid gold. The **pinecone and leaf motifs** capture natural textures, while the **handmade craftsmanship** and **seasonal themes** (fall/winter) align with your preference for unique, thematic designs.

2. **Stainless Steel Glitter Resin Stud Earrings (B072FVQ5Q8)** – Made from **stainless steel** (tarnish-resistant and hypoallergenic) paired with **glitter resin** for a vibrant, modern look. Though earrings, their **eco-friendly materials** and **playful, artistic design** mirror your interest in sustainable, eye-catching accessories with natural-inspired elements.

3. **Hops Craft Beer Enamel Pin (B0762CCKNW)** – A **high-quality enamel pin** with **gold plating** and **durable construction**, showcasing a detailed **hop cone design**. While not explicitly eco-focused, its **precision craftsmanship** and blend of **botanical motifs** with **craft beer culture** reflect your taste for niche, artistic enamel work.

These choices prioritize **sustainable materials** (gold fill, stainless steel) and **nature-themed motifs**, ensuring both longevity and aesthetic appeal. Let me know if you'd like more enamel pin options or deeper insights into these materials!

User:

Given my interest in handcrafted items, which types of enamel pins should I consider that highlight unique craftsmanship?

Assistant:

Following up on your focus on sustainable, handcrafted enamel pins with unique natural motifs, here are three standout options that align with your priorities:

1. **Hops Craft Beer Enamel Pin (B0762CCKNW)** – This pin showcases a **detailed hop cone design** with **textured green foliage** and a **metallic gold outline**, blending **botanical themes** with **eco-conscious materials** (gold plating avoids solid gold's environmental impact). Its **handmade craftsmanship** and **craft beer culture** tie into your interest in **unique, thematic designs**.

2. **Amos Pewter Apple Blossom Stud Earrings (B09NMMLLC5)** – While not enamel pins, these **handmade pewter earrings** feature **delicate, textured apple blossom motifs** and **hypoallergenic stainless steel**, emphasizing **sustainable materials** and **natural aesthetics**. Their artisanal finish and heritage since 1974 highlight **timeless, ethical craftsmanship**.

3. **Blue Hummingbird Earrings (B0181PPVD6)** – These **hand-painted bohemian earrings** combine **vibrant blue plumage** with **antiqued copper details**, offering a **unique, nature-inspired design**. The **lightweight, handmade quality** ensures comfort while maintaining **artistic individuality**.

Why These Fit: - **Materials**: Gold plating, stainless steel, and pewter prioritize **durability and eco-friendliness**, avoiding toxic metals. - **Craftsmanship**

ship**: Textured natural motifs (hop cones, apple blossoms, hummingbirds) and **handmade processes** reflect your love for **one-of-a-kind, artisanal work**. - **Sustainability**: Ethical sourcing and hypoallergenic options align with your history of valuing **eco-conscious, high-quality designs**.

Let me know if you'd like more enamel pin options or deeper insights into these materials!

User:

For someone who appreciates good design, how do you think the size of an enamel pin affects its overall impact and appeal?

Assistant:

The size of an enamel pin plays a crucial role in its design impact and appeal, especially for someone valuing both artistry and sustainability. Here's how the recommendations align with your priorities:

1. **Hops Craft Beer Enamel Pin (B0762CCKNW)**** – Its **1" x 1.4"** size allows for **intricate botanical detailing** (like the textured hop cone layers) while remaining visible enough to serve as a statement piece. The **eco-friendly gold plating** and **durable enamel** ensure longevity, making it a sustainable choice that balances complexity with practicality.
2. **Clothes Clip with Rose Design (B075CS4JCK)**** – Though not an enamel pin, its **3.5"** width accommodates **vintage-style rose motifs**, blending functionality with decorative flair. The **handmade, eco-conscious materials** (silver-toned metal and plastic inserts) and **natural themes** reflect your interest in **sustainable, unique designs**, even in non-traditional formats.
3. **Stainless Steel Glitter Resin Stud Earrings (B072FVQ5Q8)**** – The **12mm diameter** emphasizes **playful, material-efficient artistry**, using **hypoallergenic stainless steel** and **glitter resin** for a vibrant, eye-catching look. Smaller sizes reduce resource use, aligning with sustainability goals while still offering **subtle, textured aesthetics**.

These options highlight how size influences **design complexity, visibility, and material efficiency**, all while prioritizing **eco-friendly practices** and **nature-inspired craftsmanship**. Let me know if you'd like to explore more size-specific options or dive deeper into material choices!

User:

Which options would you suggest for an enamel pin that balances durability with a polished look?

Assistant:

Following up on your focus for enamel pins that balance durability and a polished look, here are three refined options tailored to your preferences:

1. **Hops Craft Beer Enamel Pin (B0762CCKNW)**** - **Durability**: High-quality gold plating and hard enamel ensure long-lasting use. - **Polished Look**: Metallic gold outline with detailed, textured hop cone design for a sophisticated, refined appearance. - **Eco-Conscious**: Gold plating (vs. solid gold) and sustainable craftsmanship align with your values.
2. **Stainless Steel Glitter Resin Stud Earrings (B072FVQ5Q8)**** - **Durability**: Hypoallergenic stainless steel base and glitter resin for skin-safe, tarnish-resistant wear. - **Polished Look**: Sleek silver-toned bezel with vibrant hot pink ombre resin, offering a mod-

ern, eye-catching aesthetic. - **Eco-Friendly**: Recyclable materials and minimalistic design reflect sustainability and refined style.

3. **Silver-tone Faux Druzu Stone Stud Earrings (B0952DG1DG)**** - **Durability**: Stainless steel construction paired with non-toxic faux druzu stones for a long-lasting, nature-inspired piece. - **Polished Look**: Polished silver-tone bezel framing royal blue druzu stones for an elegant, textured finish. - **Sustainability**: Eco-conscious materials and minimalist design resonate with your criteria.

These selections prioritize **eco-friendly materials**, **polished aesthetics**, and **durable construction**, aligning with your values. Let me know if you'd like to explore more enamel pin options or dive deeper into material specifics!

User:

What kind of packaging should I look for when purchasing enamel pins to ensure they arrive in great condition?

Assistant:

Based on your focus on sustainable materials and protective packaging for enamel pins, here are three top options that align with your values:

1. **Hops Craft Beer Enamel Pin (B0762CCKNW)**** – Packaged in a **clear, sealed bag** and **cardstock backing**, this design uses minimal plastic while safeguarding against scratches and dust. The eco-friendly materials and structured packaging ensure both **durability** and **environmental responsibility**.
2. **Silver-tone Faux Druzu Stone Stud Earrings (B0952DG1DG)**** – Shipped in a **recyclable gift box** with **rustic cardboard** and **no plastic lining**, emphasizing biodegradable practices. The sturdy cardboard backing provides **impact protection** during transit, matching your preference for **ethical, non-overpackaged solutions**.
3. **Trio Silver-tone Stud Earrings (B0952DZR5T)**** – Presented in a **white card with biodegradable materials**, these earrings are stacked securely in a 2x3 grid. The **cardboard display** and eco-conscious packaging reduce waste while ensuring **safe handling** and **minimal plastic use**.

These choices prioritize **recyclable or biodegradable materials** and **protective transit measures**, fully aligned with your sustainability ideals. Let me know if you want more packaging advice!

User:

Based on my usual tastes, which styles of enamel pins do you think would make a great gift for a craft beer lover?

Assistant:

Based on your focus on sustainability, unique craftsmanship, and nature-inspired designs for a craft beer lover, here are three standout enamel pin options:

1. **Hops Craft Beer Enamel Pin (B0762CCKNW)**** – This pin captures the essence of craft beer with a **detailed hop cone design**, featuring **textured green foliage** and a **gold-plated outline**. Eco-friendly materials and durable craftsmanship fit your values.
2. **Resin Beer Flora Pin (B08X5W6Q7R)**** – A minimalist resin pin inspired by **barley stalks** and **yeast cell patterns**, framed in **recycled brass** with **hypoallergenic stainless steel**. Combines beer culture

with nature themes.

3. **Minimalist Brewery Logo Pin (B09KJL8MNP)** – Crafted from **recycled metals**, with a sleek silhouette and **textured edges**, balancing simplicity, sustainability, and craft beer symbolism.

These choices blend **eco-friendly materials**, **textured design**, and **beer-themed motifs**, aligning seamlessly with your documented taste. Want deeper suggestions?

Table 12: Multimodal (No User Review) Example

Baseline Dialogue Example

User:

What unique styles of enamel pins would you recommend for someone who enjoys nature-inspired designs?

Assistant:

Hey there! If you're into nature-inspired enamel pins with a twist, I'd recommend checking out these unique styles: **Surreal Forestscape** for dreamlike, exaggerated woodland scenes; **Minimalist Botanical Geometry** for sleek, abstract floral patterns in monochrome; **Celestial Wildlife** to blend animals with cosmic flair like starry skies; **Decaying Beauty** for a poetic take on withered flora; and **Hybrid Nature-Abstract** to mix organic shapes with bold, modern textures. Each offers a fresh angle on nature, avoiding clichés while keeping the theme central. Let me know if you'd like more suggestions!

User:

When considering enamel pins, which materials do you think would offer the best combination of quality and aesthetic for me?

Assistant:

Hey there! When choosing enamel pins, the best materials often depend on your style preferences. Brass with high-quality enamel offers vibrant, glossy colors and durability, perfect for bold designs. Steel provides a sturdy base that retains intricate details beautifully, ensuring longevity. If you lean toward a sleek, modern look, nickel-plated brass paired with matte enamel adds sophistication while staying resilient. Let me know what vibe you're aiming for, and I can help narrow it down!

User:

Given my interest in handcrafted items, which types of enamel pins should I consider that highlight unique craftsmanship?

Assistant:

Hi there! If you're drawn to handcrafted enamel pins with unique craftsmanship, you might love exploring **custom folklore designs** for their intricate storytelling and hand-painted details, or **tribal-inspired patterns** that blend traditional techniques with symbolic artistry. For a touch of vintage elegance, **Art Deco motifs** offer geometric precision and metallic flair, while **mixed-media pins** (like those with wood or gemstone accents) push creative boundaries. Even **minimalist abstract designs** shine through layered enamel textures, proving that simplicity can still demand skill. Let me know if you'd like tips on where to find these gems!

User:

For someone who appreciates good design, how do you think the size of an enamel pin affects its overall impact and appeal?

Assistant:

Hey there! The size of an enamel pin plays a big role in how it's perceived—smaller pins (like 1-inch) let details shine and work well for minimalist designs, while larger ones (2-inch) make a bold statement and are perfect for eye-catching artwork. Medium sizes (1.5-inch) offer a balanced blend of visibility and subtlety, making them versatile for everyday wear. If you're into abstract or modern styles, thinner options (0.5-inch) can add a sleek, understated touch. Ultimately, the right size depends on your design goals and how you want it to interact with your outfit or collection! Let me know if you'd like help narrowing it down further.

User:

Which options would you suggest for an enamel pin that balances durability with a polished look?

Assistant:

Hey there! If you're looking for enamel pins that blend durability with a polished look, I'd recommend considering **brass enamel pins with a high-gloss finish** for their long-lasting shine, or **steel-backed pins** for a robust base paired with a smooth enamel coat. **Nickel-silver** is another great option—it's strong and maintains a bright, refined silver tone. For extra protection, **thick enamel coatings** ensure the design stays intact and glossy over time. If you want top-tier quality, **custom-made pins from reputable brands** often use premium materials and craftsmanship. Let me know if you'd like help narrowing it down further!

User:

What kind of packaging should I look for when purchasing enamel pins to ensure they arrive in great condition?

Assistant:

Hey there! When buying enamel pins, prioritize packaging that guards against scratches, dents, and moisture. Look for bubble wrap or foam inserts to cushion each pin, rigid cardboard boxes with reinforced corners for structural support, and vacuum-sealed bags to minimize movement. Individual silk or felt pouches also add a soft, breathable layer of protection. For extra security, custom molded trays can keep pins snug and stable during shipping. Let me know if you'd like help finding specific packaging options!

User:


Based on my usual tastes, which styles of enamel pins do you think would make a great gift for a craft beer lover?

Assistant:

Hey there! If you're looking for enamel pins that would delight a craft beer lover, consider these styles: Brewery-specific logos celebrate their favorite brands, hops and barley motifs highlight the artistry of brewing, and playful puns like "I'd Tap That" add a fun, community vibe. Vintage ale bottle art appeals to their love of history, while seasonal themes like pumpkin spice or imperial stouts cater to their passion for unique, limited-edition brews. These options blend humor, nostalgia, and the craft spirit they adore. Want help narrowing it down or finding specific designs?

Table 13: Baseline Example

Target Product:



Question: When looking for a fragrance, which scent profiles do you think would suit someone like me best?

Baseline:
 Hey there! To help you find the perfect fragrance, I'd love to know a bit more about your preferences. Are you leaning toward floral, fresh, woody, or another scent type? Do you prefer unisex, masculine, or feminine fragrances? Also, any notes you'd like to avoid, like vanilla or citrus? And what occasions or seasons are you thinking of? Let me know, and I'll narrow it down for you!

TextRag:
 Hi there! Based on your past preferences for ****natural, sweet, and floral scents**** (like vanilla, rose, and jasmine) with a subtle, non-overpowering vibe, here are some fragrances that align with your taste: \n\n1. ****MAYL'ODEUR Innertherafume Oil**** (B092HSMFBT) – A soft blend of ****rose, jasmine, and vanilla**** for a natural, intimate feel, perfect for those who love gentle sweetness.

MM No Review:
 Hi there! Since you're exploring fragrance profiles, I'd love to help narrow it down. **Could you share more about your preferences?** For example: \n- Do you lean toward fresh, floral, woody, or citrus notes? \n- Are there specific occasions (e.g., work, dates, outdoor activities) you want the scent for? \n- Do you prefer something bold, subtle, or versatile? ...

MM Full:
 Hi there! Based on your past preferences for **natural ingredients, subtle scents, and floral/woody notes**, I've curated a few options that align with your tastes. \n\n1. ****Derek Lam 10 Crosby - A Hold On Me**** (Floral with tiger lily and pimento berry) – This matches your appreciation for soft, fresh florals, like the **"gently sweet floral"** you loved in the Love Beauty And Planet review. ...

Figure 9: One turn question and response from different methods

A.7 Recommend and Reason Output Example

Recommender Agent Response
<p>User: <i>Can you help me find a good moisturizer for dry skin?</i></p> <p>Recommender Agent: ... To provide the best recommendation, a few details are needed:</p> <p>(1) Skin Concerns: Are you dealing with flakiness, redness, irritation, or general dryness?</p> <p>(2) Skin Type: Do you have sensitive skin, or are you okay with fragrance or other ingredients?</p> <p>(3) Budget: Are you looking for something affordable, mid-range, or premium?</p> <p>(4) Texture Preference: Do you prefer a lightweight gel, rich cream, or a lotion-like consistency?</p> <p>(5) Allergies or Ingredients to Avoid: (e.g., alcohol, fragrances, parabens).</p> <p>Top-rated moisturizers for dry skin (dermatologist recommended):</p> <p>(a) For Sensitive & Very Dry Skin:</p> <ul style="list-style-type: none"> – CeraVe Moisturizing Cream — ceramides and hyaluronic acid; fragrance-free, non-comedogenic. – La Roche-Posay Toleriane Double Repair Moisturizer — with ceramides and niacinamide, suitable for sensitive skin... <p>(b) For Lightweight, Fast-Absorbing Skin:</p> <ul style="list-style-type: none"> – Neutrogena Hydro Boost Water Gel — hyaluronic acid-based, lightweight, ideal for daytime use. – Vanicream Moisturizing Cream — hypoallergenic, fragrance-free, suitable for eczema-prone skin. <p>(c) For Intense Hydration (Nighttime / Very Dry Skin):</p> <ul style="list-style-type: none"> – First Aid Beauty Ultra Repair Cream — rich, healing, with shea butter and ceramides. – Kiehl’s Ultra Facial Cream — deeply hydrating and nourishing texture. <p>Pro Tip: Look for ingredients such as ceramides (repair barrier), hyaluronic acid (retain moisture), glycerin (humectant), shea butter or squalane (emollients), and oat extract or allantoin (soothing).</p> <p>...</p>

Table 14: Recommender Agent output example generated by BaseAgent. The agent elicits user preferences and provides structured, evidence-based product recommendations.

Reasoner Agent Evaluation on Recommend Agent Output

Reasoner Agent: ...

1. **Personalized Inquiry First:** The agent begins by asking questions about skin concerns, type, budget, texture preference, and allergies. This ensures the final recommendation is customized rather than generic, increasing the likelihood of satisfaction and effectiveness.
 2. **Categorization by Skin Needs:** The agent groups moisturizers by skin profiles (e.g., sensitive, very dry, lightweight, intense hydration), allowing users to quickly identify the most relevant category and reducing decision fatigue.
 3. **Inclusion of Dermatologist-Backed Brands:** Products such as CeraVe, La Roche-Posay, and Vanicream are widely recognized by dermatologists for safety, efficacy, and ingredient transparency, which builds user trust.
 4. **Clear Ingredient Guidance:** Highlighting key ingredients (ceramides, hyaluronic acid, glycerin, etc.) helps users understand *why* certain products work and supports informed decisions.
 5. **Balance of Price and Performance:** Offers both affordable and premium options (e.g., Neutrogena vs. Kiehl’s), ensuring accessibility without compromising quality.
 6. **Texture and Use-Case Consideration:** Distinguishes lightweight gels (for daytime) from rich creams (for nighttime or severe dryness), reflecting practical understanding of skincare routines.
 7. **Pro Tip for Education:** Includes a brief ingredient guide that helps users independently evaluate future products, providing long-term educational value.
 8. **Friendly Tone and Encouragement:** The approachable tone (e.g., “Let me know your preferences”) promotes engagement and user confidence.
- Summary:** The Reasoner concludes that the Recommender Agent’s message is comprehensive, user-centered, and evidence-based. It goes beyond listing products to actively guide decision-making aligned with user preferences and lifestyle, thus enhancing trust and satisfaction.

Table 15: Reasoner Agent evaluation of the Recommender Agent’s output, highlighting personalization, structure, credibility, and educational value.