DEEPER Insight into Your User: Directed Persona Refinement for Dynamic Persona Modeling

Anonymous ACL submission

Abstract

To advance personalized applications such as recommendation systems and user behavior prediction, recent research increasingly adopts large language models (LLMs) for human-readable persona modeling. In dynamic real-world scenarios, effective persona modeling necessitates leveraging streaming behavior data to continually optimize user personas. However, existing methods—whether regenerating personas or incrementally extending them with new behaviors—often fail to achieve sustained improvements in persona quality or future behavior prediction accuracy. To address this, we propose DEEPER, a novel approach for dynamic persona modeling that enables continual persona optimization. Specifically, we enhance the model's direction-search capability through an iterative reinforcement learning framework, allowing it to automatically identify effective update directions and optimize personas using discrepancies between user behaviors and model predictions. Extensive experiments on dynamic persona modeling involving 4,800 users across 10 domains highlight DEEPER 's superior persona optimization capabilities, delivering an impressive 32.2% average reduction in user behavior prediction error over four update rounds—outperforming the best baseline by a remarkable 22.92%.

1 Introduction

Recent studies increasingly utilize Large Language Models (LLMs) (OpenAI, 2023) for human-readable and interpretable persona modeling, advancing personalized applications like recommendation and behavior prediction. However, most research focuses on generating personas from static historical data, which fail to capture dynamic behaviors and evolving preferences in real-world interactive scenarios (Wang and Lim, 2023; Zhou et al., 2024). This underscores the need for dynamic persona modeling—a pivotal yet underexplored approach that iteratively updates personas using streaming user behavior data to continually enhance their quality.

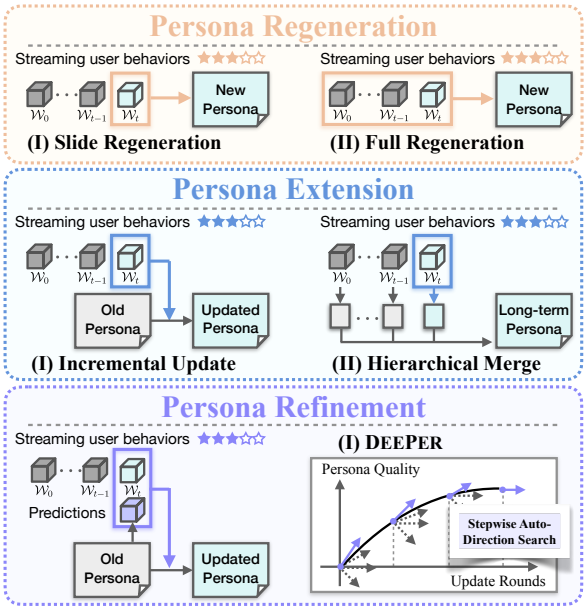

Figure 1: Comparison of dynamic persona modeling paradigms: Regeneration replaces personas, and Extension adds to them, but neither ensures optimization. Our DEEPER, based on Refinement paradigm, uses discrepancies between user behavior and model predictions to identify update directions for continuous optimization.

Existing dynamic persona modeling methods can be broadly categorized into two paradigms: (1) *Persona Regeneration*, which updates through complete replacement, rebuilds personas from scratch based on new user behaviors, either by aggregating historical and recent behaviors or by using sliding-window methods (Wu et al., 2024a; Yang et al., 2023). (2) *Persona Extension*, which updates through additive extension, incorporates new user behaviors into existing personas, either by directly integrating them or by merging short-term and long-term personas. (Liu et al., 2024; Tang et al., 2023). However, while these methods enable dynamic updates, they fail to ensure meaningful optimization due to the lack of mechanisms to evaluate update ef-

fectiveness and explicitly model the update process. Without validating whether updates enhance persona quality or predictive accuracy, both paradigms risk propagating errors and degrading performance. This highlights a critical challenge in dynamic persona modeling: *Bridging the gap between updating personas and truly optimizing them*.

To better characterize the update process and bridge this gap, we introduce the concept of *update direction*, which uniquely identifies the transformation from an existing persona to an updated one under given signals. It directly determines whether the update improves, degrades, or maintains persona quality within a specific context, serving as a core factor in persona optimization.

However, identifying an effective update direction is challenging due to the fundamental misalignment between the dense natural language persona space and the discrete user behavior space (e.g., ratings): (1) Behavior signals are insufficient, e.g., a user's 1-star movie rating does not clearly indicate whether the dissatisfaction is due to the story, pacing, or genre, making it difficult to identify specific errors in the persona. (2) Evaluating update directions is inherently complex, e.g., even if we adjust the persona to emphasize "plot complexity" or "character development," it's unclear which change would lead to better predictions.

To address the challenges, we propose **DEEPER** (**D**irec**t**ed **Pe**rsona **R**efinement), a novel approach for LLM-based dynamic persona modeling. Specifically, we introduce a new paradigm, *Persona Refinement* (Figure 1), which uses discrepancies between user behaviors and model predictions as stronger update signals to expose deficiencies in personas. To identify effective update directions, we decompose the optimization objective into three direction search goals: *Previous Preservation*, *Current Reflection*, and *Future Advancement*, ensuring stability, adaptability, and task alignment. Based on these goals, we design reward functions for clear and measurable assessments of update directions by comparing predictive errors before and after updates. Finally, we propose an iterative reinforcement learning (RL) framework with two training stages, leveraging self-sampling and DPO fine-tuning to progressively enhance the model's direction search and persona refinement capabilities, ultimately improving prediction accuracy.

Extensive experiments on over 4800 users across 10 domains demonstrate DEEPER 's strong persona optimization and direction search capability.

In summary, our contributions are as follows:

- We identify key limitations in current LLM-based dynamic persona modeling methods, emphasizing the critical gap between persona updating and optimization caused by weak update signals and unclear update direction.
- We propose DEEPER, a novel approach to dynamic persona modeling that achieves continual optimization through discrepancy-based update signals and robust direction search.
- Extensive experiments demonstrate that DEEPER successfully bridges this gap, outperforming existing methods in dynamic persona modeling.

2 Dynamic Persona Modeling

Building on prior work(Yang et al., 2023; Kang et al., 2023; Zhou et al., 2024), we formalize the concept of persona quality and the objective of dynamic persona modeling as follows:

Definition 1 (*Persona Quality*) The extent to which a persona accurately represents a user's preferences and behaviors, indicating its ability to predict future behaviors within a specific domain.

Definition 2. (*Persona Optimization*) The updated persona better represents a user than the previous persona, with improved predictive capability within a specific domain.

Objective: (*Continual Persona Optimization*) Iteratively enhance persona quality through multi-round updates, progressively enhancing its predictive capability within a specific domain.

2.1 Task Formulation

Consider a user $\mathcal{U}$ in domain $\mathcal{X}$. To capture temporal dynamics of user behaviors, we segment user's online interactions into sequential, time-ordered windows $\mathbf{W} = \{\mathcal{W}_t\}_{t=0}^{\mathcal{T}}$. Each window $\mathcal{W}_t$ contains N interactions, represented by an item list $\mathbf{I}_t = \{j_t^j\}_{j=1}^N$ and the corresponding user behaviors $\mathbf{O}_t = \{o_t^j\}_{j=1}^N$. As new data arrives at time t, the current window $\mathcal{W}_t$ captures interactions from the present period, while $\mathcal{W}_{t-1}$ reflects previous behaviors, and $\mathcal{W}_{t+1}$ outlines future interactions.

The LLM-based dynamic persona modeling pipeline consists of three stages:

- **Persona Initialization:** At time step $t = 0$, the persona $\mathcal{S}_0$ is initialized based on the user behaviors in the initial window $\mathcal{W}_0$.

- **Behavior Observation and Prediction:** In each window $\mathcal{W}_t$, previous persona $\mathcal{S}_{t-1}$ is used to predict user behaviors $\hat{\mathbf{O}}_{t|\mathcal{S}_{t-1}} = \mathcal{P}(\mathcal{S}_{t-1})$, while actual behaviors $\mathbf{O}_t$ are observed.
- **Persona Update:** At the end of each window $\mathcal{W}_t$, the persona updates using new observations.

For the first two stages, we use frozen LLM to generate initial personas and predictions across all modeling paradigms. The Persona Update stage, however, varies by paradigm and is formulated as:

- **Persona Regeneration:** Rebuild persona at the end of each window $\mathcal{W}_t$ using new behaviors $\mathbf{O}_t$:

$$\mathcal{S}_t = f_{\text{regen}}(\mathbf{O}_t). \tag{1}$$

- **Persona Extension:** Extend the previous persona $\mathcal{S}_{t-1}$ with new behaviors $\mathbf{O}_t$:

$$\mathcal{S}_t = f_{\text{exten}}(\mathcal{S}_{t-1}, \mathbf{O}_t). \tag{2}$$

- **Persona Refinement (proposed):** Refine the previous persona $\mathcal{S}_{t-1}$ with new user behaviors $\mathbf{O}_t$, and predicted results $\hat{\mathbf{O}}_{t|\mathcal{S}_{t-1}}$:

$$\mathcal{S}_t = f_{\text{refine}}(\mathcal{S}_{t-1}, \mathbf{O}_t, \hat{\mathbf{O}}_{t|\mathcal{S}_{t-1}}). \tag{3}$$

2.2 Task Evaluation

In this work, we assess persona quality indirectly through performance in a user- and domain-specific task: future behavior prediction. Prediction error, quantified by the Mean Absolute Error (MAE), serves as an indicator of *Persona Quality*:

$$\varepsilon_{t+1|\mathcal{S}_t} = \frac{1}{n}\sum_{j=1}^{N} \left|\hat{o}^j_{t+1|\mathcal{S}_t} - o^j_{t+1}\right|. \tag{4}$$

$\mathbf{O}_{t+1} = \{o^j_{t+1}\}_{j=1}^N$ represents user actual behaviors in $\mathcal{W}_{t+1}$, while $\hat{\mathbf{O}}_{t+1|\mathcal{S}_t} = \{\hat{o}^j_{t+1|\mathcal{S}_t}\}_{j=1}^N$ denotes predictions with persona $\mathcal{S}_t$.

Lower error indicates better alignment. *Persona Optimization* is realized when an updated persona reduces the prediction error for future behaviors:

$$\varepsilon_{t+1|\mathcal{S}_t} < \varepsilon_{t|\mathcal{S}_{t-1}}. \tag{5}$$

Thus, the evaluation of a dynamic persona modeling method is determined by its ability to achieve the objective of *Continual Persona Optimization*, with an effective update strategy evidenced by a progressive reduction in prediction error over time.

3 DEEPER

Existing regeneration- and extension-based methods enable dynamic updates but fall short in consistent quality improvement, resulting in a misalignment between the *update step* and the *optimization objective*. To address this, we highlight the critical role of update direction in ensuring effective updates and propose the following core proposition: *Better update directions lead to better personas.* Instead of directly searching for improved personas, we optimize refinement directions. By incorporating model predictions into the context and defining three high-level goals for direction search, we propose an iterative reinforcement learning framework with a balanced reward function, enabling effective refinement and continual persona optimization.

3.1 Refinement Step Formulation

In DEEPER, each persona refinement step at time t, can be formulated as a reinforcement learning (RL) task. The objective is to learn a policy π_θ to identify optimal refinement directions for specific contexts. For a single user $\mathcal{U}$ in domain $\mathcal{X}$, the refinement step can be formulated as:

- **State:** The previous persona $\mathcal{S}_{t-1}$, generated after the $(t-1)$-th refinement round at the end of the previous window $\mathcal{W}_{t-1}$.
- **Observation:** The observation at time step t, $\mathcal{O}_t = \{\mathbf{O}_t, \hat{\mathbf{O}}_{t|\mathcal{S}_{t-1}}\}$, where $\mathbf{O}_t$ and $\hat{\mathbf{O}}_{t|\mathcal{S}_{t-1}}$ represent the actual and predicted behaviors in the current window $\mathcal{W}_t$.
- **Action:** The refined persona $\mathcal{S}_t$, generated after the t-th refinement process based on the corresponding $(\mathcal{S}_{t-1}, \mathcal{O}_t)$ of the user.
- **Policy Model:** The refinement model π_θ, maps the state and observation to refined persona $\mathcal{S}_t$:

$$\pi_\theta : (\mathcal{S}_{t-1}, \mathcal{O}_t) \to \mathcal{S}_t. \tag{6}$$

- **Reward:** The reward r_t quantifies the effectiveness of the refinement process.

3.2 Direction and Goal Definition

In this work, we formally define the persona refinement direction and its goals as follows:

Definition 3. (*Persona Refinement Direction*) Identify the directed path of a specific persona refinement step, denoted as $\mathcal{D}_t$, which is uniquely determined by the previous persona $\mathcal{S}_{t-1}$, the current observation $\mathcal{O}_t$, and the refined persona $\mathcal{S}_t$:

$$\mathcal{D}_t \leftrightarrow (\mathcal{S}_{t-1}, \mathcal{O}_t; \mathcal{S}_t). \tag{7}$$

We define three high-level goals for direction search, ensuring comprehensive guidance with temporal insights from past, present, and future.

Goal 1. (*Previous Preservation*): Retain stable persona traits from historical behaviors to ensure consistency and preserve critical information.

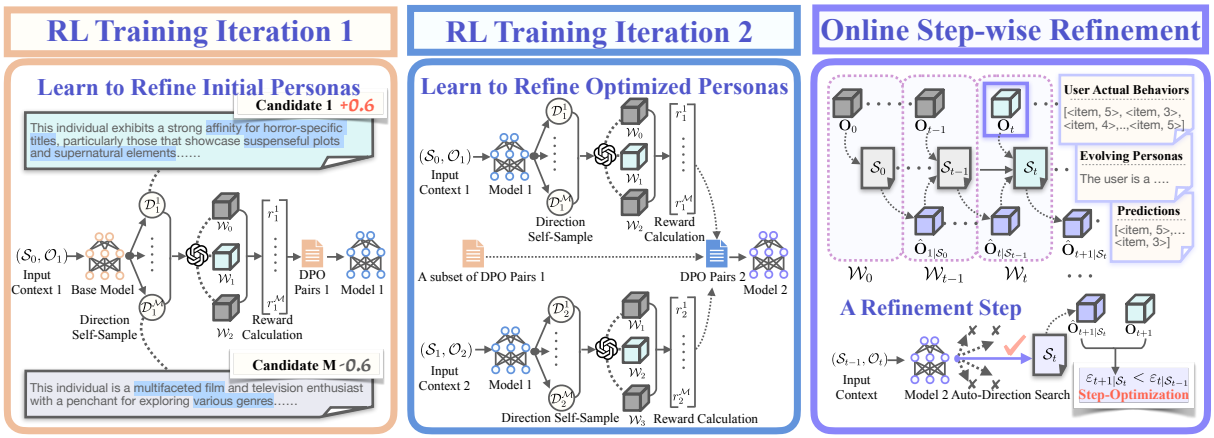

Figure 2: Framework of DEEPER. Grounded in three high-level goals for direction search, the iterative RL framework progressively enhances the model's refinement capability through two rounds of self-sampling and training. Applied online in multi-round updates, it enables step-wise persona optimization via directed refinement.

Goal 2. (*Current Reflection*): Adapt to recent user behaviors by incorporating dynamic changes and correcting errors in the previous persona.

Goal 3. (*Future Advancement*): Enhance the persona's predictive capability for future behaviors.

3.3 Reward Function Design

Given the unique correspondence between $\mathcal{D}_t$ and the triplet $(\mathcal{S}_{t-1}, \mathcal{O}_t; \mathcal{S}_t)$, the quality of $\mathcal{D}_t$ directly determines the refined persona's quality and process effectiveness within context $(\mathcal{S}_{t-1}, \mathcal{O}_t)$. Accordingly, we formalize three goals of *Direction Quality* as reductions in prediction error from refinement across past, current, and future windows, represented by rewards r_t^{prev}, r_t^{curr}, and r_t^{fut}.

$$\begin{aligned} r_t^{prev} &= \varepsilon_{t-1|\mathcal{S}_{t-1}} - \varepsilon_{t-1|\mathcal{S}_t} \\ r_t^{curr} &= \varepsilon_{t|\mathcal{S}_{t-1}} - \varepsilon_{t|\mathcal{S}_t} \\ r_t^{fut} &= \varepsilon_{t+1|\mathcal{S}_{t-1}} - \varepsilon_{t+1|\mathcal{S}_t}. \end{aligned} \tag{8}$$

$\varepsilon_{t-1|\mathcal{S}_{t-1}}$, $\varepsilon_{t|\mathcal{S}_{t-1}}$, and $\varepsilon_{t+1|\mathcal{S}_{t-1}}$ are prediction errors with previous persona $\mathcal{S}_{t-1}$ across $\mathcal{W}_{t-1}$, $\mathcal{W}t$, and $\mathcal{W}_{t+1}$, respectively, while $\varepsilon_{t-1|\mathcal{S}_t}$, $\varepsilon_{t|\mathcal{S}_t}$, and $\varepsilon_{t+1|\mathcal{S}_t}$ are errors with refined persona $\mathcal{S}_t$.

The total reward for a refinement step is:

$$r_t = r_t^{prev} + r_t^{curr} + r_t^{fut}. \tag{9}$$

3.4 Iterative Training Framework

DEEPER employs an iterative training framework (Figure 2): Iteration 1 fine-tunes the base model to refine initial personas (Model 1), while Iteration 2 further enhances it to refine pre-optimized personas (Model 2). Direct Preference Optimization (DPO) is used to seamlessly integrate rewards into preference pairs, enabling the model to identify better directions through explicit comparisons and supporting scalable iterative fine-tuning.

Iteration 1: Learn to Refine Initial Personas Iteration 1 formulates the first refinement step at $t = 1$ as an RL task, where we fine-tune the base model to refine initial personas $\mathcal{S}_0$, establishing a baseline policy for direction search and refinement.

Context Data Construction First, we initialize personas $\mathcal{S}_0$ for users across multiple domains with their behaviors in window $\mathcal{W}_0$, serving as initial states for refinement processes. The prediction model then predicts user behaviors in $\mathcal{W}_1$ based on $\mathcal{S}_0$. Combining predicted and actual behaviors, we construct observations $\mathcal{O}_1$. Together, $(\mathcal{S}_0, \mathcal{O}_1)$ form the context of the first refinement step.

Direction Sampling and Reward Calculation For each context input $(\mathcal{S}_0, \mathcal{O}_1)$, the base model samples M candidate refined personas $\{\mathcal{S}_1^k\}_{k=1}^M$, where each candidate direction $\mathcal{D}_t^k$ is represented by $(\mathcal{S}_0, \mathcal{O}_1; \mathcal{S}_1^k)$. Rewards for these directions as calculated as specified in Equation (9).

Preference Pairs Construction and Training Refined personas are partitioned into a positive set $\mathcal{S}_1^+$ (rewards $r_t \geq \tau^+$) and a negative set $\mathcal{S}_1^-$ (rewards $r_t \leq \tau^-$) based on reward thresholds. To ensure a clear distinction, we enforce a margin δ, derived from the reward distribution, such that $r_t^w - r_t^l \geq \delta$. The base model is then fine-tuned using DPO with these preference pairs. Following (Gui et al., 2024), a Supervised Fine-Tuning (SFT) loss is incorporated into the standard DPO objective to maintain alignment with high-quality refinements:

$$\mathcal{L}(\pi_\theta; \pi_{\text{ref}}) = \mathcal{L}_{\text{DPO}}(\pi_\theta; \pi_{\text{ref}}) + \alpha \mathcal{L}_{\text{SFT}}(\pi_\theta). \tag{10}$$

Iteration 2: Learn to Refine Optimized Personas Iteration 2 extends the model's refinement capabilities to handle pre-optimized personas, addressing increased complexity of nuanced refinement tasks.

Context Data Construction Includes: (1) Contexts from Iteration 1 $(\mathcal{S}_0, \mathcal{O}_1)$. (2) New contexts $(\mathcal{S}_1, \mathcal{O}_2)$ constructed based on the second refinement step, using $\mathcal{S}_1$ as initial states, with predicted and actual behaviors in $\mathcal{W}_2$ as observations.

Direction Sampling and Reward Calculation Model 1 is used to sample candidates, following the same procedure of Iteration 1.

Preference Pairs Construction and Training Similarly to Iteration 1, we construct preference pairs with consistent boundaries for positive and negative sets, with a larger margin δ to accommodate refined reward distribution and model performance. Model 1 is then fine-tuned with the same combined loss as in Iteration 1, incorporating a subset of preference pairs from Iteration 1 to prevent forgetting and ensure continual learning.

4 Experiment

4.1 Experiment Setup

Dataset and Task Data Construction We evaluate DEEPER on four real-world datasets across 10 domains, including *MovieLens 20M*(Harper and Konstan, 2015), *Food.com Recipes*(Majumder et al., 2019), *Google Local Reviews*(Yan et al., 2023; Li et al., 2022), and *Amazon Reviews (2018)*(Ni et al., 2019). From six domains, we sample 14,959 users with over 50 ratings (10,800 for training and 4,159 for testing). To assess generalization, an auxiliary test set of 650 users from four unseen domains is constructed. User interactions are segmented into five 10-rating windows, with $\mathcal{W}_0$ used for initial persona generation.

Evaluation As described in Section 2, we evaluate the effectiveness of persona update methods based on their ability to achieve *Continual Persona Optimization*, quantified by the reduction in future prediction error $\varepsilon_{t+1|\mathcal{S}_t}$ across update rounds.

Baselines We compare against baselines from two paradigms: 1. **Persona Regeneration**: - *SlideRegen*: Rebuilds personas using only the latest window of behaviors (Yang et al., 2023). - *FullRegen*: Reconstructs personas by leveraging all historical and recent behaviors (Zhou et al., 2024). 2. **Persona Extension**: - *IncUpdate*: Incrementally integrates new behaviors into existing personas (Yuan et al., 2024). - *HierMerge*(Liu et al., 2024): Hierarchically merges short-term and long-term personas.

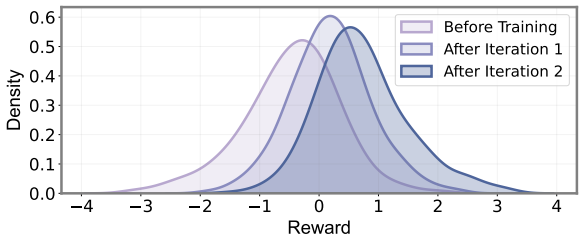

Figure 3: KDE plot illustrating changes in reward distribution across test sets before and after training.

DEEPER Training Details For DEEPER, we use Llama-3.1-8B-Instruct as the base policy model, trained iteratively on data from 10,809 users. Each iteration samples 15 candidate personas per input, with reward boundaries $\tau^+ = 0.5$ and $\tau^- = 0$. Iteration 1 applies a margin $\delta = 0.5$, producing 34,782 DPO pairs. Iteration 2 increases δ to 0.8 and incorporates 5,000 pairs from Iteration 1, resulting in 33,612 pairs. Both iterations use LoRA for fine-tuning with a learning rate of 5×10^{-6}, 4 training epochs, and a batch size of 128. The SFT loss coefficient (α) is set to 0.1. Figure 3 illustrates the reward distribution improvements across iterations.

Global and Baseline Settings We use the frozen, powerful LLM, GPT-4o-mini, to generate initial personas and predictions in a zero-shot setting for both training and evaluation, ensuring consistent initial persona quality and unbiased predictions. Additionally, it serves as the backbone for all baselines, offering a robust foundation for comparison.

4.2 Main Results

Figure 4 compares performance of DEEPER and baseline methods over four update rounds across 10 domains in the dynamic persona modeling task.

DEEPER helps continual persona optimization. DEEPER consistently achieves substantial MAE reductions across all 10 domains over four update rounds, with an average decrease of 32.2%, significantly outperforming extension-based baselines such as *IncUpdate* (9.28%) and *HierMerge* (3.92%). Notably, in the unseen domain *Arts Crafts and Sewing*, DEEPER achieves the largest improvement, reducing MAE from 0.76 to 0.40 (47.1%). In contrast, regeneration-based baselines like *FullRegen* and *SlideRegen* often exhibit minimal or negative gains, highlighting their inability to meet the task objective.

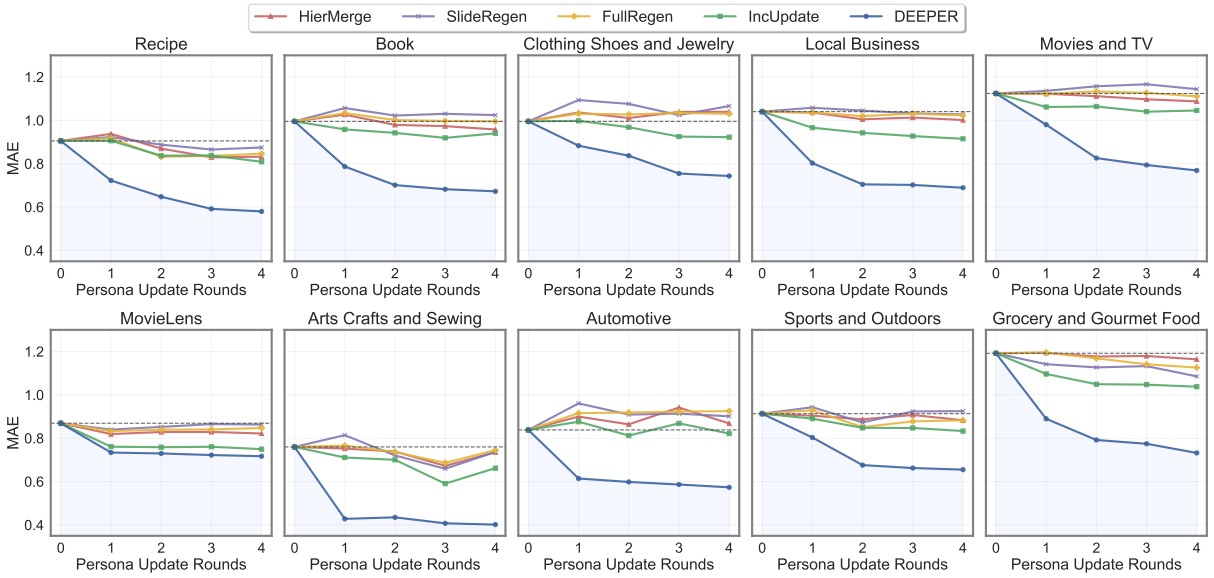

Figure 4: Performance of different methods in dynamic persona modeling over 4 rounds across 10 domains. The first six ((A) *Recipe*, (B) *Book*, (C) *Clothing Shoes and Jewelry*, (D) *Local Business*, (E) *Movies and TV*, (F) *MovieLens*) are seen during training, while ((G) *Arts Crafts and Sewing*, (H) *Automative*, (I) *Sports and Outdoors*, (J) *Grocery and Gourmet Food*) are unseen. In subsequent figures, domains are referred to by their corresponding letters.

Generalized capability and domain-specific dynamic. DEEPER achieves an average MAE reduction of 29.4% in seen domains and 36.4% in unseen domains. This emphasizes its generalized optimization capability to diverse and new scenarios. Figure 4 also reveals domain-specific variations in optimization speed, convergence patterns, and improvement potential. For instance, domains like *Automotive* exhibit faster optimization with earlier convergence, while *Movies and TV* shows slower progress and prolonged refinement. These suggest potential influences from varying persona modeling complexities, behavior predictability, and interest stability across domains.

5 In-depth Analysis

5.1 What enables DEEPER's effectiveness

Direction search enables optimization. We first evaluate the necessity of direction search by comparing DEEPER with frozen models (GPT-4o-mini and Llama-3.1-8B-Instruct (Naive Model)) which refine personas directly. As shown in Figure 5(a), both baselines exhibit significant error increases after refinement, underscoring the critical role of direction search in effective optimization.

Balanced goals drive better optimization. The assessment of Direction Quality is critical to optimization performance. We compare DEEPER 's balanced reward setting(equally weights previous, current, and future goals), with a future-focused reward and a decayed reward (decay factor = 0.5) prioritizing recent goals. As shown in Figure 5(b), DEEPER consistently outperforms both baselines across all domains. These findings underscore the importance of balanced, goal-driven direction search in enabling effective persona refinement.

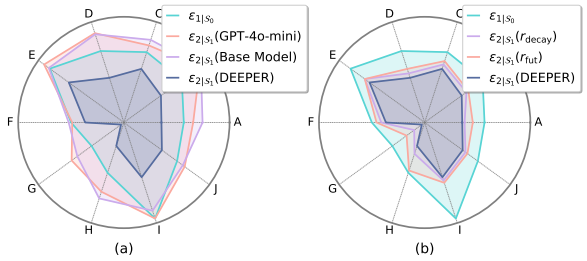

Figure 5: (a) Refinement performance of DEEPER compared to frozen models across ten domains, using $\varepsilon_{1|\mathcal{S}_0}$ as pre-refinement baseline. (b) Refinement under different reward settings. Smaller areas indicate reduced errors and improved refinement relative to baseline.

DEEPER excels in identifying high-quality directions. Building on previous insights, we further evaluate DEEPER 's ability to identify high-quality refinement directions by analyzing its performance across three goals (Table 1). DEEPER demonstrates outstanding performance: minimizing previous forgetting with the smallest average MAE increment of 0.062 (*Previous Preservation*); reducing current errors by 0.572 on average (*Current Reflection*); and improving future predictions with an average

Domain	Pre-Update	Post-Update				
	$\mathcal{S}_{old}$	SlideRegen	FullRegen	IncUpdate	HierMerge	DEEPER
*Previous Window Prediction ($\varepsilon_{prev\|\mathcal{S}_{old/new}}$) - **Previous Preservarion***						
Recipe	0.57	0.95 (0.38↑)	0.83 (0.26↑)	0.83 (0.26↑)	0.71 (0.14↑)	0.70 (0.13↑)
Book	0.78	1.09 (0.31↑)	0.94 (0.16↑)	0.91 (0.13↑)	0.88 (0.10↑)	0.76 (0.02↓)
Clothing Shoes and jewelry	0.63	1.13 (0.50↑)	0.96 (0.33↑)	0.94 (0.31↑)	0.77 (0.14↑)	0.82 (0.19↑)
Local Business	0.63	1.10 (0.47↑)	0.95 (0.32↑)	0.91 (0.28↑)	0.74 (0.11↑)	0.73 (0.10↑)
Movies and TV	0.92	1.17 (0.25↑)	1.03 (0.11↑)	1.00 (0.08↑)	0.98 (0.06↑)	0.85 (0.07↓)
MovieLens	0.76	0.89 (0.13↑)	0.83 (0.07↑)	0.80 (0.04↑)	0.80 (0.04↑)	0.74 (0.02↓)
Arts Crafts and Sewing	0.49	0.81 (0.32↑)	0.74 (0.25↑)	0.68 (0.19↑)	0.59 (0.10↑)	0.46 (0.03↓)
Automotive	0.55	1.00 (0.45↑)	0.93 (0.38↑)	0.82 (0.27↑)	0.66 (0.11↑)	0.63 (0.08↑)
Sports and Outdoors	0.56	0.99 (0.43↑)	0.87 (0.31↑)	0.85 (0.29↑)	0.67 (0.11↑)	0.66 (0.10↑)
Grocery and Gourmet Food	0.63	1.13 (0.50↑)	1.00 (0.37↑)	0.95 (0.32↑)	0.71 (0.08↑)	0.79 (0.16↑)
Average	**0.652**	**1.026** (0.374↑)	**0.908** (0.256↑)	**0.869** (0.217↑)	**0.751** (0.099↑)	**0.714** (0.062↑)
*Current Window Prediction ($\varepsilon_{curr\|\mathcal{S}_{old/new}}$) - **Current Reflection***						
Recipe	0.91	0.78 (0.13↓)	0.84 (0.07↓)	0.41 (0.50↓)	0.80 (0.11↓)	0.44 (0.47↓)
Book	1.00	0.92 (0.08↓)	0.97 (0.03↓)	0.41 (0.59↓)	0.91 (0.09↓)	0.35 (0.65↓)
Clothing Shoes and Jewelry	1.00	0.90 (0.10↓)	0.96 (0.04↓)	0.48 (0.52↓)	0.90 (0.10↓)	0.51 (0.49↓)
Local Business	1.04	0.9 (0.14↓)	0.99 (0.05↓)	0.29 (0.75↓)	0.93 (0.11↓)	0.36 (0.68↓)
Movies and TV	1.12	1.00 (0.12↓)	1.07 (0.05↓)	0.47 (0.65↓)	1.02 (0.10↓)	0.45 (0.67↓)
MovieLens	0.87	0.78 (0.09↓)	0.82 (0.05↓)	0.30 (0.57↓)	0.80 (0.07↓)	0.43 (0.44↓)
Arts Crafts and Sewing	0.76	0.76 (0.00↑)	0.77 (0.01↑)	0.39 (0.37↓)	0.72 (0.04↓)	0.26 (0.50↓)
Automotive	0.84	0.81 (0.03↓)	0.88 (0.04↑)	0.38 (0.46↓)	0.81 (0.03↓)	0.27 (0.57↓)
Sports and Outdoors	0.91	0.79 (0.12↓)	0.84 (0.07↓)	0.37 (0.54↓)	0.82 (0.09↓)	0.36 (0.55↓)
Grocery and Gourmet Food	1.19	0.93 (0.26↓)	1.08 (0.11↓)	0.46 (0.73↓)	1.05 (0.14↓)	0.49 (0.70↓)
Average	**0.964**	**0.857** (0.107↓)	**0.922** (0.042↓)	**0.396** (0.568↓)	**0.876** (0.088↓)	**0.392** (0.572↓)
*Future Window Prediction ($\varepsilon_{fut\|\mathcal{S}_{old/new}}$) - **Future Advancement***						
Recipe	0.91	0.92 (0.01↑)	0.92 (0.01↑)	0.91 (0.00↑)	0.94 (0.03↑)	0.72 (0.19↓)
Book	1.01	1.06 (0.05↑)	1.03 (0.02↑)	0.96 (0.05↓)	1.03 (0.02↑)	0.79 (0.22↓)
Clothing Shoes and Jewelry	1.03	1.09 (0.06↑)	1.03 (0.00↑)	1.00 (0.03↓)	1.04 (0.01↑)	0.88 (0.15↓)
Local Business	1.04	1.06 (0.02↑)	1.04 (0.00↑)	0.97 (0.07↓)	1.04 (0.00↑)	0.80 (0.24↓)
Movies and TV	1.18	1.14 (0.04↓)	1.12 (0.06↓)	1.06 (0.12↓)	1.12 (0.06↓)	0.98 (0.20↓)
MovieLens	0.85	0.84 (0.01↓)	0.83 (0.02↓)	0.76 (0.09↓)	0.82 (0.03↓)	0.73 (0.12↓)
Arts Crafts and Sewing	0.75	0.81 (0.06↑)	0.77 (0.02↑)	0.71 (0.04↓)	0.75 (0.00↑)	0.43 (0.32↓)
Automotive	0.86	0.96 (0.10↑)	0.92 (0.06↑)	0.88 (0.02↑)	0.90 (0.04↑)	0.61 (0.25↓)
Sports and Outdoors	0.97	0.94 (0.03↓)	0.93 (0.04↓)	0.89 (0.08↓)	0.90 (0.07↓)	0.80 (0.17↓)
Grocery and Gourmet Food	1.25	1.14 (0.11↓)	1.20 (0.05↓)	1.10 (0.15↓)	1.19 (0.06↓)	0.89 (0.36↓)
Average	**0.985**	**0.996** (0.011↑)	**0.979** (0.006↓)	**0.924** (0.061↓)	**0.973** (0.012↓)	**0.763** (0.222↓)

Table 1: MAE results of previous, current, and future window prediction tasks using personas Pre- and Post- the first update with different methods. This table illustrates how well each method achieves the three high-level goals: Previous Preservation, Current Reflection, and Future Advancement. It presents the changes in MAE ($\left|\varepsilon_{t|\mathcal{S}_{old}} - \varepsilon_{t|\mathcal{S}_{new}}\right|$) relative to the old persona, with upward arrows (↑) indicating error increases and downward arrows (↓) indicating error reductions. Average results are highlighted in **bold**, and the best results are underlined

reduction of 0.222 (*Future Advancement*). Notably, DEEPER surpasses all baselines across domains for *Future Advancement*, demonstrating its capacity step-wise optimization. These results highlight DEEPER 's ability to balance three goals for better direction search and continual optimization.

Iterative RL enhances persona refinement. Guided by stage-specific objectives, DEEPER 's two-stage iterative RL framework incrementally enhances refinement capabilities by leveraging progressively higher-quality self-sampled data and expanded preference margins. Results (Figure 6(a)) show accelerated improvements in the second iteration, highlighting effects of iterative training.

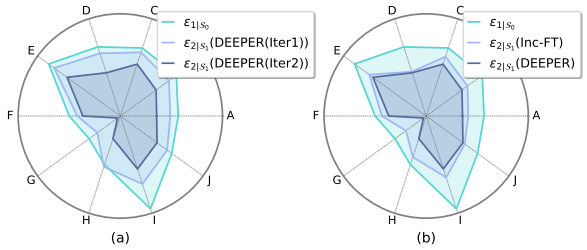

Figure 6: (a) Refinement performance across two RL iterations of DEEPER. (b) Comparison of DEEPER and fine-tuned *IncUpdate* (Inc-FT).

Prediction discrepancy facilitates direction search. We finally analyze paradigm's role in direction search by employing DEEPER 's training framework into *IncUpdate* (the best-performing

baseline). Figure 6(b) show that while direction search training improves *IncUpdate*'s performance, it still falls short of DEEPER. This underscores prediction discrepancy's role in enabling context-specific search and more precise refinement.

5.2 Persona Probing

We further conduct an preliminary analysis of refined personas, termed *persona probing*, to explore additional insights and applications of DEEPER.

Update Method	$\mathcal{S}_0$	$\mathcal{S}_1$	$\mathcal{S}_2$	$\mathcal{S}_3$	$\mathcal{S}_4$
DEEPER	245.0	316.8	353.5	393.2	429.4
IncUpdate	245.0	390.1	459.3	500.4	526.4
HierMerge	245.0	325.3	393.5	462.2	509.1

Table 2: Average persona token count across rounds.

Dynamic persona evolution across rounds. We first analyze persona dynamics during refinement process. Table 2 highlights DEEPER 's controlled length growth, balancing representation efficiency and informativeness. Figure 7(a) reveals diminishing persona changes over time, with substantial shifts in early updates ($\mathcal{S}_0 \rightarrow \mathcal{S}_1$) and increasing stability in later rounds ($\mathcal{S}_1 \rightarrow \mathcal{S}_4$), indicating convergence and improved contextual alignment.

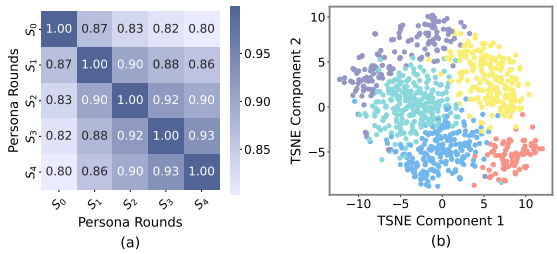

Figure 7: (a) Cosine similarity among personas across rounds; (b) User clusters based on final personas(Book).

Profiling Dimensions(Book)	User Count
Story & Plot	871
Emotion & Experience	878
Genre & Theme	878
Social & Cultural Context	680
User Behavior Traits	862
Author & Character	701
Personality & Values	867
Relationship & Connection	716

Table 3: Key profiling dimensions in Book domain.

Insights from final optimized personas. Refined personas from DEEPER also enable in-depth, domain-specific exploration. In Book domain, we uncover **group-level preferences** by clustering final persona embeddings, identifying five user groups characterized by unique high-frequency adjectives (e.g., "romantic" and "practical") (Figure 7(b)). We also extract **domain-specific patterns** by organizing high-frequency terms into eight dimensions using GPT-4o (Table 3), highlighting critical factors for modeling Book domain users. These attempts show DEEPER 's potential to support strategic user insights exploration.

6 Related Work

Persona Modeling Persona modeling in personalized applications captures user preferences and behaviors from behavioral data or dialogue history, with advancements driven by LLMs (Li and Zhao, 2021; Tan and Jiang, 2023; Tseng et al., 2024). Most studies focus on one-time persona generation from static user behavior or profile data (Ji et al., 2023; Wang and Lim, 2023; Zhou et al., 2024; Wu et al., 2024a; Wang et al., 2024; Xu et al., 2024; Lyu et al., 2023; Salemi et al., 2023). To address real-world challenges, dynamic persona modeling using streaming user data has emerged (Lian et al., 2022; Wang et al., 2020; Yin et al., 2023; Qin et al., 2024). Departing from regeneration- and extension-based approaches, our method refines personas by integrating user behaviors and model predictions for more accurate and effective updates.

LLM for Recommendation and Behavior Prediction LLMs are increasingly applied in personalized systems like recommendation engines (Wang et al., 2023; Wu et al., 2024b; Zhang et al., 2023). Some studies integrate LLMs into traditional frameworks to enhance user modeling and contextual understanding (Liu et al., 2024; Zhang et al., 2024; Li et al., 2023b,a), while others employ LLMs directly for generating recommendations or predicting future behaviors, leveraging their persona modeling capabilities for greater adaptability and precision (Liu et al., 2023; Lyu et al., 2023; Gao et al., 2023; Dai et al., 2023). This work leverages LLMs for dynamic persona modeling and behavior prediction to capture users' evolving preferences.

7 Conclusion

In this paper, we introduce DEEPER, an effective approach to dynamic persona modeling that leverages iterative reinforcement learning and discrepancy-based refinement to continuously enhance persona quality and predictive accuracy. Comprehensive experiments demonstrate DEEPER 's effectiveness across diverse domains in dynamic user modeling. We hope DEEPER marks a significant advancement in personalized applications.

Limitations

First, this study focuses on dynamic persona modeling using discrete, quantifiable user behaviors, such as ratings, as DEEPER relies on prediction discrepancies for updates and reward computation. Other data forms, such as natural language interactions, are beyond its scope. Second, due to data availability constraints, we validate DEEPER using user rating prediction tasks, which are widely applicable and provide ample real-world sequential behavior data across domains. Nevertheless, the DEEPER framework is adaptable to broader user interaction scenarios. Finally, the insights derived from the book domain are specific to the dataset and model used, and their generalizability to other datasets and models remains uncertain.

Ethics Statement

Risks First, the datasets used in this work are publicly available and anonymized. However, we acknowledge that user behavior data, even in aggregate form, may raise privacy concerns if not handled properly. Second, our model relies on datasets that may not fully represent all user groups or domains, leading to potential biases in persona refinement and prediction. The proposed method could potentially be misused for excessive user behavior tracking or manipulative personalization. Developers and practitioners should ensure ethical use in line with user privacy regulations.

References

Sunhao Dai, Ninglu Shao, Haiyuan Zhao, Weijie Yu, Zihua Si, Chen Xu, Zhongxiang Sun, Xiao Zhang, and Jun Xu. 2023. Uncovering chatgpt's capabilities in recommender systems. In *Proceedings of the 17th ACM Conference on Recommender Systems*, pages 1126–1132.

Yunfan Gao, Tao Sheng, Youlin Xiang, Yun Xiong, Haofen Wang, and Jiawei Zhang. 2023. Chat-rec: Towards interactive and explainable llms-augmented recommender system. *arXiv preprint arXiv:2303.14524*.

Lin Gui, Cristina Gârbacea, and Victor Veitch. 2024. Bonbon alignment for large language models and the sweetness of best-of-n sampling. *arXiv preprint arXiv:2406.00832*.

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

Yu Ji, Wen Wu, Hong Zheng, Yi Hu, Xi Chen, and Liang He. 2023. Is chatgpt a good personality recognizer? a preliminary study. *arXiv preprint arXiv:2307.03952*.

Wang-Cheng Kang, Jianmo Ni, Nikhil Mehta, Maheswaran Sathiamoorthy, Lichan Hong, Ed Chi, and Derek Zhiyuan Cheng. 2023. Do llms understand user preferences? evaluating llms on user rating prediction. *arXiv preprint arXiv:2305.06474*.

Jiacheng Li, Jingbo Shang, and Julian McAuley. 2022. Uctopic: Unsupervised contrastive learning for phrase representations and topic mining. *arXiv preprint arXiv:2202.13469*.

Ruyu Li, Wenhao Deng, Yu Cheng, Zheng Yuan, Jiaqi Zhang, and Fajie Yuan. 2023a. Exploring the upper limits of text-based collaborative filtering using large language models: Discoveries and insights. *arXiv preprint arXiv:2305.11700*.

Sheng Li and Handong Zhao. 2021. A survey on representation learning for user modeling. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pages 4997–5003.

Xiangyang Li, Bo Chen, Lu Hou, and Ruiming Tang. 2023b. Ctrl: Connect tabular and language model for ctr prediction. *CoRR*.

Ruixue Lian, Che-Wei Huang, Yuqing Tang, Qilong Gu, Chengyuan Ma, and Chenlei Guo. 2022. Incremental user embedding modeling for personalized text classification. In *Icassp 2022-2022 ieee international conference on acoustics, speech and signal processing (icassp)*, pages 7832–7836. IEEE.

Junling Liu, Chao Liu, Peilin Zhou, Renjie Lv, Kang Zhou, and Yan Zhang. 2023. Is chatgpt a good recommender? a preliminary study. *arXiv preprint arXiv:2304.10149*.

Qijiong Liu, Nuo Chen, Tetsuya Sakai, and Xiao-Ming Wu. 2024. Once: Boosting content-based recommendation with both open-and closed-source large language models. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, pages 452–461.

Hanjia Lyu, Song Jiang, Hanqing Zeng, Yinglong Xia, Qifan Wang, Si Zhang, Ren Chen, Christopher Leung, Jiajie Tang, and Jiebo Luo. 2023. Llm-rec: Personalized recommendation via prompting large language models. *arXiv preprint arXiv:2307.15780*.

Bodhisattwa Prasad Majumder, Shuyang Li, Jianmo Ni, and Julian McAuley. 2019. Generating personalized recipes from historical user preferences. *arXiv preprint arXiv:1909.00105*.

Jianmo Ni, Jiacheng Li, and Julian McAuley. 2019. Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint*

conference on natural language processing (EMNLP-IJCNLP), pages 188–197.

OpenAI. 2023. Gpt-4 technical report. *Preprint*, arXiv:2303.08774.

Weicong Qin, Yi Xu, Weijie Yu, Chenglei Shen, Xiao Zhang, Ming He, Jianping Fan, and Jun Xu. 2024. Enhancing sequential recommendations through multi-perspective reflections and iteration. *arXiv preprint arXiv:2409.06377*.

Alireza Salemi, Sheshera Mysore, Michael Bendersky, and Hamed Zamani. 2023. Lamp: When large language models meet personalization. *arXiv preprint arXiv:2304.11406*.

Zhaoxuan Tan and Meng Jiang. 2023. User modeling in the era of large language models: Current research and future directions. *arXiv preprint arXiv:2312.11518*.

Yihong Tang, Bo Wang, Miao Fang, Dongming Zhao, Kun Huang, Ruifang He, and Yuexian Hou. 2023. Enhancing personalized dialogue generation with contrastive latent variables: Combining sparse and dense persona. *arXiv preprint arXiv:2305.11482*.

Yu-Min Tseng, Yu-Chao Huang, Teng-Yun Hsiao, Yu-Ching Hsu, Jia-Yin Foo, Chao-Wei Huang, and Yun-Nung Chen. 2024. Two tales of persona in llms: A survey of role-playing and personalization. *arXiv preprint arXiv:2406.01171*.

Lei Wang and Ee-Peng Lim. 2023. Zero-shot next-item recommendation using large pretrained language models. *arXiv preprint arXiv:2304.03153*.

Pengyang Wang, Kunpeng Liu, Lu Jiang, Xiaolin Li, and Yanjie Fu. 2020. Incremental mobile user profiling: Reinforcement learning with spatial knowledge graph for modeling event streams. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 853–861.

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

Yancheng Wang, Ziyan Jiang, Zheng Chen, Fan Yang, Yingxue Zhou, Eunah Cho, Xing Fan, Xiaojiang Huang, Yanbin Lu, and Yingzhen Yang. 2023. Recmind: Large language model powered agent for recommendation. *arXiv preprint arXiv:2308.14296*.

Jiaxing Wu, Lin Ning, Luyang Liu, Harrison Lee, Neo Wu, Chao Wang, Sushant Prakash, Shawn O'Banion, Bradley Green, and Jun Xie. 2024a. Rlpf: Reinforcement learning from prediction feedback for user summarization with llms. *arXiv preprint arXiv:2409.04421*.

Likang Wu, Zhi Zheng, Zhaopeng Qiu, Hao Wang, Hongchao Gu, Tingjia Shen, Chuan Qin, Chen Zhu, Hengshu Zhu, Qi Liu, Hui Xiong, and Enhong Chen. 2024b. A survey on large language models for recommendation.

Rui Xu, Xintao Wang, Jiangjie Chen, Siyu Yuan, Xinfeng Yuan, Jiaqing Liang, Zulong Chen, Xiaoqing Dong, and Yanghua Xiao. 2024. Character is destiny: Can large language models simulate persona-driven decisions in role-playing? *arXiv preprint arXiv:2404.12138*.

An Yan, Zhankui He, Jiacheng Li, Tianyang Zhang, and Julian McAuley. 2023. Personalized showcases: Generating multi-modal explanations for recommendations. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2251–2255.

Fan Yang, Zheng Chen, Ziyan Jiang, Eunah Cho, Xiaojiang Huang, and Yanbin Lu. 2023. Palr: Personalization aware llms for recommendation. *arXiv preprint arXiv:2305.07622*.

Bin Yin, Junjie Xie, Yu Qin, Zixiang Ding, Zhichao Feng, Xiang Li, and Wei Lin. 2023. Heterogeneous knowledge fusion: A novel approach for personalized recommendation via llm. In *Proceedings of the 17th ACM Conference on Recommender Systems*, pages 599–601.

Xinfeng Yuan, Siyu Yuan, Yuhan Cui, Tianhe Lin, Xintao Wang, Rui Xu, Jiangjie Chen, and Deqing Yang. 2024. Evaluating character understanding of large language models via character profiling from fictional works. *arXiv preprint arXiv:2404.12726*.

An Zhang, Yuxin Chen, Leheng Sheng, Xiang Wang, and Tat-Seng Chua. 2024. On generative agents in recommendation. In *Proceedings of the 47th international ACM SIGIR conference on research and development in Information Retrieval*, pages 1807–1817.

Junjie Zhang, Ruobing Xie, Yupeng Hou, Wayne Xin Zhao, Leyu Lin, and Ji-Rong Wen. 2023. Recommendation as instruction following: A large language model empowered recommendation approach. *arXiv preprint arXiv:2305.07001*.

Joyce Zhou, Yijia Dai, and Thorsten Joachims. 2024. Language-based user profiles for recommendation. *arXiv preprint arXiv:2402.15623*.

A Dynamic Persona Modeling Details

A.1 Persona Initialization

In this study, we employ the frozen LLM, GPT-4o-mini, to initialize user personas based on their first 10 ratings ($\mathcal{W}_0$) during the initial stage of dynamic persona modeling. The prompt used for persona initialization is presented in Table 4.

```
TASK: Infer the user's persona based
on their ratings of item_type items.
Instructions:
Below is a list of
{item_type}s that the user has rated.
Each rating ranges from 1 to 5:
{user_ratings}

Based on these, generate a user persona
without mentioning item names or rating
scores.
User Persona(at least **200 words**):
```

Table 4: Persona Initialization prompt template.

A.2 Behavior Observation and Prediction

For all user behavior prediction task, we use GPT-4o-mini to role-play the given input persona and predict user ratings on the given item list. Table 5 shows the prompt template for the prediction task.

```
TASK: Role-play the given persona and
predict what score (out of 5) you would
give to the following {item_type} list.
Instructions:  Based on the persona
{persona}, predict ratings for each item
in the list below.
{items}
## Output format:
‘‘‘json
[
{{"item_name":...,
"predict_rating":...}},
{{"item_name":...,
"predict_rating":...}},
...
]
‘‘‘
```

Table 5: Behavior Prediction prompt template.

A.3 Persona Update

In this work, the formulation of persona update stage varies by corresponding paradigms. For all baselines, we use GPT-4o-mini as backbone, while for DEEPER, we use the fine-tune model via the iterative RL training framework based on Llama3.1-8b-Instruct. For each persona update method, we design corresponding update prompt template as follows.

DEEPER The DEEPER approach based on a refinement-based paradigm with predicted and actual user ratings. Table 6 shows prompt template for persona update with DEEPER.

```
TASK:
Refine the old user persona based on
differences between predicted and
actual ratings of {item_type} items.
Instructions:
Below is the existing persona inferred
from past behavior:
{old_persona}
Below is the comparison of predicted
ratings (based on the old persona)
versus actual ratings:
{predict_and_actual_user_ratings}
Reflect on these differences and
generate a refined user persona without
mentioning item names or rating scores.
Refined User Persona:
```

Table 6: DEEPER Persona Refinement prompt template.

FullRegen In the FullRegen, we fully regenerate the user's persona whenever new ratings are provided. This method does not consider the prior persona and instead creates a fresh representation based all observed ratings. Table 7 shows the prompt template for persona update with FullRegen.

```
TASK: Infer the user's persona based
on their ratings of item_type items.
Instructions:
Below is a list of
{item_type}s that the user has rated.
Each rating ranges from 1 to 5:
{Full_user_ratings}

Based on these, generate a user persona
without mentioning item names or rating
scores.
User Persona:
```

Table 7: FullRegen Persona Update prompt template.

SlideRegen In the SlideRegen method, we regenerate personas based on their recent ratings of {item_type} items(latest window). Table 8 shows the prompt template for persona update with SlideRegen.

```
TASK: Infer the user's persona based
on their ratings of item_type items.
Instructions:
Below is a list of
{item_type}s that the user has rated.
Each rating ranges from 1 to 5:
{Slide_user_ratings}

Based on these, generate a user persona
without mentioning item names or rating
scores.
User Persona:
```

Table 8: SlideRegen Persona Update prompt template.

IncUpdate In the IncUpdate, the user's persona is dynamically updated by integrating new ratings with their existing persona. Table 9 shows prompt template for persona update with IncUpdate.

```
TASK: Integrate the user's most recent
ratings of {item_type}
items into their existing persona to
generate an updated persona.
Instructions:
Below is the existing persona based on
prior behaviors:
{old_persona}
Below is a list of recent {item_type}s
that the user has rated.
Each rating ranges from 1 to 5:
{user_ratings}
Based on these, integrate the new
features from the recent ratings into
the existing persona.
Updated Persona:
```

Table 9: IncUpdate Persona Update prompt template.

HierMerge The HierMerge method combines both long-term personas and short-term personas hierarchically. Table 10 shows prompt template for persona update with HierMerge.

```
# Prompt 1:
TASK: Infer the user's persona based on
their ratings of {item_type} items.
Instructions:
Below is a list of
{item_type}s that the user has rated.
Each rating ranges from 1 to 5:
{user_ratings}
Based on these, generate a user
persona without mentioning item names
or rating scores.
User Persona:

# Prompt 2:  TASK: Update the long-term
persona by merging it with the
newly generated short-term persona.
Instructions:
Below is the existing long-term
persona based on prior
behaviors:{long_term_persona}
Below is the newly generated
short-term persona based on recent
behaviors:{short_term_persona}
Merge the short-term persona into
the long-term persona to capture
both historical stability and recent
dynamics.
The updated persona should reflect both
long-term
preferences and recent changes without
losing consistency.
Updated Long-Term Persona:
```

Table 10: HierMerge Persona Update prompt template.

Dataset	Abbreviation	Usage	# Users in Train	# Users in Eval	# Train Examples
Food.com Recipes - and Interactions	Recipe	Train/Eval	1000	356	A
Amazon Review Data (2018) - Books	Book	Train/Eval	3000	897	B
Amazon Review Data (2018) - Clothing Shoes and Jewelry	Clothing Shoes and Jewelry	Train/Eval	300	243	C
Google Local Data (2021) - New York	Local Business	Train/Eval	2500	826	D
MovieLens - 20M Dataset	MovieLens	Train/Eval	3000	1000	E
Amazon Review Data (2018) - Art Crafts and Sewing	Art Crafts and Sewing	Eval	-	86	F
Amazon Review Data (2018) - Automative	Automative	Eval	-	143	G
Amazon Review Data (2018) - Sports and Outdoors	Sports and Outdoors	Eval	-	236	H
Amazon Review Data (2018) - Grocery and Gourmet Food	Grocery and Gourmet Food	Eval	-	185	I

Table 11: Details of Datasets Used in Experiments.

B Dataset Details

B.1 User Details

We utilize four publicly available and well-known datasets, selecting a total of 10 domains. From six domains, we randomly sampled a total of 14,959 users with at least 50 ratings. Among these, 10,800 users are used for constructing the training data, and 4,159 users are used for constructing the testing data. Additionally, to evaluate the generalization ability of the methods, we sampled 650 users with at least 50 ratings from four unseen domains to construct an additional test set. Each user's 50 rating behaviors are sorted by timestamp and divided into five sequences of length 10, simulating multi-round online user interactions. The detailed user sampling statistics are in Table 11:

B.2 Training Data Construction

In **Iteration 1**, a total of 10,800 context data points are constructed, each corresponding to the first persona refinement step for each user. For each context, 15 candidate personas are randomly sampled using the Llama3.1-8b-Instruct model, with inference parameters set as follows: `temperature=1` (to ensure diversity among the candidates), `top_p=0.4` (to control the cumulative probability of tokens), and `repetition_penalty=1.1` (to prevent repetition in the generated output). The boundaries for positive and negative reward sets are set to 0.5 and 0, with a margin of 0.5. In total, 34,782 DPO preference pairs are constructed, with 10% randomly selected for the validation set. This data is used to train **Model 1**.

In **Iteration 2**, **Model 1** is first used to generate outputs for the 10,800 context data points from Iteration 1, completing the first persona update for each user. These results, in turn, are used to construct a second set of 10,800 context data points for the second persona refinement. These are then combined with the 10,800 context data points constructed in the first iteration, resulting in a total of 21,600 context data points for sampling in the second iteration. For each context, 15 candidate personas are again randomly sampled using **Model 1**, with the same inference parameters as in Iteration 1. The boundaries for positive and negative reward sets are set to 0.5 and 0, with a margin of 0.8. A total of 28,612 new DPO preference pairs are generated. Additionally, 5,000 preference pairs with a margin greater than 0.8 are randomly selected from Iteration 1 to be included in the training set, mitigating the issue of catastrophic forgetting. This results in a total of 33,612 DPO preference pairs, with 10% randomly selected for the validation set, used to train **Model 2**.

Parameter	Value	Description
Model Name or Path	`Llama-3.1-8B-Instruct`	Path to the model
Finetuning Type	`lora`	Type of finetuning
Training Stage	`dpo`	Current training stage
LoRA Target	`all`	LoRA target layers
LoRA Rank	16	LoRA rank
LoRA Alpha	32	LoRA alpha
LoRA Dropout	0.2	LoRA dropout rate
Preference Beta	0.2	Preference loss beta
Preference Loss Type	`sigmoid`	Type of preference loss
Preference Finetune Rate	0.1	Preference finetuning rate
Maximum Sequence Length	2048	Maximum input sequence length
Training Batch Size	4	Batch size per device during training
Gradient Accumulation Steps	8	Steps for gradient accumulation
Learning Rate	5.0e-06	Learning rate
Number of Epochs	4.0	Total number of training epochs
Learning Rate Scheduler	`cosine`	Learning rate scheduling strategy
Warmup Steps	250	Warmup steps before full learning rate
Maximum Gradient Norm	1.0	Maximum norm for gradient clipping
BF16 Precision	`true`	Use BF16 precision
Optimizer	`adamw_torch`	Type of optimizer
Validation Size	0.1	Fraction of data used for validation
Evaluation Batch Size	4	Batch size per device during evaluation
Evaluation Strategy	`steps`	Evaluation scheduling strategy
Evaluation Steps	100	Steps between evaluations

Table 12: Hyperparameter Details for Training

C Training Details

C.1 Hyperparameter Details

The hyperparameters used in the training process are summarized in Table 12.

C.2 Loss Function Details

In this section, we provide the detailed formulations of the training loss functions combined with DPO loss and SFT loss.

DPO Loss The DPO loss optimizes the model by leveraging user preference signals to align persona refinements with higher rewards. The loss is defined as:

$$\mathcal{L}_{\text{DPO}}(\pi_\theta; \pi_{\text{ref}}) = -\,\mathbb{E}_{(x,y_w,y_l)\sim\mathcal{D}} \left[\log \sigma \left(\beta \log \frac{\pi_\theta(y_w \mid x)}{\pi_{\text{ref}}(y_w \mid x)} - \beta \log \frac{\pi_\theta(y_l \mid x)}{\pi_{\text{ref}}(y_l \mid x)}\right)\right]. \quad (11)$$

The policy distribution of the model being trained, parameterized by θ. It represents the probability of generating specific outputs conditioned on the input x. The reference model's policy distribution, used as a baseline for comparison. It is typically a pretrained model or a checkpoint used to stabilize training. A tuple sampled from the dataset $\mathcal{D}$, where:

- x: The input prompt or context.
- y_w: better personas, corresponding to more optimal update directions.
- y_l: poor personas, corresponding to less effective update directions.

A scaling factor that controls the sensitivity of the loss function to the difference between the preferred and less preferred outputs. Larger values emphasize the contrast between the two. The conditional probabilities of the preferred (y_w) and less preferred (y_l) outcomes under the current model. The conditional probabilities of the preferred and less preferred outcomes under the reference model. The loss is computed as an average over the entire dataset $\mathcal{D}$, which contains human-annotated preference pairs (x, y_w, y_l).

SFT Loss The SFT loss is used to aline the model output with high-quality refined-persona candidates. The loss is computed as the negative log-likelihood of the reference outputs:

$$\mathcal{L}_{\text{SFT}}(\pi_\theta) = -\mathbb{E}_{(x,y_w)\sim\mathcal{D}}\big[\log \pi_\theta(y_w|x)\big]. \quad (12)$$

where $\mathcal{D}$ is the dataset of supervised examples, x represents the input context, and y is the corresponding ground truth persona refinement.

Combined Loss for Iterative Training To achieve robust refinement across iterations, we combine the SFT loss and DPO loss :

$$\mathcal{L}(\pi_\theta; \pi_{\text{ref}}) = \mathcal{L}_{\text{DPO}}(\pi_\theta; \pi_{\text{ref}}) + \mathcal{L}_{\text{SFT}}(\pi_\theta). \quad (13)$$

D Dynamic Persona Modeling Task

In the main experiment, we focused on the dynamic persona modeling task, where different methods are employed to perform four rounds of updates on the test set. These iterative updates provided specific values for predicting future user behavior, enabling us to assess the accuracy and effectiveness of each method in forecasting user actions. By evaluating the Mean Absolute Error (MAE) across various domains before and after refinement, we are able to determine the improvements achieved through each method. The results, detailed in Table 13,Table 14,Table 15,Table 16,Table 17, highlight the performance gains and validate the superiority of the proposed approaches in enhancing prediction accuracy.

Domain	$\varepsilon_{1\|\mathcal{S}_0}$	$\varepsilon_{2\|\mathcal{S}_1}$	$\varepsilon_{3\|\mathcal{S}_2}$	$\varepsilon_{4\|\mathcal{S}_3}$	$\varepsilon_{5\|\mathcal{S}_4}$
Art Crafts and Sewing	0.76	0.43	0.44	0.41	0.40
Automative	0.84	0.61	0.60	0.59	0.57
Book	1.00	0.79	0.70	0.68	0.67
Clothing Shoes and Jewelry	1.00	0.88	0.84	0.75	0.74
Grocery and Gourmet Food	1.19	0.89	0.79	0.77	0.73
Local Business	1.04	0.80	0.70	0.70	0.69
Movie	0.87	0.73	0.73	0.72	0.72
Movies and TV	1.12	0.98	0.83	0.79	0.77
Recipe	0.91	0.72	0.65	0.59	0.58
Sports and Outdoors	0.91	0.80	0.68	0.66	0.66

Table 13: Deeper Results

Domain	$\varepsilon_{1\|\mathcal{S}_0}$	$\varepsilon_{2\|\mathcal{S}_1}$	$\varepsilon_{3\|\mathcal{S}_2}$	$\varepsilon_{4\|\mathcal{S}_3}$	$\varepsilon_{5\|\mathcal{S}_4}$
Art Crafts and Sewing	0.76	0.77	0.74	0.69	0.75
Automative	0.84	0.92	0.92	0.92	0.93
Book	1.00	1.03	1.00	1.00	1.00
Clothing Shoes and Jewelry	1.00	1.03	1.03	1.03	1.03
Grocery and Gourmet Food	1.19	1.20	1.17	1.14	1.13
Local Business	1.04	1.04	1.02	1.03	1.02
Movie	0.87	0.83	0.84	0.84	0.85
Movies and TV	1.12	1.12	1.13	1.13	1.11
Recipe	0.91	0.92	0.83	0.84	0.85
Sports and Outdoors	0.91	0.93	0.85	0.88	0.88

Table 14: FullRegen (GPT-4o-mini) Results

Domain	$\varepsilon_{1\|\mathcal{S}_0}$	$\varepsilon_{2\|\mathcal{S}_1}$	$\varepsilon_{3\|\mathcal{S}_2}$	$\varepsilon_{4\|\mathcal{S}_3}$	$\varepsilon_{5\|\mathcal{S}_4}$
Art Crafts and Sewing	0.76	0.81	0.72	0.66	0.74
Automative	0.84	0.96	0.91	0.91	0.90
Book	1.00	1.06	1.02	1.03	1.02
Clothing Shoes and Jewelry	1.00	1.09	1.08	1.02	1.07
Grocery and Gourmet Food	1.19	1.14	1.13	1.13	1.09
Local Business	1.04	1.06	1.05	1.03	1.03
Movie	0.87	0.84	0.85	0.87	0.86
Movies and TV	1.12	1.14	1.16	1.17	1.14
Recipe	0.91	0.92	0.89	0.87	0.87
Sports and Outdoors	0.91	0.94	0.87	0.92	0.93

Table 15: SlideRegen (GPT-4o-mini) Results

Domain	$\varepsilon_{1\|\mathcal{S}_0}$	$\varepsilon_{2\|\mathcal{S}_1}$	$\varepsilon_{3\|\mathcal{S}_2}$	$\varepsilon_{4\|\mathcal{S}_3}$	$\varepsilon_{5\|\mathcal{S}_4}$
Art Crafts and Sewing	0.76	0.71	0.70	0.59	0.66
Automative	0.84	0.88	0.81	0.87	0.82
Book	1.00	0.96	0.94	0.92	0.94
Clothing Shoes and Jewelry	1.00	1.00	0.97	0.93	0.92
Grocery and Gourmet Food	1.19	1.10	1.05	1.05	1.04
Local Business	1.04	0.97	0.94	0.93	0.91
Movie	0.87	0.76	0.76	0.76	0.75
Movies and TV	1.12	1.06	1.06	1.04	1.05
Recipe	0.91	0.91	0.84	0.84	0.81
Sports and Outdoors	0.91	0.89	0.85	0.85	0.83

Table 16: IncUpdate (GPT-4o-mini) Results

Domain	$\varepsilon_{1\|\mathcal{S}_0}$	$\varepsilon_{2\|\mathcal{S}_1}$	$\varepsilon_{3\|\mathcal{S}_2}$	$\varepsilon_{4\|\mathcal{S}_3}$	$\varepsilon_{5\|\mathcal{S}_4}$
Art Crafts and Sewing	0.76	0.75	0.74	0.67	0.74
Automative	0.84	0.88	0.86	0.94	0.87
Book	1.00	1.03	0.98	0.97	0.96
Clothing Shoes and Jewelry	1.00	1.04	1.01	1.04	1.04
Grocery and Gourmet Food	1.19	1.19	1.17	1.18	1.16
Local Business	1.04	1.04	1.00	1.01	1.00
Movie	0.87	0.82	0.83	0.83	0.82
Movies and TV	1.12	1.12	1.11	1.10	1.09
Recipe	0.91	0.94	0.87	0.83	0.83
Sports and Outdoors	0.91	0.90	0.89	0.91	0.88

Table 17: HierMerge (GPT-4o-mini) Results

E What Enables DEEPER's Effectiveness

Below, we present an in-depth analysis of the mechanisms underlying DEEPER 's effectiveness.

E.1 Proving Effectiveness of Direction search

Firstly, we prove the effectiveness of the direction search method by comparing its performance with a direct refinement using the frozen model(GPT-4o-mini and the base model, Llama3.1-8b-Instruct), through a single round of refinement. The details of the experimental results are as follows Label 18.

- **Baseline 1** Directly Refine personas with the base model (Llama3.1-8b-Instruct)
- **Baseline 2** Directly Refine personas with the more powerful model (GPT-4o-mini)
- **DEEPER** Refine personas with auto-direction search mechanism

	Before Update	After Update		
Domain	$\varepsilon_{1\|\mathcal{S}_0}$	$\varepsilon_{2\|\mathcal{S}_1}$ (DEEPER)	$\varepsilon_{2\|\mathcal{S}_1}$ (GPT-4o-mini)	$\varepsilon_{2\|\mathcal{S}_1}$ (Llama3.1-8b-Instruct)
A	0.91	0.72	0.99	1.07
B	1.01	0.79	1.20	1.21
C	1.03	0.88	1.09	1.14
D	1.04	0.80	1.20	1.19
E	1.18	0.98	1.24	1.19
F	0.85	0.73	0.84	0.87
G	0.74	0.43	0.95	0.89
H	0.85	0.61	1.02	1.08
I	1.26	0.89	1.26	1.19
J	0.96	0.80	1.04	1.02

Table 18: Future behaviour prediction errors before and after one-step refinement with DEEPER and the frozen models.

E.2 Proving the Effectiveness of Balanced Reward

In this analysis, we aim to demonstrate the effectiveness of the balanced reward strategy by comparing it against two baseline reward settings. Specifically, we evaluate how different reward configurations influence the performance of the model during the refinement process.

Baseline Reward Settings We establish two baseline configurations to assess the impact of reward settings:

- **Baseline 1: Future Advancement Only**
  In this setting, the reward at each timestep t is solely based on future advancement. Mathematically, this is defined as:

$$r_t = r_{\text{fut}} = r_t^{\text{fut}}. \tag{14}$$

- **Baseline 2: Decayed Rewards**
  Here, we incorporate past, current, and future rewards with decay factors applied to past and current rewards. The reward at timestep t is calculated as:

$$r_t = r_{\text{decay}} = 0.25 \cdot r_t^{\text{prev}} + 0.5 \cdot r_t^{\text{curr}} + r_t^{\text{fut}}. \tag{15}$$

  where the decay factor $y = 0.5$ is applied to both past and current rewards.

Our Reward Setting: Balanced Rewards Our proposed reward setting balances the three components—past, current, and future—without applying decay factors. The reward at timestep t is defined as:

$$r_t = r_t^{\text{prev}} + r_t^{\text{curr}} + r_t^{\text{fut}}. \tag{16}$$

This approach ensures that all three goals are equally considered during the refinement process.

Experimental Results The experimental results comparing the baseline reward settings with our balanced reward strategy are presented in Table 19, which showcase future prediction errors across various domains before and after one-step refinement under different reward configurations.

	Before Update	After Update		
Domain	$\varepsilon_{1\|\mathcal{S}_0}$	$\varepsilon_{2\|\mathcal{S}_1}$ (DEEPER)	$\varepsilon_{2\|\mathcal{S}_1}$ (r_{fut})	$\varepsilon_{2\|\mathcal{S}_1}$ (r_{decay})
A	0.91	0.72	0.81	0.74
B	1.01	0.79	0.86	0.84
C	1.03	0.88	0.95	0.95
D	1.04	0.80	0.88	0.84
E	1.18	0.98	1.03	1.03
F	0.85	0.73	0.81	0.77
G	0.74	0.43	0.59	0.51
H	0.85	0.61	0.83	0.68
I	1.26	0.89	0.94	0.92
J	0.96	0.80	0.85	0.83

Table 19: Balanced reward (DEEPER) vs. Baseline reward settings results

E.3 Proving the Effectiveness of Iterative RL Training

In this analysis, we aim to demonstrate the effectiveness of iterative RL training by comparing the performance of the model after one iteration of refinement versus two iterations. This comparison helps to understand whether additional refinement iterations contribute to improved model performance.

Baseline To evaluate the impact of iterative RL training, we establish two baseline configurations:

- **Baseline 1: Single Iteration**
  The model undergoes one iteration of training.

- **Baseline 2: Two Iterations**
  The model undergoes two consecutive iterations of training to assess whether additional refinement leads to further performance gains.

Experimental Results The experimental results comparing single and double iterations of RL-based refinement are presented in Table 20.

	Before Update	After Update	
Domain	$\varepsilon_{1\|\mathcal{S}_0}$	$\varepsilon_{2\|\mathcal{S}_1}$ (DEEPER (Iter1))	$\varepsilon_{2\|\mathcal{S}_1}$ (DEEPER (Iter2))
A	0.91	0.84	0.72
B	1.01	0.93	0.79
C	1.03	0.99	0.88
D	1.04	0.98	0.80
E	1.18	1.12	0.98
F	0.85	0.78	0.73
G	0.74	0.65	0.43
H	0.85	0.87	0.61
I	1.26	1.03	0.89
J	0.96	0.91	0.80

Table 20: Iterative RL Training (DEEPER) Results comparison: Single vs. Double Training Iterations

E.4 Proving the Effectiveness of Introducing Prediction Results in Refinement Paradigm

In this analysis, we aim to demonstrate the effectiveness of incorporating prediction results into the paradigm. Specifically, we leverage the iterative RL training framework of DEEPER to enhance IncUpdate(Inc-FT), which is the best performing baseline and based on paradigm of Persona Extension. This enable auto-direction search in traditional dynamic persona paradigm which does not involve prediction results into observations. The comparison between DEEPER and Inc-FT helps to understand whether integrating prediction results helps direction search.

Experimental Results The experimental results are presented in Table 21.

	Before Update	After Update	
Domain	$\varepsilon_{1\|\mathcal{S}_0}$	$\varepsilon_{2\|\mathcal{S}_1}$ (DEEPER)	$\varepsilon_{2\|\mathcal{S}_1}$ (Inc-FT))
Recipe	0.91	0.72	0.77
Book	1.01	0.79	0.85
Clothing Shoes and Jewelry	1.03	0.88	0.95
Local Business	1.04	0.80	0.81
Movies and TV	1.18	0.98	1.02
Movie	0.85	0.73	0.79
Art Crafts and Sewing	0.74	0.43	0.62
Automative	0.85	0.61	0.78
Grocery and Gourmet Food	1.26	0.89	0.97
Sports and Outdoors	0.96	0.91	0.81

Table 21: Effectiveness of Introducing Prediction Results: DEEPER vs. IncUpdate (Inc-FT)

Dimensions	High-Frequency Terms and Frequency
Story & Plot	story (759), experience (596), reader (579), narrative (566), storytelling (445), development (440), plot (286), adventure (196), fantasy (187), suspense (156), mystery (155), action (150), passion (149), thriller (109), journey (92), arc (75), drama (66), protagonist (58), redemption (54)
Genre & Theme	theme (655), genre (647), romance (388), aspect (406), level (379), content (329), complexity (325), depth (325), world (277), novel (212), topic (189), element (231), idea (189), nuance (148), literature (137), nature (118), issue (122), setting (88), balance (97), thought (96)
Author & Character	author (143), quality (155), character (600), characteristic (93), identity (70), protagonist (58)
Emotion & Experience	affinity (217), appreciation (700), experience (596), willingness (685), desire (563), love (410), enthusiasm (377), resonance (352), emotion (185), escapism (203), curiosity (182), expectation (167), favor (153), enjoyment (136), excitement (121), comfort (117)
User Behavior Traits	range (568), star (518), growth (318), perspective (279), engagement (250), title (247), tendency (185), habit (155), exploration (140), investment (86), variety (67)
User Personality & Values	willingness (685), preference (581), individual (388), value (221), self (183), discerning (183), personality (132), creativity (94), empathy (85), adaptability (90)
Social & Cultural Context	life (198), community (164), time (181), boundary (99), need (112), culture (77), knowledge (88), learning (95), justice (61)
Relationship & Connection	connection (474), relationship (411), choice (98), family (67), interaction (57)

Table 22: Important Profiling Dimensions in the Book Domain

F Persona Probing

F.1 Important Profiling Dimensions in Book Domain

Table 22 summarizes the key profiling dimensions for users in the book domain, along with the high-frequency terms and their frequencies within each dimension. These dimensions include "Story & Plot," "Genre & Theme," "Author & Character," among others, which encapsulate critical aspects of user preferences and behaviors. The table highlights the most commonly used terms, such as "story," "experience," and "reader" under the "Story & Plot" dimension, providing insights into what users value when engaging with book-related content.

F.2 Insights into User Group Characteristics

Table 23 illustrates the unique adjectives frequently associated with specific user groups, providing a detailed view of the preferences that distinguish these groups. For instance, Group 1 exhibits traits such as "romantic" and "dedicated," while Group 4 emphasizes "practical" and "cultural" preferences. These findings underscore the variation in user characteristics, enabling targeted persona optimization based on group-specific attributes.

User Groups	Unique High-Frequency Adjectives
Group 1	romantic, paranormal, voracious, dedicated, afraid
Group 2	notable, humorous, unconventional, close, entertaining
Group 3	dramatic, dedicated, resonant
Group 4	practical, spiritual, historical, cultural, likely, playful, dynamic, inspirational, close
Group 5	thoughtful, non, suspenseful, fiction, dynamic, engaging, immersive

Table 23: Group-level preference for users

G Case Study: A User in Book Domain

To deeply evaluate the performance of different persona updates methods for dynamic persona modeling, we selected a single user from the book domain. This domain provides a complex and rich context, as users often demonstrate evolving preferences, diverse genre interests, and emotional connections with books over time. In case study, we focus on the improvements of future prediction task over four update rounds and the evolution of user's personas with DEEPER as the update method.

G.1 Dynamic Persona Modeling

We first compares five persona update methods: DEEPER, FullRegen, SlideRegen, IncUpdate, and HierMerge, focusing on the evolution of the user's persona across 4 update steps and evaluate their effectiveness based on the future prediction error (MAE) at each step, as shown in Table 24. The results of this case demonstrates that DEEPER consistently reduces prediction error across refinement steps, achieving continual persona optimization, while all baseline methods not only fail to improve but also degrade persona quality over time.

Domain	$\varepsilon_{1\|\mathcal{S}_0}$	$\varepsilon_{2\|\mathcal{S}_1}$	$\varepsilon_{3\|\mathcal{S}_2}$	$\varepsilon_{4\|\mathcal{S}_3}$	$\varepsilon_{5\|\mathcal{S}_4}$
DEEPER	1.1	1.0	0.5	0.3	0.2
FullRegen	1.1	1.6	1.5	1.5	1.7
SlideRegen	1.1	1.4	1.8	1.9	1.5
IncUpdate	1.1	1.3	1.6	1.7	1.2
HierMerge	1.1	1.7	1.8	1.9	1.3

Table 24: Future prediction error across 4 persona updates.

G.2 Persona Evolution

In this section, we display the evolution of the user's persona and corresponding observations in each window over four persona update rounds with DEEPER (Table 25, Table 26, Table 27, Table 28, Table 29).

Examples of Persona Evolution of DEEPER (*A User from Book Domain*)	
$\mathcal{S}_0$ and observartions in $\mathcal{W}_1$	"The user appears to be an individual with a strong appreciation for both creative expression and practical knowledge. Their high ratings for works that blend storytelling with deeper themes suggest a reflective and imaginative personality. They likely enjoy narratives that not only entertain but also provoke thought and inspire personal growth. The consistent high ratings for books related to finance and funding indicate a keen interest in practical skills and knowledge, suggesting that they may be entrepreneurial or involved in creative industries where funding and financial acumen are crucial. This duality in their reading preferences points to a well-rounded individual who values both the arts and the practicalities of life. The user's enjoyment of stories that involve mystery and justice indicates a fascination with problem-solving and perhaps a desire for moral clarity in complex situations. Their varied interests suggest they are open-minded and curious, likely seeking to expand their horizons through diverse genres. Overall, this user embodies a blend of creativity and pragmatism, with a penchant for narratives that inspire, educate, and entertain, reflecting a holistic approach to personal and professional development."
	item_list = ["Insights of an Ordinary Man","Aura Advantage: How the Colors in Your Aura Can Help You Attain What You Desire and Attract Success","Georgia Bound", "Fifth Life of the Cat Woman", "Lady Justice and the Assassin", "Lord James Harrington and the Winter Mystery (Volume 1)", "Elemental Magic", "Enchantments of the Heart","The Grimm Diaries Prequels 1-6", "The Quest for the Crystal Skulls"] actual_ratings = [5, 5, 5, 5, 5, 5, 4, 3, 5, 5] predict_ratings = [4, 3, 3, 4, 5, 4, 3, 3, 3, 4]

Table 25: Examples of $\mathcal{S}_0$ and observartions in $\mathcal{W}_1$

Examples of Persona Evolution of DEEPER (*A User from Book Domain*)	
$\mathcal{S}_1$ and observartions in $\mathcal{W}_2$	"The user appears to be a creative and open-minded individual with a strong appreciation for diverse forms of expression. Their high ratings for a wide range of books, including those that blend fiction with spiritual or self-help themes, suggest a deep interest in personal growth and self-discovery. This inclination towards inspirational and motivational content may indicate that the user is seeking guidance or validation in their own life journey. Their consistent 5-star ratings across various genres, including mystery and fantasy, demonstrate a willingness to explore new ideas and perspectives. The user's appreciation for practical knowledge and skills, as evident in their interest in finance and funding, highlights their pragmatic side and potential entrepreneurial or creative pursuits. Their enjoyment of stories that involve justice and problem-solving suggests a strong sense of morality and a desire for fairness. Overall, this user embodies a unique blend of creativity, curiosity, and practicality, with a penchant for narratives that inspire, educate, and entertain. They seem to value diversity and are not afraid to explore unconventional ideas, making them a highly open-minded and adaptable individual."
	item_list= ["A Hidden Fire: Elemental Mysteries Book 1", "It's Just a Little Crush: A Lizzie Hart Mystery (The Lizzie Hart Mysteries) (Volume 1)","The Devil Drinks Coffee", "The Depot: When Life and Death Cross Tracks","Lady Justice and the Lottery (Volume 14)", "Wuthering Heights (Penguin Classics)","The Ambitious Card (An Eli Marks Mystery) (Volume 1)","Miracles & Mischief: Meringue for your Brain", "The Color of Greed: Raja Williams Series (Raja Williams Mystery Thriller Series)", "Thursday Morning Breakfast (and Murder) Club"] actual_ratings= [5, 5, 5, 5, 5, 5, 5, 5, 5, 5] predict_ratings = [4, 4, 3, 4, 5, 5, 4, 3, 4, 4]

Table 26: Examples of $\mathcal{S}_1$ and observartions in $\mathcal{W}_2$.

Examples of Persona Evolution of DEEPER (*A User from Book Domain*)	
$\mathcal{S}_2$ and observartions in $\mathcal{W}_3$	"The user appears to be an extremely enthusiastic and optimistic individual who thoroughly enjoys a wide range of books across various genres. Their high ratings for a vast array of titles, including those that blend fiction with spiritual or self-help themes, suggest a deep interest in personal growth and self-discovery. This inclination towards inspirational and motivational content may indicate that the user is seeking guidance or validation in their own life journey. Their consistent 5-star ratings across various genres, including mystery and fantasy, demonstrate a willingness to explore new ideas and perspectives. The user's appreciation for practical knowledge and skills, as evident in their interest in finance and funding, highlights their pragmatic side and potential entrepreneurial or creative pursuits. Their enjoyment of stories that involve justice and problem-solving suggests a strong sense of morality and a desire for fairness. However, their tendency to rate books at the maximum score, regardless of predicted ratings, suggests a more impulsive and emotional approach to evaluating content. This user seems to value diversity and is not afraid to explore unconventional ideas, making them a highly open-minded and adaptable individual. Their love for storytelling and enthusiasm for new experiences may indicate a creative or artistic personality, always seeking inspiration and joy in their reading habits. Overall, this user embodies a unique blend of creativity, curiosity, and practicality, with a penchant for narratives that inspire, educate, and entertain."
	item_list = ["Lust, Money & Murder - Book 1: A Female Secret Service Agent Takes on an International Criminal - Kindle edition", "Lust, Money & Murder - Books 1, 2 & 3: A Female Secret Service Agent Takes on an International Criminal","The Hitman's Lover (Jack Scully Thrillers) (Volume 1)", "Initiation, Human and Solar","Consciousness of the Atom", "Contagious: Why Things Catch On","The Poetical Works of Elizabeth Barrett Browning, Vol. I eBook","The Original Folk and Fairy Tales of Grimm Brothers: The Complete First Edition", "My Short, Happy Life In 'Jeopardy!'", "22/11/63"] actual_ratings= [5, 5, 5, 4, 5, 5, 5, 3, 5, 3] predict_ratings = [5, 5, 5, 5, 5, 5, 5, 5, 5, 5]

Table 27: Examples of $\mathcal{S}_2$ and observartions in $\mathcal{W}_3$

Examples of Persona Evolution of DEEPER (*A User from Book Domain*)	
$\mathcal{S}_3$ and observartions in $\mathcal{W}_4$	"The user appears to be a highly enthusiastic and optimistic individual who enjoys a wide range of books across various genres. Their love for storytelling and enthusiasm for new experiences suggests a creative or artistic personality, always seeking inspiration and joy in their reading habits. However, their tendency to rate books at the maximum score, regardless of predicted ratings, suggests a more impulsive and emotional approach to evaluating content.Their consistent 5-star ratings for a variety of titles, including those that blend fiction with spiritual or self-help themes, suggest a deep interest in personal growth and self-discovery. This inclination towards inspirational and motivational content may indicate that the user is seeking guidance or validation in their own life journey. Their appreciation for practical knowledge and skills, as evident in their interest in finance and funding, highlights their pragmatic side and potential entrepreneurial or creative pursuits.However, their 3-star rating for"The Original Folk and Fairy Tales of Grimm Brothers" and 3-star rating for"22/11/63" suggests that they may not always appreciate complex or darker themes, and may prefer more lighthearted or uplifting content. This preference for happier endings and more optimistic themes may be a key aspect of their reading preferences.Their enjoyment of stories that involve justice and problem-solving suggests a strong sense of morality and a desire for fairness. However, their tendency to rate books at the maximum score, regardless of predicted ratings, suggests that they may be more focused on the emotional resonance of a story rather than its literary merit or technical quality.Overall, this user embodies a unique blend of creativity, curiosity, and practicality, with a penchant for narratives that inspire, educate, and entertain. They appear to be a highly open-minded and adaptable individual, always seeking new experiences and inspiration in their reading habits."
	item_list= ["Into the Light (Easyread Large Bold Edition): Real Life Stories About Angelic Visits, Visions of the Afterlife, and Other Pre-Death Experiences", "Catnapped: A Klepto Cat Mystery (Volume 1)", "Dancing In The Moonlight (The Cowboys of Cold Creek)", "11/22/63 (Thorndike Press Large Print Core)","The Medium: An Emily Chambers Spirit Medium Novel (Volume 1)","Passion, Power & Sin - Book 1: The Victim of a Global Internet Scam Plots Her Revenge - Kindle edition", "Passion, Power & Sin - Books 1-5 (Book 1 Free)", "Season of the Witch","Murder in the South of France: A Maggie Newberry Mystery, Vol. 1 (Maggie Newberry Mysteries)","World hunger: Twelve myths"] actual_ratings= [5, 4, 5, 3, 5, 5, 5, 5, 4, 5] predict_ratings=[5, 5, 5, 3, 5, 5, 5, 5, 5, 4]

Table 28: Examples of $\mathcal{S}_3$ and observartions in $\mathcal{W}_4$

Examples of Persona Evolution of DEEPER (*A User from Book Domain*)	
$\mathcal{S}_4$ and observartions in $\mathcal{W}_5$	"Based on the differences between predicted and actual ratings, the user reveals a more nuanced and complex individual. They appear to be a highly enthusiastic and optimistic person, always seeking inspiration and joy in their reading habits. Their love for storytelling and enthusiasm for new experiences suggest a creative or artistic personality, but with a notable emphasis on emotional resonance and personal connection.Their consistent 5-star ratings for a wide range of titles, including those with inspirational and motivational themes, suggest a deep interest in personal growth and self-discovery. However, their ratings also indicate a willingness to engage with more lighthearted and entertaining content, such as romance and mystery novels. This preference for happier endings and more optimistic themes may be a key aspect of their reading preferences.Their enjoyment of stories that involve justice and problem-solving suggests a strong sense of morality and a desire for fairness. However, their tendency to rate books at the maximum score, regardless of predicted ratings, suggests that they may be more focused on the emotional impact of a story rather than its literary merit or technical quality.Interestingly, their ratings also suggest a willingness to engage with complex and thought-provoking topics, such as world hunger and global issues. This inclination towards inspirational and motivational content, combined with a desire for practical knowledge and skills, highlights their pragmatic side and potential entrepreneurial or creative pursuits.They may not always appreciate complex or darker themes, and may prefer more uplifting content. However, their willingness to engage with more serious topics, such as world hunger, suggests a depth and nuance to their personality that is not immediately apparent.Overall, this user embodies a unique blend of creativity, curiosity, and practicality, with a penchant for narratives that inspire, educate, and entertain. They appear to be a highly open-minded and adaptable individual, always seeking new experiences and inspiration in their reading habits. Their refined persona reveals a more complex and multifaceted individual, with a deep appreciation for emotional resonance, personal growth, and practical knowledge."
	item_list= ["The Quickening of America: Rebuilding Our Nation, Remaking Our Lives", "The Da Vinci Code (Robert Langdon)","The Accidental Cop (Ben Colder Mystery)", "Lingering Echoes", "Rumors (Lingering Echoes)","Murder in Paris (The Maggie Newberry Mystery Series)", "Stilettos & Scoundrels", "Bitty And The Naked Ladies (The Sherri Travis Mystery Series) eBook", "Sati and the Rider: A Satyana Mystery (Volume 1)","Fleur deKey: a French Quarter Mystery (The Foundation Mystery Series) (Volume 1)"] actual_ratings= [5, 4, 5, 5, 5, 5, 4, 5, 5, 5] predict_ratings=[5, 5, 4, 5, 5, 5, 4, 5, 5, 5]

Table 29: Examples of $\mathcal{S}_4$ and observartions in $\mathcal{W}_5$