Personalize Your LLM: Fake it then Align it

Yijing Zhang [1], Dyah Adila[1], Changho Shin[1], Frederic Sala[1]

[1] University of Wisconsin - Madison
{yzhang2637, adila, cshin23, fredsala} @wisc.edu

Abstract

Personalizing large language models (LLMs) is essential for delivering tailored interactions that improve user experience. Many existing personalization methods require fine-tuning LLMs for each user, rendering them prohibitively expensive for widespread adoption. Although retrieval-based approaches offer a more compute-efficient alternative, they still depend on large, high-quality datasets that are not consistently available for all users. To address this challenge, we propose **CHAMELEON**, a scalable and efficient personalization approach that uses (1) self-generated personal preference data and (2) representation editing to enable quick and cost-effective personalization. Our experiments on various tasks, including those from the LaMP personalization benchmark, show that CHAMELEON efficiently adapts models to personal preferences, improving instruction-tuned models and outperforms two personalization baselines by an average of 40% across two model architectures.

1 Introduction

Large language models (LLMs) have transformed natural language processing (NLP), achieving excellent performance across a wide range of tasks. Their use has already expanded into diverse domains and user bases (Gururangan et al., 2020; Shi et al., 2024; Xu et al., 2024a,b). This has motivated the need for personalization, i.e. tailoring these models to individual user preferences and specific contexts (Kirk et al., 2023).

Current personalization methods are often impractical for large-scale deployment. Fine-tuning approaches (Li et al., 2024b; Tan et al., 2024; Clarke et al., 2024) are resource-intensive, making it prohibitively expensive to customize models for each individual user. In contrast, retrieval-based methods (Salemi et al., 2024; Di Palma, 2023; Fan et al., 2024) offer greater computational efficiency but suffer from a significant drawback: they rely on large high-quality datasets that are not consistently available for all users. These limitations impede the effective scaling of personalization, especially given the diverse and rapidly evolving nature of user preferences.

To achieve scalable personalization, we argue that two essential conditions must be met: (1) data efficiency, which enables effective personalization with minimal user interaction, and (2) compute efficiency, allowing for deployment across a large user base. We propose CHAMELEON, a new approach that fulfills both requirements by using synthetic, self-generated data to capture user preferences and uses representation editing to tailor its behavior to each user's unique preferences (Adila et al., 2024).

For each user, we begin with a small amount of historical data—sometimes as little as a single sample. Using this data, we prompt the LLM to generate two characteristic descriptions: one that reflects the user's personal preferences based on their history and another that represents a contrasting or non-personalized profile (e.g., "funny" versus "formal"). From these descriptions, we create synthetic user preference data. We then identify two distinct embedding spaces—personalized and non-personalized—derived from the synthetic preference data. Finally, we edit the LLM's embeddings to enhance the influence of the personalized subspace while diminishing the influence of the non-personalized subspace.

With this data- and compute-efficient approach, we improve instruction-tuned models and two LLM personalization baselines by an average of 40% in the LaMP personalization benchmark (Salemi et al., 2024). In summary, our contributions are:

1. We introduce CHAMELEON, an LLM personalization framework that leverages self-generated user preference data and embedding editing techniques, providing **scalable, user-tailored personalization that is nearly**

cost-free.

2. On extensive evaluation using the LaMP benchmark (Salemi et al., 2024), we show that CHAMELEON improves upon instruction-tuned models and two LLM personalization benchmarks by an average of 40% on two model architectures.

3. CHAMELEON can effectively personalize for new, unseen users without user history by leveraging profiles from other users with similar characteristics and preferences.

2 Related Work

Our work seeks to address the personalization problem for LLMs using representation editing as an efficient technique to align models with user preferences. We give a brief overview of related areas.

Personalized LLMs. Unlike general LLMs that produce uniform responses for all users, personalized LLMs adapt to the specific linguistic and communication preferences of individual users (Clarke et al., 2024). Fine-tuning is a common method for achieving this, by training models on user-specific or task-specific data to personalize their behavior (Woźniak et al., 2024). Approaches like P-RLHF (Li et al., 2024b), Persona-Plug (Liu et al., 2024a), and ALOE (Wu et al., 2024) exemplify this strategy. However, fine-tuning is resource-intensive, making it impractical to personalize models for individual users at scale. Parameter-efficient fine-tuning (PEFT) (Tan et al., 2024) reduces the computational burden but still requires large amounts of user data, which is often scarce and difficult to obtain in user personalization task (Zollo et al., 2024).

Retrieval-based methods personalize model outputs by incorporating user-specific information retrieved at inference time (Dai et al., 2023; Kang et al., 2023; Liu et al., 2023; Wang et al., 2023; Zhiyuli et al., 2023; Salemi et al., 2024). While these methods avoid the need for tuning, they struggle with LLMs' limited context lengths, especially when dealing with long user histories. Although long-context models (Dubey et al., 2024; Reid et al., 2024; Liu et al., 2024b) allow for processing larger user histories, this incurs a high cost as many models are charged per token. Attempts to address this issue by summarizing retrieved information have been made (Richardson et al., 2023; Liu et al., 2024c). However, these approaches are vulnerable to distractions from irrelevant information (Shi et al., 2023), particularly when user behavior or preferences shift (Carroll et al., 2024; Franklin et al., 2022).

The closest work to ours is LLM-REC (Lyu et al., 2024), a prompt-based approach that personalizes LLMs using summaries of selected top user history data. Our method takes this a step further by generating self-preference data, identifying embedding spaces that capture personalized versus non-personalized preferences, and performing personalization through representation editing. This enables a more data- and compute-efficient personalization process, making it possible to adapt models at scale to evolving user preferences quickly. Our approach represents a significant step toward scalable, real-time personalization that caters to dynamic user preference data.

Representation Editing for Personalization. Representation editing has become an important technique for model alignment, involving the direct manipulation of a model's latent representations to improve its performance and align it with desired attributes (Wang et al., 2024a; Kong et al., 2024). For example, Han et al. (2024) demonstrated that steering LLM text embeddings can guide model output *styles*. Similarly, (Li et al., 2024a; Han et al., 2023a) show that adjusting embeddings during inference can enhance specific attributes, such as honesty or truthfulness, in the generated outputs. Liang et al. (2024) found that representation editing can control aspects of text generation, such as safety, sentiment, thematic consistency, and *linguistic style*. These findings highlight the potential of using representation editing to guide models for personalization tasks. For visual generation models like Stable Diffusion, embedding-based personalization has long been recognized as an established technique (Han et al., 2023b; Arar et al., 2024; Alaluf et al., 2023; Yang et al., 2024).

Despite the growing interest in representation editing, little research has explored its application for personalizing LLMs, as proposed in our work. The most closely related study is Adila et al. (2024), where the authors use embedding editing for general, rather than personalized, alignment to broad human preferences, relying on self-generated synthetic data. Our approach advances this notion by introducing a tailored mechanism that generates personalized synthetic data for each user and adapts embedding editing techniques for both individual and group-based personalization.

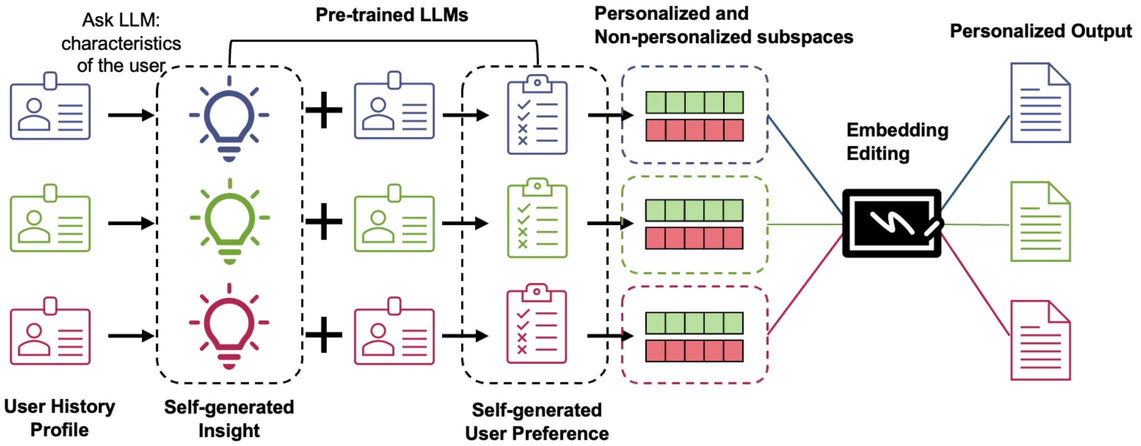

Figure 1: CHAMELEON identifies two separate subspaces, one personalized and one non-personalized, from self-generated user characteristic insights. Based on these subspaces, we modify the LLM embeddings during inference.

3 CHAMELEON: Personalization through Representation Editing

We present CHAMELEON, an almost **cost-free** alignment personalization framework with representation editing using self-generated synthetic user preference data. Figure 1 illustrates our technique. We achieve personalization with two stages: (1) self-generating user preference data (Section 3.1), and (2) representation editing using the self-generated data (Section 3.2). Additionally, we extend CHAMELEON to support **scalable user groups**, enabling efficient alignment at a group level (Section 3.3).

3.1 Self-generated Preference Data

Our method for generating self-preference data uses generic, non-personalized LLMs to identify user-specific characteristics and preferences from the available user history. Using these identified characteristics, we prompt the model to generate tailored responses for each user. This process consists of three key steps: (1) selecting relevant user history, (2) generating insights from the selected history, and (3) producing synthetic user preference data guided by these insights.

User History Selection. User's historical behavior usually contains important information regarding their characteristics, linguistic patterns, and preferred interactions. However, not all historical behaviors serve as reliable indicators of user preferences. Adapting the model using redundant and generic user behavior may not result in high-quality personalized LLMs. Selecting and filtering for representative user historical behavior is thus important. Although recent studies showed success in using retrieval-based re-rankers (Zhuang et al., 2024) and encoder-based user history selection (Liu et al., 2024a), they can struggle when user preferences shift rapidly or when there's limited historical data. To address this, we focus on a more lightweight and adaptable approach to user history selection.

Since our approach relies on embedding editing to adapt the model, we need to identify user-representative historical data. The first step is to define what makes this data "representative." We leverage sentence embeddings for their strong ability to capture both the meaning and context of text (Reimers and Gurevych, 2019). Our goal is to find the most informative and relevant embedding pieces that reflect key user preferences. A lightweight approach to find such data is to perform principal component analysis (PCA) on the embeddings (Gewers et al., 2021). Specifically, for each user u, given a set of user history $\mathcal{H}_u = \{h_u^i\}$ where each h_u^i represents an individual user history sample with index i, we have

$$e_u^i = \textbf{SentenceEmbedder}(h_u^i). \qquad (1)$$

Then, we have that $\boldsymbol{W}_u$ are the top k principal components of $\boldsymbol{E_u} = [e_u^1, e_u^2, \ldots, e_u^N]^\top$ and the projection of each embedding is $z_u^i = e_u^i \boldsymbol{W}_u$. We next find the top k history data embeddings:

$$\boldsymbol{E}_u^k = \underset{i\in[1,\ldots,N]}{\arg\text{top-}k} \left\| z_u^i \right\|, \qquad (2)$$

and get top k history data $\boldsymbol{H}_u^k = \{h_u^i : i \in \boldsymbol{E}_u^k\}$.

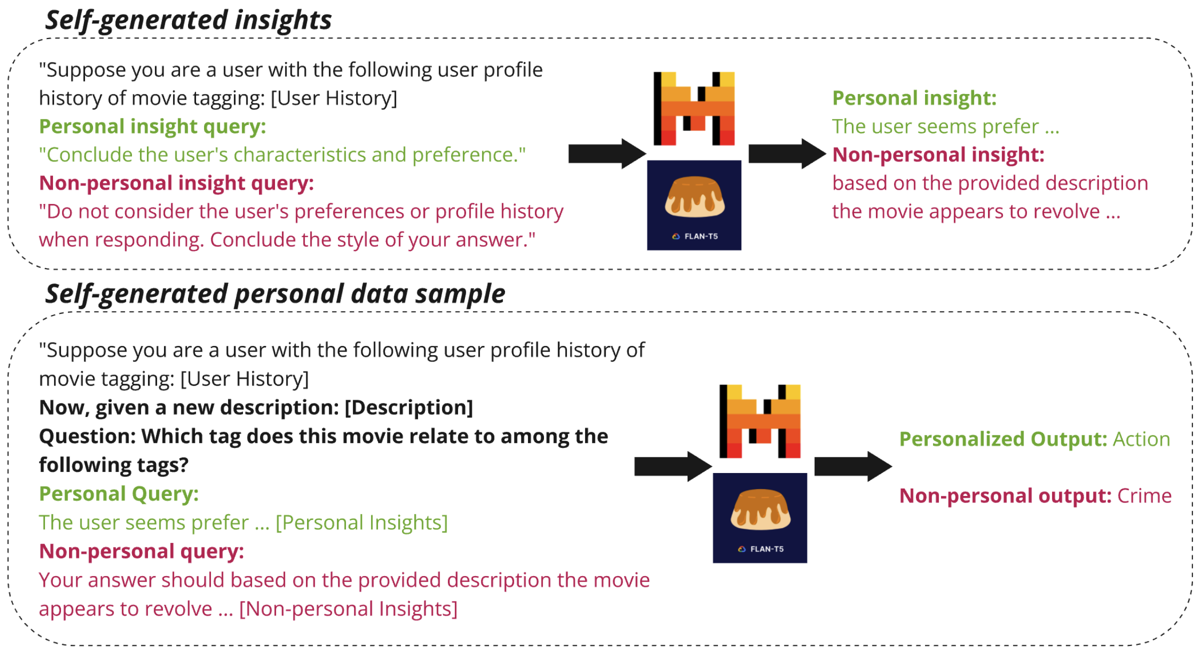

Figure 2: Self-generated user preference data: we use the generated conclusion of user characteristics to guide the personal answer generation.

Insight Generation. We query an instruction-tuned general-purpose LM to analyze and infer characteristics specific to individual users. For each user u, given the selected set of user history $\boldsymbol{H}_u^k$ from the previous step, we query the LM (denoted as ω) and generate two distinct styles of responses: one as a personalized agent (C^P) and the other as a non-personalized/neutral agent (C^N). The personalized agent (C^P) draws on the user's historical data $\boldsymbol{H}_u^k$, concluding insights about the user's preferences, behaviors, and style. The neutral agent (C^N) is asked to give characteristics of impersonal and general responses. It represents the standard behavior of the model when user personalization is absent. Then, for each user u, we have an personalized-neutral insights pair (c_u^P, c_u^N).

Generating Synthetic User Preference Data Once the insights are generated, we use the insight pairs as prompt guidance to generate synthetic user preference data. For each user u and each user query q_u, given the pre-selected history set $\mathcal{H}_u$ and insight pair $(c_u^{i,P}, c_u^{i,N})$, we have our general-purpose LM (ω) separately generate personalized and neutral preference outputs $(\hat{y}_u^{i,P}, \hat{y}_u^{i,N})$ to query q_u^i conditioned on $(c_u^{i,P}, c_u^{i,N})$ respectively. We then concatenate the outputs $(\hat{y}_u^{i,P}, \hat{y}_u^{i,N})$ with user history $\mathcal{H}_u$ and obtain the self-generated preference pair $(p_u^{i,P}, p_u^{i,N})$ for each user query q_u^i. By applying this procedure to all user queries, we obtain self-generated preference data pairs (P_u^P, P_u^N).

Note that we do not apply any prompt tuning; rather, we use a predefined set of prompt templates and a frozen LLM for all generations. Figure 2 illustrates the full process, with prompting details in Appendix A.3.

3.2 Representation Editing

Next, using the self-generated user preference data, we align the model with users' preferences with a technique inspired by ALIGNEZ (Adila et al., 2024). We first identify two subspaces in the model's embedding space (denoted as vector $\theta \in \mathbb{R}^d$ in LM ω's latent space) that correspond with the users' preferences. We use singular value decomposition (SVD) on the preference data embeddings to capture directions of the personalized embeddings $\theta_{l,u}^P$. Next, we employ CCS-based identification (Burns et al., 2023) to find the hyperplane that best separates the non-personalized embeddings from the personalized ones and denote the directions of the hyperplane as $\theta_{l,u}^N$. A detailed explanation is provided in Appendix A.4.

With the personalized and non-personalized subspaces θ^P and θ^N, we perform embedding editing on the MLP outputs of the most impactful decoder layers (i.e. layers that have lowest average CSS loss) during the inference phase to adapt the LLM to users' preferences. More concretely, given x_l, the output of the MLP of layer $l \in L$, where L is the set of layers with lowest average CSS loss, we strengthen the personalized direction by

$$\hat{x}_{l,u} \leftarrow x_l + \frac{\langle x_l, \theta^P_{l,u} \rangle}{\langle \theta^P_{l,u}, \theta^P_{l,u} \rangle} \theta^P_{l,u}$$

and remove the non-personalized direction by

$$\hat{x}_{l,u} \leftarrow \hat{x}_{l,u} - \frac{\langle \hat{x}_{l,u}, \theta^N_{l,u} \rangle}{\langle \theta^N_{l,u}, \theta^N_{l,u} \rangle} \theta^N_{l,u}.$$

These edits are performed for each user query.

3.3 Group-scale Personalization

Individually aligning the model for multiple users is inefficient when scaling to a large user base (Dai et al., 2024). To overcome this, we extend CHAMELEON to group-scale alignment. Instead of aligning for each user separately, we combine the history data of all users into a single group and perform collective alignment. Specifically, we aggregate the synthetic self-preference data for all users into one set, $(P^P, P^N) = \{(p^{i,P}_u, p^{i,N}_u) \in (P^P_u, P^N_u) | u \in U\}$, where U is the set of users in the group. (P^P, P^N) is then used to find direction vectors for representation editing.

This approach enables efficient personalization by processing all users simultaneously, leading to faster alignment. In Section 4.4, we show that group-scale personalization outperforms the single-user setting. Furthermore, this method allows us to leverage data from other users for those with no available history, enabling personalization for new or unseen users (see Experiment 4.2).

4 Experiments

We begin by detailing our experimental setup in Section 4.1, followed by experiments to validate the following key claims about CHAMELEON:

- Aligns LLMs to user-specific preferences (Section 4.2),
- Generalizes to unseen users (Section 4.3),
- Group-scale personalization improves performance (Section 4.4),
- Outperforms compute extensive methods like DPO in time-constrained scenarios (Section 4.5).

In Section 4.6, we perform ablation study to understand the effect of the number of user history data to CHAMELEON performance.

4.1 Experimental Setup

Datasets and Tasks. We evaluate CHAMELEON using the LaMP language model personalization benchmark (Salemi et al., 2024). Our evaluation focuses on three specific personalization tasks: (1) Personalized Movie Tagging (LaMP 2), (2) Personalized Product Rating (LaMP 3), and (3) Personalized Tweet Paraphrasing (LaMP 7). We adhered to the user-based data split provided by the LaMP benchmark, using the default training and test splits. Additional details about the datasets and tasks can be found in Appendix A.2.

Evaluation Metrics. We use the evaluation metrics established by the LaMP benchmark for each task. For Personalized Movie Tagging (LaMP 2), we measure Accuracy (Acc.) and F-1 Score (F-1). For Personalized Product Rating (LaMP 3), we assess performance using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). For Personalized Tweet Paraphrasing (LaMP 7), we apply the ROUGE-1 (R-1) and ROUGE-L (R-L) metrics.

Baseline 1: Non-personalized Instruction-tuned Models. We evaluate CHAMELEON against two general purpose instruction-tuned models: Mistral-7B-v0.3-Instruct (Jiang et al., 2023) and Flan-T5 XXL (Chung et al., 2022). Both models are assessed using the same set of user queries as CHAMELEON, following the same prompt format and using the same pre-selected user history profile—excluding any insights. Additional prompt details can be found in Appendix A.3.

Baseline 2: Personalization Methods. We also compare CHAMELEON against two personalization techniques, namely LLM-REC (Lyu et al., 2024), a prompting-engineering personalization method, and ALOE (Wu et al., 2024), a supervised Fine-tuning (SFT) personalization method.

Group Personalization Setup. To implement group-scale personalization (Section 3.3), we randomly select 100 users from the training split of the LaMP benchmark. Using PCA-based history selection (Section 3.1), we choose up to 10 user history entries per profile. For each user, we generate personalized and neutral insight pairs along with self-generated preference data. Any data where the personalized and non-personalized outputs are identical is discarded. We then combine the self-generated preference data for all users, perform

Models →		Mistral Instruct				Flan T5 XXL			
Dataset	Metric	Instruct Model	LLM -REC	ALOE	CHAMELEON	Instruct Model	LLM -REC	ALOE	CHAMELEON
LaMP2	Acc. ↑	0.198	0.262	0.307	**0.396**	0.238	0.214	0.333	**0.420**
	F-1 ↑	0.236	0.309	0.220	**0.349**	0.171	0.146	0.255	**0.311**
LaMP3	MAE ↓	0.497	0.484	0.423	**0.407**	0.456	0.798	0.427	**0.400**
	RMSE ↓	0.944	0.976	0.888	**0.815**	0.818	1.439	0.786	**0.714**
LaMP7	R-1 ↑	0.354	0.183	0.362	**0.381**	0.333	0.225	0.376	**0.429**
	R-L ↑	0.295	0.144	0.313	**0.334**	0.292	0.196	0.331	**0.385**

Table 1: CHAMELEON outperforms all baselines in personalization for users with history. Best performance is highlighted in **bold**. Metrics where higher values indicate better performance are shaded in blue cells, while metrics where lower values are preferable are marked with green cells.

Models →		Mistral Instruct		Flan T5 XXL	
Dataset	Metric	ALOE	CHAMELEON	ALOE	CHAMELEON
LaMP2	Acc. ↑	0.227	**0.363**	0.109	**0.390**
	F-1 ↑	0.177	**0.338**	0.040	**0.304**
LaMP3	MAE ↓	0.522	**0.442**	0.544	**0.413**
	RMSE ↓	0.906	**0.903**	1.030	**0.839**
LaMP7	R-1 ↑	0.185	**0.377**	0.251	**0.420**
	R-L ↑	0.155	**0.331**	0.206	**0.373**

Table 2: CHAMELEON performance compared ALOE on new unseen users.

group-scale alignment, and evaluate the personalized model on unseen user queries from the LaMP test split (Section 4.3). This process is repeated for different random sets of 100 users, and we report the average performance.

4.2 Aligns LLMs to user-specific preferences

Setup. We compare CHAMELEON with the previously mentioned baselines. In the self-insight generation process, user history data is fed directly to the models using simple prompts (see Appendix A.3), without access to human annotations.

Results. As shown in Table 1, CHAMELEON consistently outperforms all baselines. Remarkably, these improvements are achieved with minimal user history data and without any training and fine-tuning, surpassing an SFT-based method (ALOE). **These results validate our claim that CHAMELEON can effectively align LLMs to individual user preferences**.

4.3 Generalizes to unseen users

Setup. We also assess CHAMELEON's ability to personalize for new, unseen users who have no prior history. In this evaluation, we run both CHAMELEON and ALOE on the LaMP training split and evaluate their performance on test samples from users not included in the training data. This experimental setup is not applicable to instruct models and LLM-REC, as both of these methods use prompt-based personalization and do not differentiate between seen and unseen users.

Results. Table 2 demonstrates that CHAMELEON achieves strong personalization performance even with new, unseen users, **validating our claim that CHAMELEON can effectively generalize to users without prior history**. In contrast, ALOE struggles in this scenario, suggesting that it may overfit to the characteristics of users in the training set.

4.4 Group-scale personalization improves performance

Setup. To assess the effectiveness of group-scale personalization compared to single-user personalization, we run CHAMELEON on groups of varying sizes. We experiment with group sizes of $\{1, 20, 40, 60, 80, 100\}$ on both LaMP2 and LaMP3 tasks, while keeping the amount of generated insights and preference data per user constant.

Results. Figure 3 reveals a clear trend: as the number of users in the group increases, personalization performance consistently improves. This is evident both when shifting from a single-user setup

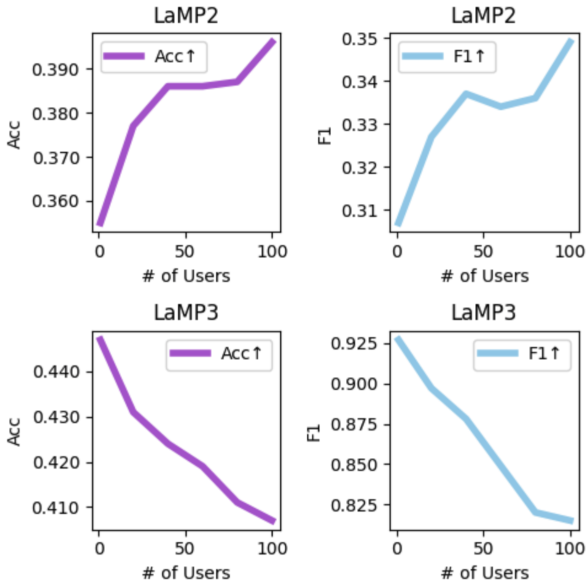

Figure 3: The change of performance when different number of users are given to CHAMELEON

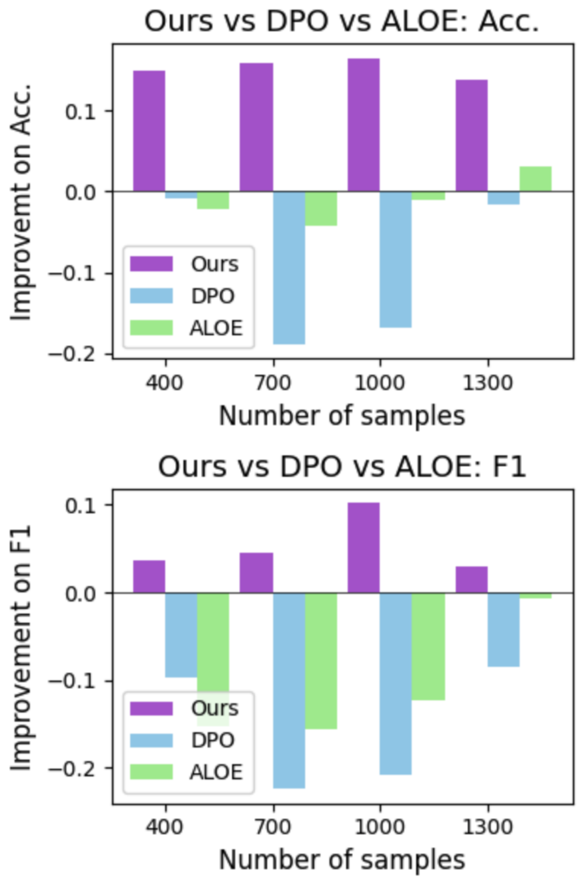

Figure 4: CHAMELEON compared with DPO and ALOE in time-constrained scenarios. The columns denotes the improvement from the instruction-tuned model.

(left-most point, where number of users = 1) to group personalization, and as the group size grows. **These results support our claim that group personalization offers performance gain compared to single-user personalization**.

4.5 Outperforms DPO in time-constrained scenario

Setup. We compare CHAMELEON with DPO (Rafailov et al., 2024) and ALOE (Wu et al., 2024), a tuning-based alignment and SFT-based personalization methods, in a time-constrained scenario where alignment must be performed quickly. In this setup, we fix the time allowed for all methods and get the number of samples for each method within that time. This setup reflects real-world situations where instant personalization is required for new users with little to no available data. Hyperparameter details for DPO and ALOE are provided in Appendix A.5.

Results As shown in Figure 4, CHAMELEON consistently delivers stable personalization gains in the time-constrained scenario, whereas both ALOE and DPO struggle with limited sample availability. This supports our claim that **CHAMELEON is more suitable than resource-intensive approaches in time-sensitive scenarios**.

4.6 Ablations

Setup. To examine the impact of the amount of user history data on performance, we run CHAMELEON on the LaMP2 task, varying the number of history per user as $\{5, 10, 15, 20, 25\}$, while keeping the number of users in the group constant.

Results. Figure 5 illustrates that when the amount of user history data is small, the performance improvement of CHAMELEON is limited. This limitation likely arises from the difficulty in generating accurate personalization insights with insufficient data. Conversely, when the amount of history data is too big, the performance of CHAMELEON declines. We hypothesize that this deterioration occurs because too many history profiles may introduce unrelated or outdated samples, hindering effective personalization.

5 Discussion

Limitations. While CHAMELEON successfully delivers scalable personalization with minimal costs, it has some limitations. A key challenge is its dependence on the quality of the self-generated preference data. Although aligning the model with

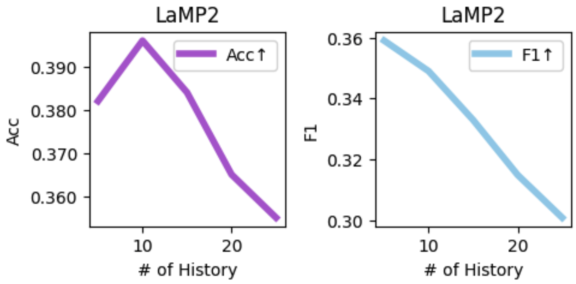

Figure 5: The change of performance when different number of history data per user are given to CHAMELEON

this data yields promising results, the effectiveness of the personalization largely depends on how accurately and comprehensively user preferences are captured by the base LLM. Future research could focus on developing more refined metrics to capture personal characteristics better, ensuring more precise and reliable self-alignment.

One potential risk with CHAMELEON is the possibility of malicious input in user history. Since CHAMELEON relies on a limited amount of user history to generate self-preference data for alignment, harmful or biased history inputs could unintentionally lead the model to produce toxic or malicious responses. This highlights the need for strong safeguards, such as thorough filtering and ethical review processes, to prevent the model from aligning with or reinforcing negative behaviors while still delivering effective personalization.

Ethical Considerations. Privacy has long been a problem for LLM personalization, as personalizing LLMs usually require large-scale personal data and preferredly (human) labeled, which could lead to potential privacy leaks. Though personalization dataset, like LaMP benchmark dataset used in our experiments, is publicly accessible an does not raise privacy concerns, personal data collection and usage still presents significant challenge in personalizing LLMs. With our approach, we only acquire a very small portion of user historical data and resolve data labeling problem with self-generation technique. And since self-generated user preference data are fake synthetic data for performing alignment, it can possibly reduce the risk of privacy leaks.

Conclusion. We present CHAMELEON, a novel light-weight, scalable approach for personalizing LLMs without access to large-scale human-annotated personal data and individual fine-tuning. By leveraging the ability to conclude and capture user characteristics and preferences, CHAMELEON adjusts the model embeddings during inference to generate outputs that are more aligned with user preferences. Our experiments show that CHAMELEON significantly enhance the personalization ability of base language models using only a small portion of real user data, and it is able to adapt models with multiple user expectations within one single alignment process.

This work represents an initial step toward achieving cost-free, rapid, group-scale personalization that current personalization methods struggle to address.

References

Dyah Adila, Changho Shin, Yijing Zhang, and Frederic Sala. 2024. Is free self-alignment possible? *Preprint*, arXiv:2406.03642.

Yuval Alaluf, Elad Richardson, Gal Metzer, and Daniel Cohen-Or. 2023. A neural space-time representation for text-to-image personalization. *ACM Trans. Graph.*, 42(6).

Moab Arar, Andrey Voynov, Amir Hertz, Omri Avrahami, Shlomi Fruchter, Yael Pritch, Daniel Cohen-Or, and Ariel Shamir. 2024. Palp: Prompt aligned personalization of text-to-image models. *Preprint*, arXiv:2401.06105.

Collin Burns, Haotian Ye, Dan Klein, and Jacob Steinhardt. 2023. Discovering latent knowledge in language models without supervision. In *The Eleventh International Conference on Learning Representations*.

Micah Carroll, Davis Foote, Anand Siththaranjan, Stuart Russell, and Anca Dragan. 2024. Ai alignment with changing and influenceable reward functions. *Preprint*, arXiv:2405.17713.

Hyung Won Chung, Le Hou, Shayne Longpre, Barret Zoph, Yi Tay, William Fedus, Yunxuan Li, Xuezhi Wang, Mostafa Dehghani, Siddhartha Brahma, Albert Webson, Shixiang Shane Gu, Zhuyun Dai, Mirac Suzgun, Xinyun Chen, Aakanksha Chowdhery, Alex Castro-Ros, Marie Pellat, Kevin Robinson, Dasha Valter, Sharan Narang, Gaurav Mishra, Adams Yu, Vincent Zhao, Yanping Huang, Andrew Dai, Hongkun Yu, Slav Petrov, Ed H. Chi, Jeff Dean, Jacob Devlin, Adam Roberts, Denny Zhou, Quoc V. Le, and Jason Wei. 2022. Scaling instruction-finetuned language models. *Preprint*, arXiv:2210.11416.

Christopher Clarke, Yuzhao Heng, Lingjia Tang, and Jason Mars. 2024. Peft-u: Parameter-efficient fine-tuning for user personalization. *Preprint*, arXiv:2407.18078.

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.

Zhenlong Dai, Chang Yao, WenKang Han, Ying Yuan, Zhipeng Gao, and Jingyuan Chen. 2024. Mpcoder: Multi-user personalized code generator with explicit and implicit style representation learning. *Preprint*, arXiv:2406.17255.

Dario Di Palma. 2023. Retrieval-augmented recommender system: Enhancing recommender systems with large language models. In *Proceedings of the 17th ACM Conference on Recommender Systems*, pages 1369–1373.

Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, et al. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

Wenqi Fan, Yujuan Ding, Liangbo Ning, Shijie Wang, Hengyun Li, Dawei Yin, Tat-Seng Chua, and Qing Li. 2024. A survey on rag meeting llms: Towards retrieval-augmented large language models. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 6491–6501.

Matija Franklin, Hal Ashton, Rebecca Gorman, and Stuart Armstrong. 2022. Recognising the importance of preference change: A call for a coordinated multidisciplinary research effort in the age of ai. *arXiv preprint arXiv:2203.10525*.

Felipe L. Gewers, Gustavo R. Ferreira, Henrique F. De Arruda, Filipi N. Silva, Cesar H. Comin, Diego R. Amancio, and Luciano Da F. Costa. 2021. Principal component analysis: A natural approach to data exploration. *ACM Computing Surveys*, 54(4):1–34.

Suchin Gururangan, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A. Smith. 2020. Don't stop pretraining: Adapt language models to domains and tasks. *Preprint*, arXiv:2004.10964.

Chi Han, Jialiang Xu, Manling Li, Yi Fung, Chenkai Sun, Nan Jiang, Tarek Abdelzaher, and Heng Ji. 2023a. Lm-switch: Lightweight language model conditioning in word embedding space. *arXiv preprint arXiv:2305.12798*.

Chi Han, Jialiang Xu, Manling Li, Yi Fung, Chenkai Sun, Nan Jiang, Tarek Abdelzaher, and Heng Ji. 2024. Word embeddings are steers for language models. *Preprint*, arXiv:2305.12798.

Inhwa Han, Serin Yang, Taesung Kwon, and Jong Chul Ye. 2023b. Highly personalized text embedding for image manipulation by stable diffusion. *arXiv preprint arXiv:2303.08767*.

Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7b. *Preprint*, arXiv:2310.06825.

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*.

Hannah Rose Kirk, Bertie Vidgen, Paul Röttger, and Scott A Hale. 2023. Personalisation within bounds: A risk taxonomy and policy framework for the alignment of large language models with personalised feedback. *arXiv preprint arXiv:2303.05453*.

Lingkai Kong, Haorui Wang, Wenhao Mu, Yuanqi Du, Yuchen Zhuang, Yifei Zhou, Yue Song, Rongzhi Zhang, Kai Wang, and Chao Zhang. 2024. Aligning large language models with representation editing: A control perspective. *Preprint*, arXiv:2406.05954.

Kenneth Li, Oam Patel, Fernanda Viégas, Hanspeter Pfister, and Martin Wattenberg. 2024a. Inference-time intervention: Eliciting truthful answers from a language model. *Advances in Neural Information Processing Systems*, 36.

Xinyu Li, Zachary C. Lipton, and Liu Leqi. 2024b. Personalized language modeling from personalized human feedback. *Preprint*, arXiv:2402.05133.

Xun Liang, Hanyu Wang, Yezhaohui Wang, Shichao Song, Jiawei Yang, Simin Niu, Jie Hu, Dan Liu, Shunyu Yao, Feiyu Xiong, and Zhiyu Li. 2024. Controllable text generation for large language models: A survey. *Preprint*, arXiv:2408.12599.

Jiongnan Liu, Yutao Zhu, Shuting Wang, Xiaochi Wei, Erxue Min, Yu Lu, Shuaiqiang Wang, Dawei Yin, and Zhicheng Dou. 2024a. Llms + persona-plug = personalized llms. *Preprint*, arXiv:2409.11901.

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*.

Nelson F Liu, Kevin Lin, John Hewitt, Ashwin Paranjape, Michele Bevilacqua, Fabio Petroni, and Percy Liang. 2024b. Lost in the middle: How language models use long contexts. *Transactions of the Association for Computational Linguistics*, 12:157–173.

Qijiong Liu, Nuo Chen, Tetsuya Sakai, and Xiao-Ming Wu. 2024c. 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, Chris Leung, Jiajie Tang, and Jiebo Luo. 2024. LLM-rec: Personalized recommendation via prompting large language models. In *Findings of the Association for Computational Linguistics: NAACL 2024*, pages 583–612, Mexico City, Mexico. Association for Computational Linguistics.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D. Manning, and Chelsea Finn. 2024. Direct preference optimization: Your language model is secretly a reward model. *Preprint*, arXiv:2305.18290.

Machel Reid, Nikolay Savinov, Denis Teplyashin, Dmitry Lepikhin, Timothy Lillicrap, Jean-baptiste Alayrac, Radu Soricut, Angeliki Lazaridou, Orhan Firat, Julian Schrittwieser, et al. 2024. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. *arXiv preprint arXiv:2403.05530*.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *Preprint*, arXiv:1908.10084.

Chris Richardson, Yao Zhang, Kellen Gillespie, Sudipta Kar, Arshdeep Singh, Zeynab Raeesy, Omar Zia Khan, and Abhinav Sethy. 2023. Integrating summarization and retrieval for enhanced personalization via large language models. *Preprint*, arXiv:2310.20081.

Alireza Salemi, Sheshera Mysore, Michael Bendersky, and Hamed Zamani. 2024. LaMP: When large language models meet personalization. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7370–7392, Bangkok, Thailand. Association for Computational Linguistics.

Freda Shi, Xinyun Chen, Kanishka Misra, Nathan Scales, David Dohan, Ed H Chi, Nathanael Schärli, and Denny Zhou. 2023. Large language models can be easily distracted by irrelevant context. In *International Conference on Machine Learning*, pages 31210–31227. PMLR.

Wenqi Shi, Ran Xu, Yuchen Zhuang, Yue Yu, Hang Wu, Carl Yang, and May D Wang. 2024. Medadapter: Efficient test-time adaptation of large language models towards medical reasoning. *arXiv preprint arXiv:2405.03000*.

Zhaoxuan Tan, Qingkai Zeng, Yijun Tian, Zheyuan Liu, Bing Yin, and Meng Jiang. 2024. Democratizing large language models via personalized parameter-efficient fine-tuning. *Preprint*, arXiv:2402.04401.

Danqing Wang, Kevin Yang, Hanlin Zhu, Xiaomeng Yang, Andrew Cohen, Lei Li, and Yuandong Tian. 2023. Learning personalized story evaluation. *arXiv preprint arXiv:2310.03304*.

Liang Wang, Nan Yang, Xiaolong Huang, Linjun Yang, Rangan Majumder, and Furu Wei. 2024a. Improving text embeddings with large language models. *Preprint*, arXiv:2401.00368.

Xingyao Wang, Yangyi Chen, Lifan Yuan, Yizhe Zhang, Yunzhu Li, Hao Peng, and Heng Ji. 2024b. Executable code actions elicit better llm agents. *Preprint*, arXiv:2402.01030.

Stanisław Woźniak, Bartłomiej Koptyra, Arkadiusz Janz, Przemysław Kazienko, and Jan Kocoń. 2024. Personalized large language models. *Preprint*, arXiv:2402.09269.

Shujin Wu, May Fung, Cheng Qian, Jeonghwan Kim, Dilek Hakkani-Tur, and Heng Ji. 2024. Aligning llms with individual preferences via interaction. *arXiv preprint arXiv:2410.03642*.

Ran Xu, Hejie Cui, Yue Yu, Xuan Kan, Wenqi Shi, Yuchen Zhuang, May Dongmei Wang, Wei Jin, Joyce Ho, and Carl Yang. 2024a. Knowledge-infused prompting: Assessing and advancing clinical text data generation with large language models. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 15496–15523.

Ran Xu, Wenqi Shi, Yue Yu, Yuchen Zhuang, Bowen Jin, May D Wang, Joyce C Ho, and Carl Yang. 2024b. Ram-ehr: Retrieval augmentation meets clinical predictions on electronic health records. *arXiv preprint arXiv:2403.00815*.

Yitong Yang, Yinglin Wang, Jing Wang, and Tian Zhang. 2024. Prompt-softbox-prompt: A free-text embedding control for image editing. *Preprint*, arXiv:2408.13623.

Aakas Zhiyuli, Yanfang Chen, Xuan Zhang, and Xun Liang. 2023. Bookgpt: A general framework for book recommendation empowered by large language model. *arXiv preprint arXiv:2305.15673*.

Yuchen Zhuang, Haotian Sun, Yue Yu, Rushi Qiang, Qifan Wang, Chao Zhang, and Bo Dai. 2024. Hydra: Model factorization framework for black-box llm personalization. *Preprint*, arXiv:2406.02888.

Thomas P Zollo, Andrew Wei Tung Siah, Naimeng Ye, Ang Li, and Hongseok Namkoong. 2024. Personalllm: Tailoring llms to individual preferences. *arXiv preprint arXiv:2409.20296*.

A Appendix

A.1 Glossary

Table 3 shows glossary of terms used in this paper.

A.2 Dataset and Task Details

The LaMP dataset is a publicly available dataset for personalizing LLMs. We only used LaMP dataset for the purpose of running the experiments.

The tasks of LaMP we experimented with are as follows:

Symbol	Definition
y	Ground truth output
$\hat{y}$	Model prediction
$\mathcal{H}_u$	Set of user history for user u
h_u^i	i-th user history for user h (i-th data data in $\mathcal{H}_u$)
e_u^i	Sentence embedding of h_u^i
$\boldsymbol{E}_u$	Embedding matrix of user history for user u
$\boldsymbol{H}_u^k$	Top k selected history data
C^P	Personalized agent
C^N	Non-personalized agent
c_u^P	Personalized insights for user u
c_u^N	Non-personalized insights for user u
$c_u^{i,P}$	i-th personalized insight for user u
$c_u^{i,N}$	i-th non-personalized insight for user u
$\hat{y}_u^{i,P}$	Model prediction conditioned on $c_u^{i,P}$
$\hat{y}_u^{i,N}$	Model prediction conditioned on $c_u^{i,N}$
q_u^i	i-th query for user u
$p_u^{i,P}$	Personalized preference for user query q_u^i
$p_u^{i,N}$	Non-personalized preference for user query q_u^i
P_u^P	Set of personalized preferences for user u
P_u^N	Set of non-personalized preferences for user u
θ^P	Personalized embedding direction
θ^N	Non-personalized embedding direction
$\theta_{l,u}^P$	Personalized embedding direction for user u at layer l
$\theta_{l,u}^N$	Non-personalized embedding direction for user u at layer l
x_l	Representation (embedding) at layer l
$\hat{x}_{l,u}$	Personalized representation for user u at layer l

Table 3: Glossary of variables and symbols used in this paper.

1. **LaMP 2: Personalized Movie Tagging.** Given a user profile of user history tagging along with the movie description, you are tasked to predict the movie tag given a new movie description.

2. **LaMP 3: Personalized Product Rating.** Given a user profile of user history product rating along with the product reviews, you are tasked to predict the rating of a product given a new product review wrote by the user.

3. **LaMP 7: Personalized Tweet Paraphrasing.** Given a user profile of user history tweets you are tasked to predict how the user may paraphrase a new given tweet.

Details of LaMP dataset is presented in Table 4. [*italic text*] presents actual data.

A.3 Prompt Template

Following is the history and prompt template used to query the base LM to generate preference samples for different LaMP task. History prompt format follows the format used by LaMP benchmark (Salemi et al., 2024).

LaMP 2: Personalized Movie Tagging

Personalize prompt: Suppose you are a user with the following user profile history of movie tagging: [HISTORY]

Now, given a new description: [QUERY]

Question: Which tag does this movie relate to among the following tags? Just answer with only ONE tag name without further explanation. tags: [sci-fi, based on a book, comedy, action, twist ending, dystopia, dark comedy, classic, psychology, fantasy, romance, thought-provoking, social commentary, violence, true story]

You are a helpfully personalized assistant. You try to predict the movie tagging that the user preferred based on their history. The user prefers [INSIGHT]. Answer only with one tag name (sci-fi, based on a book, comedy, action, twist ending, dystopia, dark comedy, classic, psychology, fantasy, romance, thought-provoking, social commentary, violence, true story).

Your answer: [OUTPUT]

Non-personalize/Neutral prompt: Suppose you are a user with the following user profile history of movie tagging: [HISTORY]

Now, given a new description: [QUERY]

Question: Which tag does this movie relate to among the following tags? Just answer with only ONE tag name without further explanation. tags: [sci-fi, based on a book, comedy, action, twist ending, dystopia, dark comedy, classic, psychology, fantasy, romance, thought-provoking, social commentary, violence, true story]

You are a generic and impersonal assistant. You do not consider the user's preferences or profile history when responding. Your answer shoulds [INSIGHT]. Answer only with one tag name (sci-fi, based on a book, comedy, action, twist ending, dystopia, dark comedy, classic, psychology, fantasy, romance, thought-provoking, social commentary, violence, true story).

Your answer: [OUTPUT]

History format:

1. The tag for movie: "[DESCRIPTION 1]" is "[TAG 1]".
2. The tag for movie: "[DESCRIPTION 2]" is "[TAG 2]".
3. ...

LaMP 3: Personalized Product Rating

Personalize prompt: Suppose you are a user with the following user profile history of product rating based on the user's review of the product: [HISTORY]

Now, given a new review by the user: [QUERY]

Question: What is the rating score of the following review on a scale of 1 to 5? Just answer with 1, 2, 3, 4, or 5 without further explanation.

You are a helpfully personalized assistant. You try to predict the rating of the product based on the user history ratings. The user prefers [INSIGHT]. Just answer with 1, 2, 3, 4, or 5 without further explanation.

Your answer: [OUTPUT]

Non-personalize/Neutral prompt: Suppose you are a user with the following user profile history of product rating based on the user's review of the product: [HISTORY]

Now, given a new review by the user: [QUERY]

Question: What is the rating score of the following review on a scale of 1 to 5? Just answer with 1, 2, 3, 4, or 5 without further explanation.

You are a generic and impersonal assistant. You do not consider the user's preferences or profile history when responding. Your answer should [INSIGHT].

Your answer: [OUTPUT]

History format:

1. [SCORE 1] is the rating score for product: "[TEXT 1]".

Table 4: LaMP Dataset Detail

Task	Input	Output
LaMP 2	ID: [*id*] Input: Which tag does this movie relate to among the following tags? Just answer with the tag name without further explanation. tags: [sci-fi, based on a book, comedy, action, twist ending, dystopia, dark comedy, classic, psychology, fantasy, romance, thought-provoking, social commentary, violence, true story] description: [description] Profile: {id: [*id*], tag: [*tag*], description: [*description*] }, . . .	[*tag*]
LaMP 3	ID: [*id*] Input What is the score of the following review on a scale of 1 to 5? just answer with 1, 2, 3, 4, or 5 without further explanation. review: [review], Profile {id: [*id*], tag: [*text*], description: [*score*] }, . . .	[*score*]
LaMP 7	ID: [*id*] Input: Paraphrase the following tweet without any explanation before or after it: [*tweet*] Profile: {id: [*id*], tag: [*text*]}, . . .	[*tweet*]

2. [SCORE 2] is the rating score for product: "[TEXT 2]".
3. ...

LaMP 7: Personalized Tweet Paraphrasing

Personalize prompt: Suppose you are a twitter user with the following user profile history that shows their preferred way of speaking: [HISTORY]

Now, given a new twitter post: [QUERY]

Question: Paraphrase the tweet in the style the user likes without any explanation before or after it.

You are a helpfully personalized assistant. You try to paraphrase the tweet in the style the user likes based on the history. The user prefers [INSIGHT].

Your answer: [OUTPUT]

Non-personalize/Neutral prompt: Suppose you are a twitter user with the following user profile history that shows their preferred way of speaking: [HISTORY]

Now, given a new twitter post: [QUERY]

Question: Paraphrase the tweet in the style the user likes without any explanation before or after it.

You are a generic and impersonal assistant. You do not consider the user's preferences or profile history when responding. Your answer should [INSIGHT].

Your answer: [OUTPUT]

History format:

1. [TWEET 1]
2. [TWEET 2]
3. ...

A.4 Details on Representation Editing

We provide the details of Section 3.2. We identify personalized and non-personized directions using singular value decomposition (SVD) or contrast consistent search (CCS) on the preference data embeddings. Let Φ_l represent the function that maps an input sentence to the LM embedding space at layer l. For each pair $(p_u^{i,P}, p_u^{i,N})$, we obtain their corresponding representations $\Phi_{l,u}^{i,P}$ and $\Phi_{l,u}^{i,P}$, respectively. To begin, we construct an embedding matrix for personalized direction for user u, denoted as $\mathbf{H}_{l,u}^{P}$, using these representations:

$$\mathbf{H}_{l,u}^{P} := \left[\Phi_{l,u}^{1,P}\middle|\ldots\middle|\Phi_{l,u}^{K,P}\right]^{T},$$

where K is the total number of self-generated data. Similarly, we create the non-personalized preferences embedding matrix $\mathbf{H}_{l,u}^{N}$.

SVD-Based Identification. Our approach for identifying personalized embedding directions involves using singular value decomposition (SVD) on the preference data embeddings. We extract the top right singular vector of $\mathbf{H}_{l,u}^{P}$ as $\theta_{l,u}^{P}$. Intuitively, we view θ as the direction that best captures the underlying personalized characteristics. We identify the personalized embedding direction for user u as follows:

$$\mathbf{H}^{P}_{l,u} = \mathbf{U}\Sigma\mathbf{V}$$
$$\theta^{P}_{l,u} := \mathbf{V}_{0,*}. \quad (3)$$

Here, $\mathbf{U}$ and $\mathbf{V}$ represent the left and right unitary matrices produced by running SVD, respectively, and Σ is the diagonal matrix of singular values. We define $\theta^{P}_{l,u}$ as the first row of $\mathbf{V}$, corresponding to the top right singular vector of $\mathbf{H}^{P}_{l,u}$. The non-personalized direction $\theta^{N}_{l,u}$ is defined similarly.

CCS-Based Identification (Burns et al., 2023). Another approach for identifying these directions is by finding a hyperplane in the latent space that separates personalized data embeddings from non-personalized ones. Typically, this is achieved by training lightweight probes $\theta_{l,u}$ that maps $\Phi^{P}_{l,u}$ and $\Phi^{N}_{l,u}$ to their respective classification labels (Li et al., 2024a). However, we face the challenge of avoiding overfitting to the noise inherent in self-generated data, which limits the applicability of supervised classifier loss in our context. To mitigate this issue, we employ the unsupervised Contrast-Consistent Search (CCS) loss $\mathcal{L}_{CCS}$ proposed in (Burns et al., 2023). Adapting the definition from (Burns et al., 2023) to our notations, $\mathcal{L}_{CCS}$ for each user u can be expressed as:

$$\mathcal{L}_{consistency}(g_{\theta,b}, \Phi^{i,P}_{l,u}, \Phi^{i,N}_{l,u})))$$
$$:= \left[g_{\theta,b}(\Phi^{N}_{l,u}) - (1 - g_{\theta,b}(\Phi^{P}_{l,u}))\right]^2$$
$$\mathcal{L}_{confidence}(g_{\theta,b}, \Phi^{i,P}_{l,u}, \Phi^{i,N}_{l,u})))$$
$$:= \min\left\{g_{\theta,b}(\Phi^{N}_{l,u}), g_{\theta,b}(\Phi^{P}_{i,u})\right\}$$
$$\mathcal{L}_{CCS}(g_{\theta,b}) := \frac{1}{K}\sum_{i=1}^{K}(\mathcal{L}_{consistency}(g_{\theta,b}, \Phi^{i,P}_{l,u}, \Phi^{i,N}_{l,u})$$
$$+ \mathcal{L}_{confidence}(g_{\theta,b}, \Phi^{i,P}_{l,u}, \Phi^{i,N}_{l,u})),$$

where $g_{\theta,b}(v) = \frac{1}{1+e^{-(\theta^\top v+b)}}$. Training $\theta^{N}_{l,u}$ with the L_{CCS} objective aims to find a separating hyperplane without fitting any labels with $\mathcal{L}_{consistency}$ and concurrently promoting maximum separation with $\mathcal{L}_{confidence}$.

Hybrid Identification. While both SVD-based or CCS-based identification can be used to identify both of personalized and non-personalized directions, our exploration revealed that the best results are achieved by combining the two approaches. Specifically, we use SVD to identify $\theta^{P}_{l,u}$ and CCS to determine $\theta^{N}_{l,u}$. This combined approach leverages the strengths of both techniques: SVD's ability to capture the primary direction of personalized embeddings and CCS's effectiveness in identifying the hyperplane that best separates non-personalized embeddings from personalized ones.

A.5 Time-constrained experiment Set Up

CHAMELEON The approximation for the time taken for our experiment is 10, 20, 30 and 40 minutes.

DPO DPO experiment is trained on 40%, 60%, 80%, 100% of the LaMP2 partition to get the approximate same time. The hyperparameters we used consist of 1 training epoch, a batch size of 16, a gradient accumulation step of 1, a learning rate of 5e-5, a max grad norm of 0.3, a warmup ratio of 0.1, a precision of bfloat16, a memory saving quantize flag of "bnb.nf4", a learning rate scheduler type of cosine, and an optimizer of AdamW with PEFT configurations of a r of 256, a α of 128, a dropout of 0.05 and a task type of causal language modeling"

ALOE We trained ALOE with 7%, 23%, 39%, 55% of the LaMP2 training partition with a relatively equal percentage of CodeAct data (Wang et al., 2024b) as described by ALOE (Wu et al., 2024). We used parameters provided in their SFT hyperparameters, which contains 1 training epoch, a per device train batch size of 1, a gradient accumulation step of 48, a learning rate of 1e-5, and a max sequence length of 8192.

A.6 Computing Resources

All experiments are trained on an Amazon EC2 Instances with eight NVIDIA A100-SXM4-40GB.