# End-to-end Training for Recommendation with Language-based User Profiles

Zhaolin Gao, Joyce Zhou, Yijia Dai and Thorsten Joachims

*Cornell University, Ithaca, NY, USA*

### Abstract

There has been growing interest in natural language-based user profiles, as they promise enhanced transparency and scrutability of recommender systems. While existing work has shown that language-based profiles from standard LLMs can be effective, such generalist LLMs are unlikely to be optimal for this task. In this paper, we introduce LangPTune, the first end-to-end learning method for training LLMs to produce language-based user profiles that optimize recommendation effectiveness. We evaluate LangPTune with various training configurations against a variety of baseline approaches, finding that our method significantly outperforms other profile-based methods and is on par with less transparent recommender methods. A short presentation can be found at https://youtu.be/neH5JtkPuVs.

### Keywords
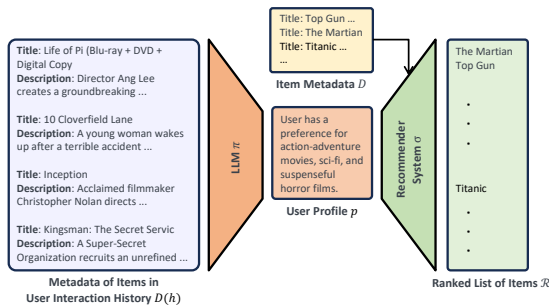
Large Language Models, Recommendation, Alignment

**Figure 1:** Illustration of LangPTune Pipeline

## 1. Introduction

The rapid advancement in large language models (LLMs) enables the use of natural language for improving the transparency, scrutability, and ultimately steerability of recommendation platforms [1, 2]. In particular, recent work has investigated language-based user profiles, where the user's preferences are represented as human-readable free-form text. Such language-based user profiles promise enhanced steerability by allowing users to better understand and adjust how recommendation systems perceive them [3, 4, 5, 6]. These LLM-generated profiles can also serve as additional features in traditional recommender systems, boosting their performance [7, 8, 9]. However, none of these works have addressed the problem of training LLMs to generate effective user profiles – they generally use zero-shot or few-shot inference, prompt tuning, or separate feature extraction algorithms instead.

In this paper, we introduce the first end-to-end training pipeline designed to enable LLMs to generate more effective user profiles for recommendation. The inference process of our pipeline is illustrated in Fig. 1. Given a list of items that the user has interacted with, our goal is to produce a high-quality user profile via an LLM that maximizes the downstream performance of the recommender system. Drawing on insights from Reinforcement Learning from Human Feedback (RLHF), we propose a novel approach called *Reinforcement Learning for System Optimization* (**RLSO**).

For each user, RLSO samples multiple profiles and uses the downstream recommender as a reward model to evaluate these generated profiles. Using a sequence of least-squares optimization problems that relate to Natural Policy Gradient steps, the LLM is then optimized based on these rewards, as illustrated in Fig. 2. Furthermore, we jointly optimize the downstream embedding model of the recommender alongside the LLM, which completes our end-to-end training framework that we call **LangPTune** (*Lang*uage-based *Profile Tune*r).

We evaluate LangPTune against a variety of baseline methods spanning different recommendation approaches. Our method with active training significantly outperforms other profile-based methods and exhibits comparable performance to the conventional recommendation models. Additionally, we provide detailed ablation studies on profile length and training dataset size.

## 2. Related Work

**LLM for Recommender Systems.** Our research falls within the broad spectrum of using LLMs for recommendation, which can largely be categorized into two main approaches. The first approach involves directly training LLMs for item or rating prediction [10, 11, 12, 13, 14, 15, 16], or otherwise using them as system augmentation [17, 18]. With this approach, user interaction histories and item metadata are provided as inputs to the LLM, which is then trained to predict the items most relevant to the user. These works take advantage of LLM's ability to learn causal patterns for item recommendation, and generally do not prioritize transparency or explainability as much. The second approach leverages LLMs to generate additional user or item profiles/descriptions based on existing features [19, 7, 8, 9]. These works demonstrate that incorporating additional features generated by LLMs could enhance the performance of conventional recommendation systems.

**Natural language-based profile methods.** We focus specifically on the use of LLMs for generating user profiles, which has garnered increasing attention for enhancing the transparency and explainability of recommendation systems [3]. Yang et al. [4] constructs user profiles based on observed user behaviors, which are subsequently used to prompt an LLM to perform personalized recommendations. Additionally, Ramos et al. [5] explores various prompting strategies aimed at improving the quality
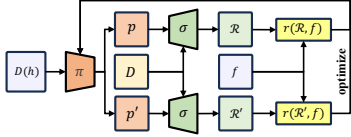
**Figure 2: Language Model Optimization Pipeline**

of LLM-generated user profiles. Their work also examines the system's scrutability by observing how modifications to the user profile affect the resulting recommendations. Zhou et al. [6] demonstrates that generating a compact and human-readable summary often performs comparably with or better than direct LLM prediction. However, none of these approaches allows end-to-end training to optimize profiles.

# 3. End-to-End Training with Language-Based Profiles

In the following, we develop our end-to-end training framework. Consider a recommendation system $\sigma$ that uses language-based profiles with $U$ users and $I$ items. We denote $h \in H$ as the user interaction history that consists of a list of items, e.g. $h = [i_1, i_2, \ldots, i_{|h|}]$, and $f \in F$ as the list of items that the user will interact with in the future. We use $d_i \in \mathcal{D}$ to denote the metadata for item $i$ and $p_u$ to denote the user profile for user $u$. Typically, $p_u$ conveys the metadata of items that are in the user interaction history $h_u$, i.e. $p_u = [d_{i_1}, \ldots, d_{i_{|h_u|}}]$.

The goal of the system $\sigma$ is to produce a ranked list of $I$ items given the user profile $p$ and all item metadata $\mathcal{D}$ such that the items the user will interact with in the future, $f$, will be at the top of the ranked list. Formally, we denote $\mathcal{R} = \sigma(p, \mathcal{D})$ as the ranked list and $r(\mathcal{R}, f)$ as the reward where $r$ could be some metric like Mean Reciprocal Rank (MRR), Normalized Discounted Cumulative Gain (NDCG), etc. To improve the scrutability and steerability of the system, we introduce a large language model (LLM), denoted by $\pi$, into the framework, as illustrated in Fig. 1. We use the LLM to generate user profile $p$ based on the metadata of items that are in the user interaction history, $p \sim \pi(\cdot | d_{i_1}, \ldots, d_{i_{|h|}})$, while maximizing the downstream performance of the recommender system $\sigma$ in terms of $r(\mathcal{R}, f)$. We use $D(h)$ to denote $[d_{i_1}, \ldots, d_{i_{|h|}}]$, and use $\pi$ to denote a policy (i.e. LLM) that maps from $D(h)$ to a distribution over the profile space $\Delta\mathcal{P}$.

In our implementation, the text-based recommender system $\sigma$ contains a text embedding model denoted by $\sigma_\theta$. Given $p$ and $\mathcal{D}$, $\sigma_\theta$ embeds the user profile $p$ and all item metadata $d_i \in \mathcal{D}$. The ranked list is produced based on the similarity between the user embedding $\sigma_\theta(p)$ and the item embedding $\sigma_\theta(d_i)$.

## 3.1. Language Model Optimization

To optimize the policy $\pi$, at each iteration of our algorithm, we aim to solve the following KL-constrained RL problem:

$$\pi_{t+1} = \arg\max_\pi \mathbb{E}_{h,f,p \sim \pi(\cdot|D(h))} \Big[ r(\sigma(p, \mathcal{D}), f) - $$
$$\frac{1}{\eta} \text{KL}\left(\pi(\cdot|D(h)) || \pi_t(\cdot|D(h))\right) \Big] \quad (1)$$

Intuitively, we aim to maximize the metric of the recommender system, $r(\sigma(p, \mathcal{D}), f)$, while staying close to the $\pi_t$ by minimizing the Kullback–Leibler (KL) divergence between $\pi$ and $\pi_t$. From Ziebart et al. [20], there exists a closed-form solution to the above minimum relative entropy problem [21]:

$$\forall h, f, p : \pi_{t+1}(p|D(h))$$
$$= \frac{\pi_t(p|D(h)) \exp(\eta r(\sigma(p, \mathcal{D}), f))}{Z(h, f)} \quad (2)$$

$$Z(h, f) = \sum_p \pi_t(p|D(h)) \exp(\eta r(\sigma(p, \mathcal{D}), f)) \quad (3)$$

Based on Rafailov et al. [22], we can invert Eq. 2 and write the reward as a function of the policy:

$$\forall h, f, p : r(\sigma(p, \mathcal{D}), f) = \frac{1}{\eta}\left(\ln(Z(h, f)) + \ln\left(\frac{\pi_{t+1}(p|D(h))}{\pi_t(p|D(h))}\right)\right) \quad (4)$$

From Gao et al. [23], Eq. 4 can be translated into a squared loss objective

$$\left(r(\sigma(p, \mathcal{D}), f) - \frac{1}{\eta}\left(\ln(Z(h, f)) + \ln\left(\frac{\pi_{t+1}(p|D(h))}{\pi_t(p|D(h))}\right)\right)\right)^2 \quad (5)$$

and we can sample two profiles to cancel the partition function $Z$ as it does not depend on $p$:

$$\left(r(\sigma(p, \mathcal{D}), f) - r(\sigma(p', \mathcal{D}), f) - \right.$$
$$\left. \frac{1}{\eta}\left(\ln\left(\frac{\pi_{t+1}(p|D(h))}{\pi_t(p|D(h))}\right) - \ln\left(\frac{\pi_{t+1}(p'|D(h))}{\pi_t(p'|D(h))}\right)\right)\right)^2 \quad (6)$$

Here, $p \sim \pi_t(\cdot|D(h))$ and $p' \sim \pi_t(\cdot|D(h))$.

To optimize $\pi$ with the above squared loss objective, we begin by sampling two profiles, $p$ and $p'$, from $\pi_t$. Next, we obtain two ranked lists, $\mathcal{R}$ and $\mathcal{R}'$, and calculate the corresponding rewards (e.g., NDCG), $r(\mathcal{R}, f)$ and $r(\mathcal{R}', f)$. The optimization of $\pi_t$ is then guided by the relative difference between these rewards. This process is illustrated in Fig. 2. Intuitively, if profile $p$ is better than $p'$, i.e. $r(\sigma(p, \mathcal{D}), f) > r(\sigma(p', \mathcal{D}), f)$, we aim to optimize the policy to increase the likelihood of generating $p$ over $p'$, i.e. increasing $\ln\left(\frac{\pi_{t+1}(p|D(h))}{\pi_t(p|D(h))}\right) - \ln\left(\frac{\pi_{t+1}(p'|D(h))}{\pi_t(p'|D(h))}\right)$. We refer to this training algorithm as *Reinforcement Learning for System Optimization* (**RLSO**), since it trains the LLM to produce profiles that produce accurate recommendations.

## 3.2. Recommender System Optimization

The text-based recommender system $\sigma$ might not be aligned well with the LLM-generated user profile $p$. Therefore, we jointly train $\sigma$ using InfoNCE loss [24]. Given a user profile $p$, a batch of items $B$ with one positive item $i_p$, the objective is to increase the similarity between $\sigma_\theta(p)$ and $\sigma_\theta(d_{i_p})$ while decreasing the similarity between $\sigma_\theta(p)$ and all other items in the batch:

$$\sum_{p,B} -\ln \frac{\exp(sim(\sigma_\theta(p), \sigma_\theta(d_{i_p})))}{\sum_{i \in B} \exp(sim(\sigma_\theta(p), \sigma_\theta(d_i)))} \quad (7)$$

We refer to this training method as *Contrastive Learning* (**CL**) analogous to Chen et al. [25].

### 3.3. Joint Model Optimization

Combining the language model optimization and recommender system optimization, we arrive at our joint optimization pipeline, **LangPTune**. The pseudocode is provided in Appendix A. We alternate the optimization between the two models for $K$ iterations. During each iteration, we optimize the recommender system using a **CL** objective with $J$ batches and the language model using **RLSO** with $T$ batches.

## 4. Experiments

We evaluated our proposed method, **LangPTune**, against multiple baselines on the Amazon-Movie-TV dataset [26].

### 4.1. Dataset

For Amazon-Movie-TV, we construct item metadata to include the title, description, average rating, category, and price. In addition to the metadata of the items in the user interaction history, $D(h)$, we also include the user reviews of the items as additional inputs to the policy. Examples of the item metadata and the policy input are shown in Appendix B.

We filter out sessions and items that have less than 5 interactions. To satisfy the context length of the LLM and embedding model, we additionally filter out items with metadata that are longer than 512 tokens and sessions with a history longer than 1024 tokens. Each review session contains 4 user interaction history items and 1 "upcoming" item, where it is guaranteed all five items are distinct from each other, i.e. $|h| = 4$ and $|f| = 1$. Overall, we have 17,482 sessions, which we split into 84% training, 8% validation, 8% test set (14,698 training, 1,392 validation, 1,392 test sessions). There are a total of 10,533 distinct items within the filtered dataset. To evaluate top-k ranking performance for each model, we adopt four widely used metrics: Normalized Discounted Cumulative Gain (NDCG), Mean Reciprocal Rank (MRR), Recall, and Precision [27, 28, 29].

### 4.2. Baselines

We compared **LangPTune** against multiple baseline types that span a range of recommendation approaches.

**Simple methods** use basic dataset statistics. These baselines do not train anything beyond simple item counting. **Random** generates a randomly shuffled ranked list of all items. **Most Popular (MP)** ranks all items in order of descending popularity.[1]

**ID-based methods**[2] use user and item ID interaction histories without any additional item metadata and take a collaborative filtering approach without employing LLMs. **Non-negative matrix factorization (NMF)** [30] decomposes user-item interaction data into latent factors and ranks the items based on the predictions of missing preferences. **LightGCN** [31] uses Graph Convolutional Networks (GCN) [32] with Bayesian Personalized Ranking (BPR) loss [33] to learn robust user/item embeddings. **MCL** [28] builds on top of LightGCN using sample mining and a mixed-centric loss.

**LLM-based methods** use a large language model as a part of their architecture. **BERT4REC** [27] uses masked language modeling (MLM) objective in BERT [34] to learn user/item embeddings. **LLM2BERT4REC** [19] initializes then trains a BERT4REC model with embeddings obtained from a text embedding model using item descriptions. **Zero-shot Embedding Model (EMB-0)** uses a pre-trained text embedding model to generate user and item embeddings directly from user interaction history with a list of item metadata and individual item metadata respectively. This approach is analogous to our pipeline illustrated in Fig. 1, but without LLM $\pi$ and User Profile $p$. **CL Embedding Model (EMB-CL)** is comparable to **EMB-0** but the embedding model is trained with InfoNCE loss in Eq. 7.

**Profile-based methods** generate natural language-based profiles to capture user preferences, then recommend based on the profile text. **LangPTune** is our proposed algorithm where we use RLSO to optimize the LLM and InfoNCE to optimize the embedding model. We compare against multiple ablation baselines. **LangPTune-0** only performs zero-shot inference with the pre-trained models without training either of the models. **LangPTune-CL** only trains the embedding model with InfoNCE without optimizing the LLM. This baseline is analogous to previous profile-based methods in the literature. **LangPTune-RLSO** only trains the LLM with RLSO without optimizing the embedding model.

### 4.3. Implementation Details

We tune each baseline separately on the dataset to obtain the best performance. Hyperparameter details of each method are shown in Appendix C. We implement our profile-based methods using both the Gemma-2B-it[3] [35] and the Llama-3-8B-it[4] [36] LLMs to evaluate our approach across different LLM sizes and architectures. We set a maximum profile length of 512 tokens. Following Zhu et al. [37] and Gao et al. [23], we only train the last four layers of the Llama-3-8B-it while keeping other layers frozen. Gemma-2B-it is trained with Low-Rank Adaptors (LoRA) [38]. Llama-3-8B-it experiments are trained on 8 A6000 GPUs for two days, and Gemma-2B-it experiments are trained for one day. For the LLM-based and Profile-based methods excluding BERT4REC, we adopt the Mxbai[5] [39] embedding model, and perform full-parameter training on 8 A6000 GPUs with each iteration taking one hour.

In our implementation, we use NDCG as the reward function $r$ as it is bounded between 0 and 1 when $|f| = 1$. In addition, we standardize the rewards within each batch to zero-mean and unit-variance to further stabilize the learning of the policy. To ensure that $\pi_\theta$ remains close to $\pi_{\theta_0}$, we apply an additional KL penalty to the reward:

$$r(\sigma(p, \mathcal{D}), f) \leftarrow r(\sigma(p, \mathcal{D}), f) - \gamma \left( \ln \pi_\theta(p|D(h)) - \ln \pi_{\theta_0}(p|D(h)) \right) \tag{8}$$

where $\gamma$ is a hyperparameter that controls the penalty [40, 23]. Furthermore, to ensure that the generations terminate within the maximum generation length, following Huang et al. [40], we penalize any generation that exceeds this length by setting $r(\sigma(p, \mathcal{D}), f)$ to a small fixed constant, $\Gamma$.

---

[1]Popularity is observed based on the training set and user interaction history items of the validation/test set.

[2]For all ID-based baselines, we use both the training set and the user interaction history items of the validation/test set during training.

[3]HuggingFace Model Card: google/gemma-2b-it

[4]HuggingFace Model Card: meta-llama/Meta-Llama-3-8B-Instruct

[5]HuggingFace Model Card: mixedbread-ai/mxbai-embed-large-v1

| LLM | Method | NDCG@20 | MRR@20 | Recall@20 | Precision@20 |
|---|---|---|---|---|---|
| Gemma-2B-it | LangPTune-0 | 0.0289 | 0.0189 | 0.0643 | 0.0032 |
| | LangPTune-CL | <u>0.0436</u> | <u>0.0317</u> | **0.0863** | **0.0043** |
| | LangPTune-RLSO | 0.0364 | 0.0253 | 0.0746 | 0.0037 |
| | LangPTune | **0.0446** | **0.0327** | <u>0.0859</u> | **0.0043** |
| Llama-3-8B-it | LangPTune-0 | 0.0231 | 0.0146 | 0.0535 | 0.0027 |
| | LangPTune-CL | <u>0.0456</u> | <u>0.0319</u> | <u>0.0938</u> | <u>0.0047</u> |
| | LangPTune-RLSO | 0.0377 | 0.0289 | 0.0763 | 0.0040 |
| | LangPTune | **0.0494** | **0.0358** | **0.0970** | **0.0048** |

**Table 1**
**Results on Amazon-Movie-TV for Profile-based Methods.**
The best-performing methods for each metric and base model
are highlighted in bold and the next best methods are underlined.

| Method | NDCG@20 | MRR@20 | Recall@20 | Precision@20 |
|---|---|---|---|---|
| Random | 0.0006 | 0.0003 | 0.0017 | 0.0001 |
| MP | 0.0033 | 0.0013 | 0.0108 | 0.0005 |
| NMF | 0.0343 | 0.0278 | 0.0575 | 0.0029 |
| LightGCN | 0.0525 | 0.0387 | 0.1009 | 0.0050 |
| MCL | 0.0563 | 0.0417 | 0.1078 | 0.0054 |
| BERT4REC | 0.0384 | 0.0280 | 0.0758 | 0.0038 |
| LLM2BERT4REC | 0.0423 | 0.0312 | 0.0783 | 0.0039 |
| EMB-0 | 0.0372 | 0.0261 | 0.0760 | 0.0038 |
| EMB-CL | 0.0543 | 0.0408 | 0.1012 | 0.0051 |
| LangPTune (Llama) | 0.0494 | 0.0358 | 0.0970 | 0.0048 |

**Table 2**
**Results on Amazon-Movie-TV with Different Recommender Approaches**

## 4.4. Performance Comparison among Profile-based Methods

The results for the profile-based methods are shown in Table 1. While both LangPTune-CL and LangPTune-RLSO show substantial improvements over the non-learning approach LangPTune-0, LangPTune achieves the highest performance compared to other profile-based methods in terms of NDCG, MRR, and Precision. Interestingly, the zero-shot performance of Llama is worse than Gemma, while, with any form of training, Llama surpasses Gemma. This could be because the profile generated by the initial Llama model is not in the same space as the item metadata after embedding. Training helps to align the profile with the item metadata in terms of embedding space. Examples of generated user profiles are shown in Appendix D.

## 4.5. Performance Comparison with Other Methods

We compare the best performing profile-based method, LangPTune (Llama), with other state-of-the-art recommendation methods, and the results are shown in Table 2. The ID-based methods perform slightly better than our profile-based method, but they lack the interpretability and steerability of language-based profiles. Compared to EMB-CL, we expected that the requirement to produce a language-based profile is a bottleneck that leads to marginally lower performance. EMB-CL avoids the bottleneck present in our pipeline by directly utilizing the list of item metadata without incurring information loss. It is important to note that profile-based methods inherently experience some degree of information loss due to the generally shorter and more concise nature of profiles.

| Profile Length | NDCG@20 | MRR@20 | Recall@20 | Precision@20 |
|---|---|---|---|---|
| 512 | 0.0494 | 0.0358 | 0.0970 | 0.0048 |
| 256 | 0.0492 | 0.0373 | 0.0956 | 0.0048 |
| 128 | 0.0444 | 0.0313 | 0.0901 | 0.0045 |
| 64 | 0.0319 | 0.0201 | 0.0604 | 0.0030 |

**Table 3**
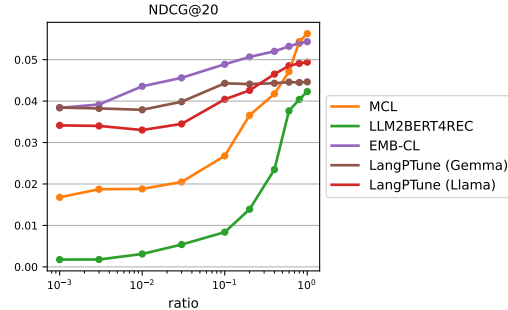**Ablation on Profile Length**



Figure 3: Ablation on Training Dataset Ratio

## 4.6. Ablation Analysis

**Profile Length.** The length of the profile plays an important role as it controls the width of the information bottleneck in the pipeline. A shorter length would cause more information loss. We ablate the length from 512 to 64 and the results are shown in Table 3. The performance exhibits an almost monotonic decrease as the profile length decreases, with the rate of performance degradation accelerating as the profile length becomes shorter. Examples of profiles with different lengths are shown in Appendix D.

**Training Dataset Size.** To assess the sensitivity of the methods to the amount of available training data, we conducted an ablation study by gradually reducing the training data from 100% down to 0.1% of the full training set size. The results are shown in Fig. 3. MCL and LLM2BERT4REC experience significant performance degradation with reduced training data, as these methods require an explicit embedding for each item. In contrast, text-based methods such as EMB-CL and LangPTune are less affected, as they operate directly on text-based user profiles and item metadata. Their performance remains nearly consistent across the different dataset sizes. This suggests that profile-based methods may be particularly suitable for cold-start scenarios, where there is little training data and where users may be particularly interested in directly steering the recommendations through edits to the profile. A full evaluation over NDCG, MRR, Recall, and Precision is shown in Appendix E.

## 5. Discussion & Future Work

There are multiple open questions highlighted by our work. One major question is how to qualitatively evaluate the generated user profiles and the downstream recommendation system for user transparency, scrutability, or steerability. It would be interesting to see a user study evaluating the user profiles throughout the training stages. An intuitive next step is to evaluate the effects of user changes to profile texts. On the technical side, we could try to combine the optimization of the language model and the recommender subsystems together to simplify the training pipeline. Finally, instead of only focusing on user profile generation, we could adapt our pipeline to obtain better item descriptions based on the item reviews for better recommendations.

# Acknowledgments

# References

[1] J. Liu, C. Liu, P. Zhou, R. Lv, K. Zhou, Y. Zhang, Is chatgpt a good recommender? a preliminary study, 2023. URL: https://arxiv.org/abs/2304.10149. arXiv:2304.10149.

[2] S. Sanner, K. Balog, F. Radlinski, B. Wedin, L. Dixon, Large language models are competitive near cold-start recommenders for language- and item-based preferences, in: Proceedings of the 17th ACM Conference on Recommender Systems, RecSys '23, Association for Computing Machinery, New York, NY, USA, 2023, p. 890–896. URL: https://doi.org/10.1145/3604915.3608845. doi:10.1145/3604915.3608845.

[3] F. Radlinski, K. Balog, F. Diaz, L. Dixon, B. Wedin, On natural language user profiles for transparent and scrutable recommendation, in: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '22, ACM, 2022. URL: http://dx.doi.org/10.1145/3477495.3531873. doi:10.1145/3477495.3531873.

[4] F. Yang, Z. Chen, Z. Jiang, E. Cho, X. Huang, Y. Lu, Palr: Personalization aware llms for recommendation, 2023. URL: https://arxiv.org/abs/2305.07622. arXiv:2305.07622.

[5] J. Ramos, H. A. Rahmani, X. Wang, X. Fu, A. Lipani, Transparent and scrutable recommendations using natural language user profiles, 2024. URL: https://arxiv.org/abs/2402.05810. arXiv:2402.05810.

[6] J. Zhou, Y. Dai, T. Joachims, Language-based user profiles for recommendation, 2024. URL: https://arxiv.org/abs/2402.15623. arXiv:2402.15623.

[7] X. Ren, W. Wei, L. Xia, L. Su, S. Cheng, J. Wang, D. Yin, C. Huang, Representation learning with large language models for recommendation, in: Proceedings of the ACM Web Conference 2024, volume 33 of *WWW '24*, ACM, 2024, p. 3464–3475. URL: http://dx.doi.org/10.1145/3589334.3645458. doi:10.1145/3589334.3645458.

[8] H. Lyu, S. Jiang, H. Zeng, Y. Xia, Q. Wang, S. Zhang, R. Chen, C. Leung, J. Tang, J. Luo, Llm-rec: Personalized recommendation via prompting large language models, 2024. URL: https://arxiv.org/abs/2307.15780. arXiv:2307.15780.

[9] S. Yang, W. Ma, P. Sun, Q. Ai, Y. Liu, M. Cai, M. Zhang, Sequential recommendation with latent relations based on large language model, 2024. URL: https://arxiv.org/abs/2403.18348. arXiv:2403.18348.

[10] Y. Zhang, F. Feng, J. Zhang, K. Bao, Q. Wang, X. He, Collm: Integrating collaborative embeddings into large language models for recommendation, 2023. URL: https://arxiv.org/abs/2310.19488. arXiv:2310.19488.

[11] J. Ji, Z. Li, S. Xu, W. Hua, Y. Ge, J. Tan, Y. Zhang, Genrec: Large language model for generative recommendation, 2023. URL: https://arxiv.org/abs/2307.00457. arXiv:2307.00457.

[12] K. Bao, J. Zhang, Y. Zhang, W. Wang, F. Feng, X. He, Tallrec: An effective and efficient tuning framework to align large language model with recommendation, in: Proceedings of the 17th ACM Conference on Recommender Systems, RecSys '23, ACM, 2023. URL: http://dx.doi.org/10.1145/3604915.3608857. doi:10.1145/3604915.3608857.

[13] W.-C. Kang, J. Ni, N. Mehta, M. Sathiamoorthy, L. Hong, E. Chi, D. Z. Cheng, Do llms understand user preferences? evaluating llms on user rating prediction, 2023. URL: https://arxiv.org/abs/2305.06474. arXiv:2305.06474.

[14] J. Liao, S. Li, Z. Yang, J. Wu, Y. Yuan, X. Wang, X. He, Llara: Large language-recommendation assistant, 2024. URL: https://arxiv.org/abs/2312.02445. arXiv:2312.02445.

[15] S. Geng, S. Liu, Z. Fu, Y. Ge, Y. Zhang, Recommendation as Language Processing (RLP): A Unified Pretrain, Personalized Prompt & Predict Paradigm (P5), 2023. URL: http://arxiv.org/abs/2203.13366. doi:10.48550/arXiv.2203.13366, arXiv:2203.13366 [cs].

[16] L. Xu, J. Zhang, B. Li, J. Wang, M. Cai, W. X. Zhao, J.-R. Wen, Prompting Large Language Models for Recommender Systems: A Comprehensive Framework and Empirical Analysis, 2024. URL: http://arxiv.org/abs/2401.04997. doi:10.48550/arXiv.2401.04997, arXiv:2401.04997 [cs].

[17] X. Li, B. Chen, L. Hou, R. Tang, CTRL: Connect Collaborative and Language Model for CTR Prediction, 2023. URL: http://arxiv.org/abs/2306.02841. doi:10.48550/arXiv.2306.02841, arXiv:2306.02841 [cs].

[18] Y. Xi, W. Liu, J. Lin, X. Cai, H. Zhu, J. Zhu, B. Chen, R. Tang, W. Zhang, R. Zhang, Y. Yu, Towards Open-World Recommendation with Knowledge Augmentation from Large Language Models, 2023. URL: http://arxiv.org/abs/2306.10933. doi:10.48550/arXiv.2306.10933, arXiv:2306.10933 [cs].

[19] J. Harte, W. Zorgdrager, P. Louridas, A. Katsifodimos, D. Jannach, M. Fragkoulis, Leveraging large language models for sequential recommendation, in: Proceedings of the 17th ACM Conference on Recommender Systems, RecSys '23, ACM, 2023. URL: http://dx.doi.org/10.1145/3604915.3610639. doi:10.1145/3604915.3610639.

[20] B. D. Ziebart, A. L. Maas, J. A. Bagnell, A. K. Dey, et al., Maximum entropy inverse reinforcement learning., in: Aaai, volume 8, Chicago, IL, USA, 2008, pp. 1433–1438.

[21] P. D. Grünwald, A. P. Dawid, Game theory, maximum entropy, minimum discrepancy and robust bayesian decision theory, The Annals of Statistics 32 (2004). URL: http://dx.doi.org/10.1214/009053604000000553. doi:10.1214/009053604000000553.

[22] R. Rafailov, A. Sharma, E. Mitchell, S. Ermon, C. D. Manning, C. Finn, Direct preference optimization: Your language model is secretly a reward model, 2023. arXiv:2305.18290.

[23] Z. Gao, J. D. Chang, W. Zhan, O. Oertell, G. Swamy, K. Brantley, T. Joachims, J. A. Bagnell, J. D. Lee, W. Sun, Rebel: Reinforcement learning via regressing relative rewards, 2024. URL: https://arxiv.org/abs/2404.16767. arXiv:2404.16767.

[24] A. van den Oord, Y. Li, O. Vinyals, Representation learning with contrastive predictive cod-

ing, 2019. URL: https://arxiv.org/abs/1807.03748. arXiv:1807.03748.

[25] T. Chen, S. Kornblith, M. Norouzi, G. Hinton, A simple framework for contrastive learning of visual representations, 2020. URL: https://arxiv.org/abs/2002.05709. arXiv:2002.05709.

[26] Y. Hou, J. Li, Z. He, A. Yan, X. Chen, J. McAuley, Bridging language and items for retrieval and recommendation, arXiv preprint arXiv:2403.03952 (2024).

[27] F. Sun, J. Liu, J. Wu, C. Pei, X. Lin, W. Ou, P. Jiang, Bert4rec: Sequential recommendation with bidirectional encoder representations from transformer, 2019. URL: https://arxiv.org/abs/1904.06690. arXiv:1904.06690.

[28] Z. Gao, Z. Cheng, F. Perez, J. Sun, M. Volkovs, Mcl: Mixed-centric loss for collaborative filtering, in: Proceedings of the ACM Web Conference 2022, WWW '22, Association for Computing Machinery, New York, NY, USA, 2022, p. 2339–2347. URL: https://doi.org/10.1145/3485447.3512106. doi:10.1145/3485447.3512106.

[29] Y. Lu, Z. Gao, Z. Cheng, J. Sun, B. Brown, G. Yu, A. Wong, F. Pérez, M. Volkovs, Session-based recommendation with transformers, in: Proceedings of the Recommender Systems Challenge 2022, RecSysChallenge '22, Association for Computing Machinery, New York, NY, USA, 2022, p. 29–33. URL: https://doi.org/10.1145/3556702.3556844. doi:10.1145/3556702.3556844.

[30] D. Lee, H. S. Seung, Algorithms for non-negative matrix factorization, in: T. Leen, T. Dietterich, V. Tresp (Eds.), Advances in Neural Information Processing Systems, volume 13, MIT Press, 2000. URL: https://proceedings.neurips.cc/paper_files/paper/2000/file/f9d1152547c0bde01830b7e8bd60024c-Paper.pdf.

[31] X. He, K. Deng, X. Wang, Y. Li, Y. Zhang, M. Wang, Lightgcn: Simplifying and powering graph convolution network for recommendation, 2020. URL: https://arxiv.org/abs/2002.02126. arXiv:2002.02126.

[32] T. N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, 2017. URL: https://arxiv.org/abs/1609.02907. arXiv:1609.02907.

[33] S. Rendle, C. Freudenthaler, Z. Gantner, L. Schmidt-Thieme, Bpr: Bayesian personalized ranking from implicit feedback, 2012. URL: https://arxiv.org/abs/1205.2618. arXiv:1205.2618.

[34] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, 2019. URL: https://arxiv.org/abs/1810.04805. arXiv:1810.04805.

[35] G. Team, Gemma 2: Improving open language models at a practical size, 2024. URL: https://arxiv.org/abs/2408.00118. arXiv:2408.00118.

[36] L. Team, The llama 3 herd of models, 2024. URL: https://arxiv.org/abs/2407.21783. arXiv:2407.21783.

[37] B. Zhu, E. Frick, T. Wu, H. Zhu, J. Jiao, Starling-7b: Improving llm helpfulness & harmlessness with rlaif, 2023.

[38] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, W. Chen, Lora: Low-rank adaptation of large language models, 2021. URL: https://arxiv.org/abs/2106.09685. arXiv:2106.09685.

[39] S. Lee, A. Shakir, D. Koenig, J. Lipp, Open source strikes bread - new fluffy embeddings model, 2024. URL: https://www.mixedbread.ai/blog/mxbai-embed-large-v1.

[40] S. Huang, R. F. J. Dossa, A. Raffin, A. Kanervisto, W. Wang, The 37 implementation details of proximal policy optimization, in: ICLR Blog Track, 2022. URL: https://iclr-blog-track.github.io/2022/03/25/ppo-implementation-details/, https://iclr-blog-track.github.io/2022/03/25/ppo-implementation-details/.

# A. Algorithm

---

**Algorithm 1 LangPTune** Optimization

---

1: **Input**: reward $r$, policy class $\Pi$, learning rate $\eta$, model class $\Sigma_\theta$, similarity function $sim(\cdot, \cdot)$
2: Initialize policy $\pi$ and model $\sigma$ which uses $\sigma_\theta$ as its embedding model.
3: **for** $k = 0$ to $K - 1$ **do**
4:    $\sigma_{\theta_0} \leftarrow \sigma_\theta$
5:    **for** $j = 0$ to $J - 1$ **do**
6:       Collect dataset $\epsilon_j = \{h, p, B\}$ where $p \sim \pi(\cdot | D(h))$, $B$ is a list of items with one positive item $i_p$.
7:       Optimize for cross entropy objective:

$$\sigma_{\theta_{j+1}} = \arg\min_{\sigma_\theta \in \Sigma_\theta} \sum_{(p,B) \in \epsilon_t} -\ln \frac{\exp(sim(\sigma_{\theta_j}(p), \sigma_{\theta_j}(d_{i_p})))}{\sum_{i \in B} \exp(sim(\sigma_{\theta_j}(p), \sigma_{\theta_j}(d_i)))}$$

8:    **end for**
9:    $\sigma_\theta \leftarrow \sigma_{\theta_J}$
10:    Set $\sigma$ to use $\sigma_\theta$ as its embedding model
11:    $\pi_0 \leftarrow \pi$
12:    **for** $t = 0$ to $T - 1$ **do**
13:       Collect dataset $\xi_t = \{h, f, p, p'\}$ where $p \sim \pi_t(\cdot | H), p' \sim \pi_t(\cdot | H)$.
14:       Solve the squared regression problem:

$$\pi_{t+1} = \arg\min_{\pi \in \Pi} \sum_{(h,f,p,p') \in \xi_t} \left( r(\sigma(p, \mathcal{D}), f) - r(\sigma(p', \mathcal{D}), f) - \frac{1}{\eta} \left( \ln\left( \frac{\pi(p|D(h))}{\pi_t(p|D(h))} \right) - \ln\left( \frac{\pi(p'|D(h)}{\pi_t(p'|D(h))} \right) \right) \right)^2$$

15:    **end for**
16:    $\pi \leftarrow \pi_T$
17: **end for**

---

# B. Dataset Details

The session description consists of a sequence of item metadatas in the session. Examples of a session description and an item metadata are shown below:

**Amazon-Movie-TV | Session Description Example**

---

**Title**: Top Gun [DVD & Digital Copy Included] [3D Blu-ray]
**Description**: (average rating: 4.7) Top Gun takes a look at the danger and excitement that awaits every pilot at the Navy's prestigious fighter weapons school. Tom Cruise is superb as Pete "Maverick" Mitchell, a daring young flyer who's out to become the best of the best. And Kelly McGillis sizzles as the civilian instructor who teaches Maverick a few things you can't learn in a classroom.
**Category**: Movies & TV, Blu-ray, Movies
**Price**: 17.4
**Review from the user**: (rating: 1.0) Always liked this movie and was excited to see that it was be re-released in 3D BluRay. Well it is 3D, but the quality is horrible. The picture in some spots looks like VHS. The picture in the best spots looks no better than DVD.<br /><br />I'm learning that I shouldn't spend good money to upgrade DVD films to 3D or Blu Ray.

**Title**: Person of Interest: Season 4 (Blu-ray + Digital Copy)
**Description**: (average rating: 4.8) Person of Interest: The Complete Fourth Season (Blu-ray) ]]>
**Category**: Movies & TV, Blu-ray, TV
**Price**: 53.66
**Review from the user**: (rating: 4.0) Good, but not as good as the first three seasons. Too much about AI battles and not enough about people.

**Title**: The Martian
**Description**: (average rating: 4.8) From legendary director Ridley Scott (Alien, Prometheus) comes a gripping tale of human strength and the will to survive, starring Matt Damon as an astronaut stranded on Mars.
**Category**: Science Fiction, Drama, Anxious, Biting
**Price**: 4.99
**Review from the user**: (rating: 5.0) A really good movie. Well done in all aspects. Great story, great acting, great visual effects. Thoroughly enjoyed it.

**Title**: The Firm: The Complete Series
**Description**: (average rating: 4.3) As a young associate Mitch brought down Bendini Lambert & Locke a prestigious Memphis law firm operating as a front for the Chicago Mob. His life was never the same. After a difficult ten years which included a stay in the Federal Witness Protection Program Mitch and his family emerge from isolation determined to reclaim their lives and their future – only to find that past dangers are still lurking and new threats are everywhere.NOTE:Disc 1 is behind the insert attached to back of front cover.
**Category**: Movies & TV, Boxed Sets, Drama
**Price**: None
**Review from the user**: (rating: 4.0) Loved most of it. Got a little predictable towards the end. Great acting.

---

**Amazon-Movie-TV | Item Metadata Example**

---

**Title**: Outlander: Season 5 [DVD]
**Description**: (average rating: 4.9) The fifth season of Outlander sees a continuation of Claire and Jamie's fight to protect those they love, as they navigate the trials and tribulations of life in colonial America. The Frasers strive to flourish within a society which is unwittingly marching toward Revolution, and Jamie must now defend the home they have built together at Fraser's Ridge while Claire seeks to put her own skills and medical expertise to use in keeping her family together and safe from harm. Meanwhile, Brianna and Roger MacKenzie struggle to find their respective places in this world and chase away the shadow cast over their lives by Stephen Bonnet as they raise their son in this brave new world. For the Frasers and their family, "home" is more than simply a site on which they live, it is the place in which they are laying the foundations for the rest of their lives.
**Category**: Movies & TV, Studio Specials, Sony Pictures Home Entertainment, All Sony Pictures Titles
**Price**: 17.61

---

**Amazon-Movie-TV | Prompt for Generation**

---

The user has rated (out of 5.0) and reviewed following movies and TV shows arranged chronologically from the oldest (top) to the newest (bottom). Please provide a high-level summary of the user preference in detail.

---

# C. Hyperparameter Details

Tuned on validation dataset.

**Amazon-Movie-TV**

| Method | Parameters | |
|---|---|---|
| Non-negative matrix factorization (NMF) | Components: 1000<br>Solver: batch HALS | Beta loss: L2 |
| LightGCN | Learning rate: 1e-3<br>Batch size: 1000<br>Dimension: 512 | Layer: 3<br>Num negative: 1 |
| MCL | Learning rate: 1e-3<br>Batch size: 1000<br>Dimension: 512<br>Layer: 3<br>Num negative: 10 | $\alpha$: 1.25<br>$\beta$: 5.0<br>$\lambda_p$: 6.5<br>$\lambda_n$: -0.5 |
| BERT4REC | Batch size: 128<br>Learning rate: 1e-3<br>Decay step: 25<br>Epochs: 100<br>Dropout: 0.1 | Hidden units: 512<br>Mask prob: 0.5<br>Num blocks: 2<br>Num heads: 4 |
| LLM2BERT4REC | Batch size: 128<br>Learning rate: 1e-3<br>Decay step: 25<br>Epochs: 100<br>Dropout: 0.1 | Hidden units: 512<br>Mask prob: 0.5<br>Num blocks: 2<br>Num heads: 4 |
| EMB-CL | CL learning rate: 1e-5<br>CL batch size: 256<br>CL weight decay: 1e-6<br>Schedule: linear decay | $K$: 5<br>$J$: 800<br>Embedding size: 512 |
| LangPTune-CL | CL learning rate: 1e-5<br>CL batch size: 256<br>CL weight decay: 1e-6<br>Schedule: linear decay | $K$: 5<br>$J$: 800<br>Embedding size: 512 |
| LangPTune-RLSO (Gemma-2B-it) | RLSO learning rate: 1e-7<br>RLSO batch size: 64<br>RLSO weight decay: 1e-6<br>Schedule: linear decay<br>LoRA rank: 1024<br>LoRA alpha: 2048<br>$T$: 200 | $K$: 5<br>Num epochs: 4<br>$\eta$: 1.0<br>$\gamma$: 0.2<br>$\Gamma$: 0<br>Embedding size: 512 |
| LangPTune-RLSO (Llama-3-8B-it) | RLSO learning rate: 3e-7<br>RLSO batch size: 64<br>RLSO weight decay: 1e-6<br>Schedule: linear decay<br>$T$: 500<br>$K$: 5 | Num epochs: 4<br>$\eta$: 1.0<br>$\gamma$: 0<br>$\Gamma$: 0<br>Embedding size: 512 |
| LangPTune (Gemma-2B-it) | RLSO learning rate: 1e-7<br>RLSO batch size: 64<br>RLSO weight decay: 1e-6<br>CL learning rate: 1e-7<br>CL batch size: 256<br>CL weight decay: 1e-6<br>Schedule: linear decay<br>LoRA rank: 1024<br>LoRA alpha: 2048 | $T$: 200<br>$K$: 5<br>$J$: 800<br>Num epochs: 4<br>$\eta$: 1.0<br>$\gamma$: 0.2<br>$\Gamma$: 0<br>Embedding size: 512 |
| LangPTune (Llama-3-8B-it) | RLSO learning rate: 3e-7<br>RLSO batch size: 64<br>RLSO weight decay: 1e-6<br>CL learning rate: 3e-7<br>CL batch size: 256<br>CL weight decay: 1e-6<br>Schedule: linear decay<br>$T$: 500 | $K$: 5<br>$J$: 2500<br>Num epochs: 4<br>$\eta$: 1.0<br>$\gamma$: 0<br>$\Gamma$: 0<br>Embedding size: 512 |

# D. Example Generations

In this section, we show different model generations given the following input prompt:

**Input Prompt**

The user has rated (out of 5.0) and reviewed following movies and TV shows arranged chronologically from the oldest (top) to the newest (bottom). Please provide a high-level summary of the user preference in detail.

**Title**: Life of Pi (Blu-ray + DVD + Digital Copy)
**Description**: (average rating: 4.7) Director Ang Lee creates a groundbreaking movie event about a young man who survives a tragic disaster at sea and is hurtled into an epic journey of adventure and discovery. While marooned on a lifeboat, he forms an amazing and unexpected connection with the ship's only other survivor...a fearsome Bengal tiger.
**Category**: Movies & TV, Featured Categories, Blu-ray, Action & Adventure
**Price**: 7.99
**Review from the user**: (rating: 3.0) great story, great visuals and great message. It's a long movie and some parts might be boring but good movie overall. Also there are more subtitles than indicated... I think it's subtitled with every language in the world!

**Title**: 10 Cloverfield Lane
**Description**: (average rating: 4.7) A young woman wakes up after a terrible accident to find that she's locked in a cellar with a doomsday prepper, who insists that he saved her life and that the world outside is uninhabitable following an apocalyptic catastrophe. Uncertain what to believe, the woman soon determines that she must escape at any cost. Movie stars John Goodman, Mary Elizabeth Winstead, John Gallagher Jr and Bradley Cooper.
**Category**: Movies & TV, Featured Categories, Blu-ray, Horror **Price**: 9.99
**Review from the user**: (rating: 5.0) great movie, very suspenseful and claustrophobic. kept me guessing to the end. never seen the original Cloverfield so didn't know what to expect, so of course the ending was a bit unexpected. acting is superb.

**Title**: Inception (Rpkg/BD) [Blu-ray]
**Description**: (average rating: 4.6) Inception (Rpkg/BD) Acclaimed filmmaker Christopher Nolan directs an international cast in this sci-fi actioner that travels around the globe and into the world of dreams. Dom Cobb (Leonardo DiCaprio) is the best there is at extraction: stealing valuable secrets inside the subconscious during the mind's vulnerable dream state. His skill has made him a coveted player in industrial espionage but also has made him a fugitive and cost him dearly. Now he may get a second chance if he can do the impossible: inception, planting an idea rather than stealing one. If they succeed, Cobb and his team could pull off the perfect crime. But no planning or expertise can prepare them for a dangerous enemy that seems to predict their every move. An enemy only Cobb could have seen coming. ]]>
**Category**: Movies & TV, Blu-ray, Movies
**Price**: 12.88
**Review from the user**: (rating: 5.0) it's an awesome movie and surely most people know that already

**Title**: Kingsman: The Secret Service (Blu-ray + Digital Copy)
**Description**: (average rating: 4.7) A Super-Secret Organization recruits an unrefined but promising street kid into the agency's ultra-competitive training program just as a dire global threat emerges from a twisted tech genius. A phenomenal cast, including Academy Award Winners Colin Firth and Michael Caine and Oscar Nominee Samuel L. Jackson, leads this action-packed spy-thriller directed by Matthew Vaughn (X-Men: First Class).
**Category**: Movies & TV, Blu-ray, Movies
**Price**: 8.99
**Review from the user**: (rating: 5.0) Epic action film, recommended for action film fans.

**Gemma-2B-it** | $k = 5$, Maximum Generation Length = 512

User Preference Summary:
**Life of Pi (Blu-ray + DVD + Digital Copy)**: The user finds the movie enjoyable with visually captivating visuals, a thought-provoking story, and a positive message about courage and resilience. However, the lengthy runtime may be challenging for some viewers.
**10 Cloverfield Lane**: The user appreciates the suspenseful and claustrophobic atmosphere, as well as the unexpected ending that reveals the reality behind the apocalypse. The movie is a unique and refreshing take on a dystopian world.
**Inception (Rpkg/BD)**: The detailed and visual nature of the dream-world sequences in Inception is highly appreciated by the user. The film features exceptional acting and a thought-provoking narrative that explores the complexities of dreams and reality.
**Kingsman**: The Secret Service (Blu-ray + Digital Copy):** The user finds the movie thrilling and fast-paced, with an exhilarating action film combined with humorous comedic elements. The cast of skilled actors and the movie's fast-paced editing contribute to the overall enjoyment.
**Overall, the user has a diverse and positive taste in movies and TV shows, with a particular fondness for visual storytelling and thought-provoking narratives.**

**Llama-3-8B-it** | $k = 5$, Maximum Generation Length = 512

---

Based on the user's ratings and reviews, here is a summary of their preferences:

**Storytelling**: The user enjoys complex storylines with twists and surprises. They appreciate movies that explore themes of adventure, discovery, and suspense. The user particularly liked the unexpected connections and surprises in ""Life of Pi"" and the twists in ""10 Cloverfield Lane"".

**Visuals**: The user appreciates visually stunning movies with groundbreaking effects. They praised the visuals in ""Life of Pi"" and might have appreciated the action sequences and spy-thriller elements in ""Kingsman: The Secret Service"".

**Suspense and Thrills**: The user enjoys suspenseful movies that keep them guessing until the end. They highly rated ""10 Cloverfield Lane"" for its ability to maintain tension and uncertainty throughout the film.

**Acting:** The user values strong acting performances, praising the talents of the cast in ""10 Cloverfield Lane"" and potentially appreciating the acting in ""Kingsman: The Secret Service"".

**Blu-ray and Extras**: The user is willing to pay a premium price for Blu-ray copies of movies, suggesting that they appreciate high-quality visual and sound presentation.

**Movie Length**: The user might be willing to tolerate longer movie runtime if the story is engaging, as evidenced by their 3.0 rating for ""Life of Pi"", which they found was ""a long movie"".

**Subtitles**: The user might be sensitive to subtitles, noting that there were more subtitles than indicated in ""Life of Pi"".

**Action and Adventure**: The user enjoys action-packed movies, particularly spy-thrillers like ""Kingsman: The Secret Service"" and adventure films like ""10 Cloverfield Lane"".

**Sci-Fi and Fantasy**: The user might appreciate sci-fi elements, as evidenced by their appreciation for the themes of dreams and reality in ""Inception"".

**Classics**: The user might be willing to watch classics, but they didn't mention any particular appreciation for them.

Overall, the user enjoys complex storylines, suspenseful plots, and strong acting performances, particularly in action-packed and sci-fi thrillers.

---

**Llama-3-8B-it** | $k = 5$, Maximum Generation Length = 256

---

Based on the ratings and reviews provided, it appears that the user prefers suspenseful and action-packed movies with engaging storylines and strong suspense elements. They have rated:

Life of Pi (Blu-ray + DVD + Digital Copy) - 3/5 stars

10 Cloverfield Lane - 5/5 stars

Inception (Rpkg/BD) - 5/5 stars

Kingsman: The Secret Service (Blu-ray + Digital Copy) - 5/5 stars

The user praised the suspensefulness and claustrophobic atmosphere of 10 Cloverfield Lane, the action-packed nature of Kingsman: The Secret Service, and the engaging storytelling and suspense elements of Inception. They rated Life of Pi relatively highly, praising its story, visuals, and message, but criticized its length and the abundance of subtitles.

Overall, the user prefers movies with strong suspense elements, engaging storylines, and fast-paced action sequences.

---

**Llama-3-8B-it** | $k = 5$, Maximum Generation Length = 128

---

Based on the ratings and reviews, it is clear that the user has a strong preference for action-based Blu-ray movies, Blu-ray + Digital Copy films, and Blu-ray movies, such as ""Life of Pi"", ""10 Cloverfield Lane"", ""Inception"", ""Kingsman: The Secret Service"".

---

**Llama-3-8B-it** | $k = 5$, Maximum Generation Length = 64

---

Based on the ratings and reviews provided, it appears that the user has a preference for action-adventure movies, sci-fi, and suspenseful horror films.

---

**Llama-3-8B-it** | $k = 4$, Maximum Generation Length = 512

---

Based on the user's ratings and reviews, here is a summary of their preferences :

**Storytelling**: the user appreciates complex and suspenseful storylines, as evidenced by their ratings of 5. 0 for 10 cloverfield lane and kingsman : the secret service, and 3. 0 for life of pi. they enjoy plots that keep them guessing until the end.

**Action and Suspense**: the user is drawn to action - packed films, as demonstrated by their ratings of 5. 0 for kingsman : the secret service, 10 cloverfield lane, and inception. they seem to enjoy fast - paced plots with twists and turns.

**Visuals**: the user appreciates visually stunning films, as indicated by their ratings of 3. 0 for life of pi, which they praised for its great visuals.

**Acting**: the user values good acting, as evidenced by their ratings of 5. 0 for 10 cloverfield lane, which they praised for superb acting.

**thematic complexity**: the user may be drawn to films that deal with complex themes, as suggested by their appreciation for inception's exploration of reality and perception.

**Blu-ray Format**: the user appears to prioritize blu - ray formats, as they purchased and watched all the blu - ray movies on the list.

**Budget-friendly Options**: the user may be budget - conscious, as they seem to be willing to pay reasonable prices for blu - ray copies of movies, ranging from 7.99 to 12.88.

Overall, the user enjoys complex storylines, action - packed plots, suspenseful twists, and visually stunning films, but may be willing to overlook some pacing issues if the story and themes are compelling.

---

**Llama-3-8B-it** | $k = 3$, Maximum Generation Length = 512

---

Based on the user's ratings and reviews, here is a summary of their preferences :

**movie genre**: the user enjoys action - packed movies with a mix of adventure, suspense, and thriller elements. they appreciate movies that keep them guessing until the end and have a strong plot.

**storytelling**: the user values storytelling over special effects, as evident from their appreciation for movies like "" life of pi "" and "" 10 cloverfield lane "" that have strong narratives. they enjoy complex storylines and unexpected twists, as seen in their rating of "" 10 cloverfield lane "".

**suspense and thrills**: the user enjoys movies that are suspenseful and keep them on the edge of their seat. they appreciate claustrophobic settings, as seen in their rating of "" 10 cloverfield lane "".

**action and adventure**: the user enjoys action - packed movies with a mix of adventure and espionage, as evident from their ratings of "" kingsman : the secret service "" and "" inception "".

**language**: the user doesn't seem to have any issues with subtitles, as they didn't mention anything negative about the subtitles in their review of "" life of pi "".

overall, the user enjoys complex, suspenseful storylines with strong performances, action - packed plots, and high - quality visuals. they are willing to pay a premium for blu - ray copies of movies and appreciate movies that keep them guessing until the end.

---

**Llama-3-8B-it** | $k = 2$, Maximum Generation Length = 512

Based on the user's ratings and reviews, here is a summary of their preferences :
**movie genre**: the user enjoys action - packed movies, particularly sci - fi actioners ( inception ) and spy - thrillers ( kingsman : the secret service ). they also appreciate horror movies ( 10 cloverfield lane ) and adventure films ( life of pi ).
**storytelling**: the user values complex, thought - provoking storylines with unexpected twists and turns. they enjoyed the suspenseful plot of 10 cloverfield lane and the mind - bending narrative of inception.
**visuals**: the user appreciates visually stunning movies, noting the impressive visuals in life of pi. they also appreciated the claustrophobic settings and tense atmosphere in 10 cloverfield lane.
**acting**: the user highly praised the acting in 10 cloverfield lane, noting that the performances were superb. they also enjoyed the ensemble cast in kingsman : the secret service.
**pacing**: the user acknowledged that life of pi may have some slow parts, but overall, they enjoyed the film's epic journey. they appreciated the fast - paced action and suspenseful plot in 10 cloverfield lane and kingsman : the secret service.
**language**: the user noted that life of pi had more subtitles than indicated, which may be a consideration for future purchases.
**overall**: the user enjoys action - packed movies with complex storylines, suspenseful plots, and impressive visuals. they appreciate strong acting performances and are willing to overlook slower pacing parts in favor of enjoying the film's overall experience.

**Llama-3-8B-it** | $k = 1$, Maximum Generation Length = 512

Based on the user's ratings and reviews, here is a summary of their preferences:
**Overall Preferences**: The user enjoys action - packed movies with suspenseful plots, thrilling endings, and strong performances. They appreciate visually stunning movies with great storytelling, cinematography, and direction. The user seems to be fond of sci - fi, adventure, and spy thrillers, as seen in the movies "" life of pi "", "" 10 cloverfield lane "", "" inception "", and "" kingsman : the secret service "".
**Movie length and pacing**: The user mentioned that "" life of pi "" is a "" long movie "" and that some parts might be boring, suggesting that they prefer movies with a faster pace. However, they did not mention any issues with the pacing of the other movies, which may indicate that they are willing to invest time in a well - crafted story.
**Characteristics of well-received movies** : The user enjoyed movies with well - executed suspense, surprise endings, and strong performances. They appreciated movies with great visuals, direction, and storytelling. The user seems to value movies that keep them guessing until the end, as seen in their praise for the unexpected ending of "" 10 cloverfield lane "".
**Negative Feedback** : The user mentioned that "" life of pi "" has more subtitles than indicated, which might be a minor annoyance. They did not provide any negative feedback for the other movies, indicating that they were generally satisfied with their viewing experiences.
**Budget and Price** : The user seems to be willing to pay a premium for high - quality movies, as evidenced by their purchases of blu - ray copies and willingness to pay upwards of $ 10 for each movie.

**Llama-3-8B-it** | $k = 0$, Maximum Generation Length = 512

based on the user's reviews and ratings, here are some insights into their movie preferences :
1. **story is important**: the user appreciates a good story, as evident from their reviews of "" life of pi "" and "" 10 cloverfield lane "". they mention the importance of a great story, message, and script.
2. **action and suspense**: the user enjoys action - packed and suspenseful movies, as seen in their reviews of "" 10 cloverfield lane "", "" inception "", and "" kingsman : the secret service "". they appreciate films that keep them engaged and guessing until the end.
3. **visually impressive**: the user is impressed by visually stunning movies, as evident from their review of "" life of pi "", which they praised for its great visuals.
4. **credible acting**: the user values good acting, citing the superb acting in "" 10 cloverfield lane "" as one of the film's strengths.
5. **frequent surprises**: the user seems to enjoy unexpected twists and surprises, as seen in their reviews of "" 10 cloverfield lane "" and "" inception "", where they were kept guessing until the end.
6. **genre preferences**: the user appears to enjoy movies that fall under the action, adventure, horror, and sci - fi genres, as seen from their reviews of "" life of pi "", "" 10 cloverfield lane "", "" inception "", and "" kingsman : the secret service "".
overall, the user seems to appreciate movies that have a strong narrative, exciting action sequences, and impressive visuals, with a focus on suspense, twists, and surprises. they value good acting and are open to trying different genres and themes.
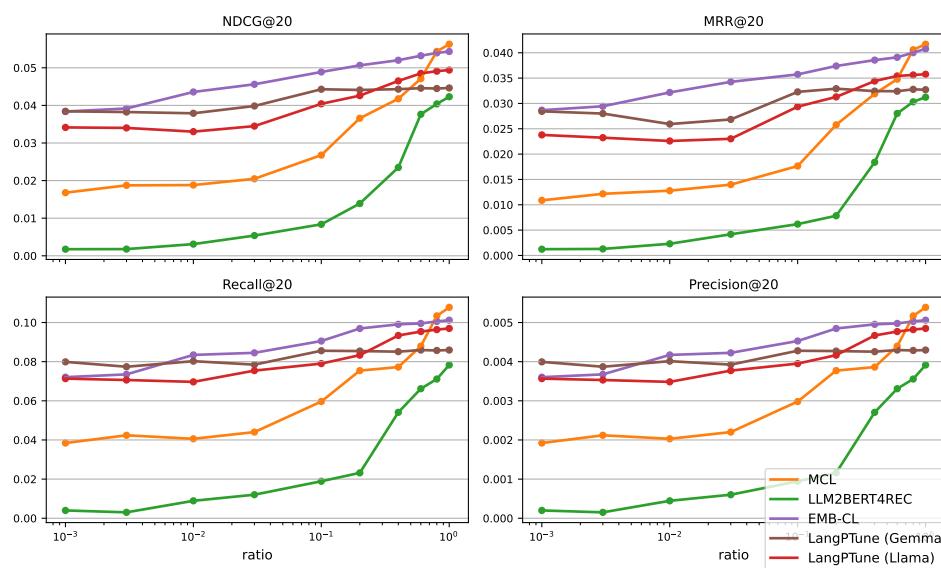
# E. Training Dataset Size Ablation



**Figure 4: Ablation on Training Dataset Ratio over all four metrics**