Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Multi-Stage Alignment of Large Language Models for Popularity Bias Mitigation in Generative Movie Recommendation

Subham Raj

subham_2221cs25@iitp.ac.in

Indian Institute of Technology Patna

Krishnakant Chourey

Indian Institute of Technology Patna

Sriparna Saha

Indian Institute of Technology Patna

Brijraj Singh

Sony Research India

Niranjan Pedanekar

Sony Research India

Research Article

Keywords: Large Language Models, DPO, Biasness, Recommendation

Posted Date: October 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10375648/v2

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Multi-Stage Alignment of Large Language Models for Popularity Bias Mitigation in Generative Movie Recommendation

Subham Raj[1*], Krishnakant Chourey[1], Sriparna Saha[1], Brijraj Singh[2], Niranjan Pedanekar[2]

[1]Indian Institute of Technology Patna, Patna, 801103, Bihar, India.
[2]Sony Research India, Bangalore, 560048, Karnataka, India.

*Corresponding author(s). E-mail(s): subham_2221cs25@iitp.ac.in;
Contributing authors: 2201ph15_krishnakant@iitp.ac.in; sriparna@iitp.ac.in;
brijraj.singh@sony.com; niranjan.pedanekar@sony.com;

Abstract

Large language models (LLMs) exhibit a strong popularity bias in recommendation tasks, often over-recommending mainstream items while under-representing user-aligned but less popular content. To address this limitation, we propose a multi-stage alignment pipeline for debiasing LLM-based movie recommenders, consisting of preference extraction from user histories, supervised fine-tuning (SFT) for behavioral grounding, and Direct Preference Optimization (DPO) for bias-corrective alignment toward less-popular yet topically relevant items. By explicitly separating grounding from bias correction, the framework enables stable and controllable improvements in exposure fairness. In addition to intrinsic alignment, we systematically compare our approach against classical post-hoc debiasing methods, including popularity-aware re-ranking, to examine trade-offs between recommendation quality and exposure balance. Experiments on MovieLens, Netflix, and Flickscore demonstrate that multi-stage alignment achieves substantial gains in novelty and catalog coverage while maintaining competitive accuracy and offering advantages over simpler exposure-adjustment baselines. These results position multi-stage LLM alignment as a principled and scalable strategy for mitigating popularity bias in generative recommender systems. Our code is publicly available at https://github.com/krishnakantx/debiased-llm-recsys.

Keywords: Large Language Models, DPO, Biasness, Recommendation

1 Introduction

The integration of large language models (LLMs) into recommendation systems represents a convergence of two previously distinct paradigms: symbolic generative modeling and latent collaborative filtering. The rapid advancements in LLMs have revolutionized numerous natural language processing (NLP) applications, including dialogue systems, content generation, and personalized recommendations [1–6]. By leveraging their deep contextual understanding and generalization capabilities, LLMs have emerged as powerful

tools for enhancing user experience in recommendation systems. Specifically, in the domain of movie recommendations, LLMs offer the ability to process textual data such as reviews, plot descriptions, and user interactions to generate tailored suggestions [7–9].

However, a critical challenge in deploying LLMs for recommendation tasks is their inherent *popularity bias*, where frequently mentioned or widely recognized content is overrepresented in recommendations, leading to limited exposure to diverse and niche preferences [10–12]. This bias is not new—it has long been studied in traditional recommender systems, where re-ranking strategies, calibrated recommendation loss functions, and diversity constraints have been proposed to mitigate it [13, 14]. The challenge becomes more pronounced in LLMs, which absorb global popularity patterns from massive pre-training corpora, causing them to disproportionately recommend mainstream items regardless of an individual user's taste. While such post-hoc strategies remain computationally efficient and widely adopted, it remains unclear whether intrinsic alignment of LLMs can provide systematic advantages over these classical exposure-adjustment methods in generative recommendation settings.

Existing LLM-based recommendation approaches primarily rely on zero-shot prompting and traditional fine-tuning, both of which have been shown to exhibit substantial popularity bias rather than explicitly mitigate it [15]. While zero-shot prompting enables models to generate recommendations without explicit task-specific adaptation, recommendation systems trained on imbalanced interaction data often overemphasize mainstream content, leading to popularity bias [16]. Prompt formulation itself is a crucial factor influencing generative behavior, and studies have shown that small changes in prompts can yield large shifts in model outputs, including reinforcement of bias [17, 18]. Traditional fine-tuning, which involves adapting LLMs using labeled recommendation datasets, provides some improvements in personalization but often fails to fully mitigate bias, as the model may still retain prior knowledge favoring popular items [19].

Beyond bias, fairness and diversity in recommendation outputs have emerged as critical quality dimensions. Several works have proposed evaluation frameworks and models for balancing relevance, user satisfaction, and content exposure [20, 21]. In the context of LLM-based recommendation, these fairness concerns are compounded by the opacity and memorization tendencies of generative models. In this work, we focus on exposure fairness in the form of popularity bias mitigation and systematically examine both intrinsic alignment and classical post-hoc mitigation strategies.

To address these challenges, we introduce a **multi-stage alignment pipeline** for adapting LLMs to fairer and more personalized movie recommendations. Unlike prior work that relies solely on prompting [22–24] or single-stage supervised fine-tuning [25, 26], our pipeline decouples *behavioral grounding* from *bias-corrective alignment*. The pipeline consists of three stages:

1. **Preference Extraction:** User interaction histories are transformed into structured natural-language preference profiles that normalize and unify input across datasets.
2. **Supervised Fine-Tuning (SFT):** The model is grounded to generate coherent, structured recommendation outputs while learning generic user–item relevance patterns.
3. **Direct Preference Optimization (DPO):** The model is aligned toward recommending *less-popular yet topically relevant* items using contrastive preference pairs that explicitly encode debiasing objectives.

We emphasize that DPO[27] itself is not a novel component of this work; rather, its role is instrumental within a broader alignment strategy. This multi-stage design is essential because the alignment objectives of recommendation grounding and bias correction are not interchangeable. Empirically, we observe that applying DPO without prior behavioral grounding leads to unstable and incoherent recommendation outputs, whereas SFT alone tends to reproduce the popularity distribution implicit in the training data. By explicitly decoupling behavioral grounding (SFT) from bias-corrective alignment (DPO), the proposed pipeline yields a stable, reproducible, and effective workflow for mitigating popularity bias in LLM-based recommender systems.

We conduct extensive experiments on three state-of-the-art LLMs—LLaMA 3.1, LLaMA 2.0, and Mistral 7B—to assess the effectiveness of zero-shot prompting, SFT, and DPO in mitigating

popularity bias while maintaining recommendation quality. Our experimental results demonstrate that the proposed multi-stage alignment strategy effectively reduces reliance on global popularity signals while maintaining competitive recommendation accuracy. In particular, bias-corrective preference optimization plays a complementary role when applied on top of a behaviorally grounded model, enabling more personalized and exposure-balanced recommendations. We further benchmark the intrinsic alignment pipeline against classical popularity-aware re-ranking baselines to evaluate trade-offs between recommendation quality and exposure balance.

Main Contributions.

- *Problem Decomposition for LLM-Based Recommendation Alignment:* We formulate popularity bias mitigation in LLM-based recommender systems as a two-stage alignment problem—behavioral grounding and bias correction—rather than a single fine-tuning objective.
- *Empirical Analysis of Single-Stage Alignment Limitations:* Through systematic comparison of zero-shot prompting, supervised fine-tuning (SFT), and preference optimization, we demonstrate that neither SFT nor DPO alone can simultaneously ensure stable recommendation behavior and effective debiasing, motivating a multi-stage design.
- *Multi-Stage Alignment Strategy for Popularity Bias Mitigation:* We propose and validate a three-stage alignment workflow—preference extraction, SFT, and DPO—that explicitly decouples behavioral grounding from bias-corrective alignment, enabling stable and reproducible debiased recommendation generation.
- *Programmatic Preference Construction for DPO in Recommendation:* We introduce a fully automated, dataset-driven method for constructing contrastive preference pairs for DPO using item popularity statistics and metadata, eliminating the need for human annotation.
- *Cross-Model and Cross-Dataset Validation:* We conduct extensive experiments across three LLM families (LLaMA 2, LLaMA 3.1, and Mistral) and three datasets with distinct popularity distributions, demonstrating the robustness and generalizability of the proposed alignment strategy.
- *Reproducible Framework and Benchmarks:* We release detailed experimental protocols and benchmarks to support transparent and reproducible research on LLM-based recommendation alignment.
- *Intrinsic vs. Extrinsic Fairness Comparison:* We provide a systematic empirical comparison between intrinsic alignment (SFT+DPO) and classical post-hoc re-ranking methods, analyzing trade-offs between exposure fairness, accuracy, and recommendation stability.

Structure of the Paper Section 2 reviews relevant literature on LLM-based recommendation and bias mitigation. Section 3 deals with Motivation, while Section 4 gives information on Problem Formulation. Section 5 describes the datasets used in this work. Section 6 outlines the LLM architectures under study. Section 7 presents our full methodology, including the multi-stage alignment pipeline. Section 8 defines evaluation metrics and discusses our experimental findings on accuracy, novelty, popularity, and diversity, including a comparison against post-hoc re-ranking baselines (Section 8.4) and an analysis of hallucination and catalogue validity (Section 8.6). Section 9 concludes with a discussion of limitations and future work, and Section 10 discusses ethical considerations.

2 Related Work

2.1 Popularity Bias and De-biasing in Classical Recommender Systems

The challenge of popularity bias in recommender systems has been extensively studied across traditional collaborative filtering and content-based approaches. Popularity bias manifests when recommendation algorithms disproportionately favor frequently rated or consumed items, inadvertently suppressing exposure to less popular but potentially more relevant content [10, 28]. This phenomenon creates a feedback loop where popular items gain even more visibility, limiting catalog diversity and user discovery opportunities.

Classical debiasing approaches include inverse propensity weighting techniques [29] and popularity-aware re-ranking strategies that explicitly adjust item exposure during the ranking

phase [13]. Subsequent work in multistakeholder recommendation [30] has highlighted the limitations of global popularity signals and emphasized the need to account for heterogeneous exposure effects across users and items. In addition, causal inference techniques[31] have been explored to model the role of exposure and confounding factors in recommendation outcomes.

Graph-based collaborative filtering methods[32] have also been shown to suffer from popularity bias, as high-degree (popular) items can dominate embedding propagation and influence user representations in interaction graphs . Prior work [33] has observed that such dynamics can exacerbate rich-get-richer effects in representation learning, motivating the need for mechanisms that promote diversity and reduce embedding redundancy. Additionally, existing surveys[34] have highlighted the close relationship between fairness, robustness, and diversity considerations in recommender systems allowing bias mitigation techniques to be analyzed across different recommendation paradigms.

While these approaches have proven effective in classical recommendation settings with explicit candidate sets and scoring functions, their assumptions do not directly translate to fully generative recommendation paradigms based on large language models. These classical post-hoc methods, particularly popularity-aware re-ranking strategies, remain strong and computationally efficient baselines for exposure control and therefore serve as important points of comparison in generative recommendation settings.

2.2 LLMs for Recommendation: Capabilities, Limitations, and Recent Advances

The integration of Large Language Models into recommendation systems represents a paradigm shift from traditional discriminative approaches to generative recommendation frameworks [35]. LLMs offer significant advantages through their natural language understanding capabilities, enabling context-rich recommendations that can handle diverse user preferences and item descriptions without requiring explicit behavioral data. Recent advances have categorized LLM-based recommenders into two main paradigms: Discriminative LLM for Recommendation (DLLM4Rec) and Generative LLM for Recommendation (GLLM4Rec) [15].

However, LLM-based recommenders face several notable limitations. Prior work has revealed that LLMs exhibit inherent instability due to position bias, where the order of presented items can significantly influence recommendation quality [36]. This sensitivity has motivated the use of calibration techniques that employ probabilistic methods to improve the reliability and consistency of LLM outputs [37]. Furthermore, LLMs are less effective at directly modeling sparse identifiers such as user and item IDs compared to traditional collaborative filtering approaches, motivating hybrid architectures that integrate LLM semantic knowledge with collaborative signals [38].

Recent innovations have focused on addressing these limitations through various approaches. Conversational recommendation systems powered by LLMs have shown promise in providing more transparent and explainable recommendations [22, 24], while multi-modal integration allows recommender systems, including recent LLM-based models, to leverage both textual and visual content for enhanced recommendation quality [39]. However, challenges persist in areas such as cold-start scenarios, computational efficiency, and the need for continuous model updates to maintain recommendation relevance [15].

Recent LLM-based recommender systems have explored both prompt-driven and hybrid retrieval-based designs. For example, Chat-Rec [22] investigates conversational and explainable recommendation through interactive prompting, while TALLRec [40] proposes a two-stage framework in which LLMs are aligned using candidate sets generated by traditional recommender components. While these approaches demonstrate the potential of LLMs in recommendation, they primarily operate within candidate-dependent or retrieval-based paradigms. In contrast, our work studies candidate-free generative recommendation, where the language model directly determines item selection from natural-language preference descriptions. We therefore position our approach as complementary to prior LLM-based recommenders, focusing on alignment strategies for mitigating popularity bias in fully generative recommendation settings.

2.3 Transferability and Fairness in Multi-Dataset, Multi-Model Contexts

The evaluation of recommender systems across multiple datasets and model architectures has highlighted challenges in achieving robust generalization under distribution shifts, particularly when data characteristics vary across domains. Prior work on cross-domain and sequential recommendation has shown that traditional approaches often struggle when overlapping users or items are limited, leading to degraded performance under domain shift [41]. Recent multimodal recommendation models have demonstrated that leveraging rich item-side content can improve robustness and representation quality, which is beneficial for generalization across diverse recommendation settings [42].

Fairness considerations in multi-dataset contexts are particularly complex, as different datasets may exhibit varying bias patterns and demographic distributions [43]. Recent work has highlighted the importance of developing fairness metrics that remain meaningful across different evaluation contexts and user populations [44]. The integration of LLMs into multi-domain recommendation scenarios further complicates fairness analysis, as these models may perpetuate biases inherited from large-scale pretraining data, even when applied across diverse datasets and deployment settings [15, 35].

Cross-domain recommendation systems have increasingly adopted federated learning approaches to address privacy concerns while maintaining model performance across distributed environments [45]. However, ensuring fairness in such privacy-preserving settings remains challenging, as federated training must balance recommendation utility, fairness objectives, and data heterogeneity across clients. Recent work has begun to explicitly address these trade-offs by proposing fairness-aware federated recommendation frameworks that calibrate model updates to reduce bias under decentralized training [46].

Despite these advances, existing work on debiasing recommendation systems—both classical and LLM-based—largely focuses on single-stage interventions, such as re-ranking, prompt engineering, or fine-tuning in isolation. Prior LLM-based approaches typically evaluate either zero-shot prompting or supervised fine-tuning, without systematically analyzing how different alignment stages interact or whether bias correction should be decoupled from behavioral grounding. In contrast, our work explicitly formulates popularity bias mitigation in LLM-based recommendation as a multi-stage alignment problem, and empirically demonstrates that separating behavioral grounding (via SFT) from bias-corrective alignment (via preference optimization) is essential for achieving stable and exposure-balanced recommendations across models and datasets.

3 Motivation

Despite extensive research in both classical and LLM-based recommendation systems, several critical gaps remain unaddressed in the current literature. First, there is a lack of systematic evaluation of de-biasing techniques across multiple LLM architectures and datasets simultaneously. While individual studies have examined specific models or datasets, comprehensive comparative analysis across diverse settings remains limited. This gap is particularly pronounced in understanding how different LLM architectures respond to identical de-biasing interventions.

Second, existing research has examined the trade-offs between recommendation accuracy and popularity bias mitigation within specific algorithmic settings [47]. However, systematic analyses comparing intrinsic alignment-based debiasing with classical post-hoc methods across different model architectures and datasets remain limited. The development of unified evaluation frameworks that can assess both recommendation quality and fairness simultaneously across multiple contexts therefore represents a significant research opportunity.

Third, current approaches to popularity bias mitigation in LLM-based systems lack standardized methodologies for prompt engineering and target selection. Although prior studies suggest that prompt modifications can influence bias patterns, their generalizability across domains and user populations remains unclear. Moreover, most existing work evaluates single-domain settings,

leaving multi-domain and cross-domain bias mitigation largely underexplored.

Finally, there is a notable absence of research examining the robustness and generalizability of de-biasing techniques across different cultural and linguistic contexts. Most existing studies are conducted on English-language datasets with Western cultural contexts, limiting the applicability of findings to global recommendation scenarios. This study addresses these gaps by providing a comprehensive, multi-model, multi-dataset evaluation framework that examines the effectiveness and transferability of LLM-based de-biasing techniques across diverse recommendation contexts, including an empirical assessment of where such techniques generalize well and where their benefits are more limited. These gaps collectively motivate the design of our proposed multi-stage alignment pipeline. Specifically, the lack of cross-model analysis calls for a systematic evaluation across multiple LLM families; the absence of joint accuracy–bias analysis necessitates unified evaluation across relevance and exposure metrics; and the limitations of single-stage debiasing approaches motivate an explicit separation between behavioral grounding and bias-corrective alignment. Our work addresses these challenges through a reproducible, multi-dataset study that isolates the roles of preference extraction, supervised fine-tuning, and preference optimization in mitigating popularity bias in LLM-based recommender systems.

4 Problem Formulation

4.1 Recommendation Setting

Let $\mathcal{U}$ denote the set of users and $\mathcal{I}$ denote the set of items (movies). Each user $u \in \mathcal{U}$ is associated with an interaction history $\mathcal{H}_u$, defined as:

$$\mathcal{H}_u = \{(i, r_{u,i}) \mid r_{u,i} \text{ is observed}\}. \tag{1}$$

where $r_{u,i}$ represents an explicit user feedback signal, such as a multi-level rating or a binary like/dislike indicator.

Given a user history $\mathcal{H}_u$, the goal of the recommendation system is to generate a ranked list of previously unseen items:

$$\mathcal{R}_u = [i_1, i_2, \ldots, i_K], \quad i_k \in \mathcal{I} \setminus \mathcal{H}_u, \tag{2}$$

such that the recommended items are relevant to the user's preferences.

4.2 LLM-Based Recommendation Paradigm

In this work, we consider a generative recommendation paradigm based on large language models (LLMs). User interaction histories $\mathcal{H}_u$ are first transformed into natural-language preference descriptions through a deterministic prompt construction function:

$$x_u = f_{\text{prompt}}(\mathcal{H}_u), \tag{3}$$

where x_u denotes the textual prompt provided as input to the LLM.

Given the prompt x_u, the LLM generates a ranked recommendation list in textual form, which is subsequently parsed into a structured list $\mathcal{R}_u$. Unlike traditional collaborative filtering approaches that rely on explicit user and item identifiers and score-based ranking, LLM-based recommenders leverage semantic representations learned during pre-training and alignment to infer user preferences from natural language descriptions.

4.3 Popularity Bias and Exposure Fairness

Popularity bias refers to the tendency of recommender systems to disproportionately recommend items with high global interaction frequency, regardless of individual user relevance. Let $p(i)$ denote the popularity of item i, measured as the total number of interactions it receives in the dataset:

$$p(i) = \sum_{u \in \mathcal{U}} \mathbb{I}[(u, i) \in \mathcal{H}_u]. \tag{4}$$

A recommendation system is said to exhibit popularity bias if the probability of recommending an item increases primarily as a function of $p(i)$ rather than its relevance to the user.

In this work, we focus on *exposure fairness*, where fairness is operationalized as balanced exposure across items with different popularity levels. We do not consider demographic or group-based fairness notions; instead, our objective is to mitigate popularity-driven exposure imbalance while preserving recommendation relevance.

4.4 Popularity-Aware Recommendation Objective

The objective of this work is to generate recommendation lists that jointly satisfy two criteria:

1. High relevance to user preferences.
2. Reduced reliance on global popularity signals.

Formally, we conceptualize the recommendation policy π as optimizing a trade-off between relevance and exposure fairness:

$$\max_{\pi} \mathbb{E}_{u\sim\mathcal{U}}\Big[\mathcal{Q}_{\text{rel}}(\mathcal{R}_u) \;-\; \lambda\,\mathcal{Q}_{\text{pop}}(\mathcal{R}_u)\Big], \qquad (5)$$

where $\mathcal{Q}_{\text{rel}}(\cdot)$ denotes a relevance-oriented quality measure, $\mathcal{Q}_{\text{pop}}(\cdot)$ captures the average popularity of recommended items, and λ controls the trade-off between relevance and exposure fairness.

In practice, this objective is evaluated using standard ranking metrics (e.g., Precision@K, nDCG@K) and exposure-oriented metrics such as novelty, catalog coverage, and average item popularity.

4.5 Scope of the Problem

We study popularity bias mitigation in the context of movie recommendation using explicit user feedback. Our formulation focuses on alignment strategies for large language models and does not introduce new recommendation objectives or loss functions. Instead, we analyze how different alignment stages influence recommendation stability, relevance, and exposure fairness across models and datasets.

5 Datasets

5.1 MovieLens 1M

The MovieLens 1M dataset contains **6,040 users**, **3,952 movies**, and **1,000,209 ratings** collected between 1998-2005 [48]. Each user has rated at least 20 movies on a 5-star scale, resulting in 95.8% sparsity. The dataset encompasses 18 genres and includes comprehensive demographic information, making it a cornerstone benchmark for recommendation systems research. The temporal concentration of ratings (49% of users complete ratings within a single day) reflects retrospective rating behavior patterns.

5.2 Netflix Prize

The Netflix Prize dataset represents a large-scale collection with **480,189 users**, **17,770 movies**, and **100,480,507 ratings** spanning 1998-2005 [49]. With 99.5% sparsity, it exhibits extreme popularity bias where 5% of movies receive over 1,000 ratings while 90% receive only hundreds. The dataset requires external metadata integration for genre information and careful preprocessing to handle its massive scale while maintaining statistical validity for de-biasing research.

5.3 Flickscore

The Flickscore dataset provides unique regional diversity with **919 users**, **2,851 movies**, and **10,000 ratings** across **18 Indian languages** [50]. Unlike traditional 5-star systems, it employs binary ratings (like/dislike), capturing preferences across diverse regional cinema. The dataset includes rich user demographics (age, gender, geographic location, language proficiency) and comprehensive movie metadata, offering insights into multilingual movie consumption patterns and cultural preferences.

5.4 Dataset Comparison

The three datasets provide complementary evaluation contexts spanning different scales, cultural contexts, and rating systems. MovieLens 1M offers a balanced benchmark with moderate scale and comprehensive metadata. Netflix Prize provides large-scale evaluation with extreme popularity bias challenges. Flickscore contributes cultural diversity and multilingual perspectives with its binary rating approach. Together, they enable robust assessment of model generalizability across Western and regional cinema, different rating scales (5-star vs. binary), and varying sparsity levels (95.8% to 99.5%), ensuring comprehensive evaluation of LLM-based recommendation systems across diverse real-world scenarios.

6 Language Model Architectures

This study evaluates three prominent open-source language models, each chosen to represent a different point in the landscape of scale, efficiency,

and generational advancement: LLaMA 3.1 8B, LLaMA 2 7B, and Mistral 7B.

6.1 LLaMA 3.1 8B

LLaMA 3.1 8B[51] represents the latest generation of powerful, general-purpose models from Meta. It comprises 8 billion parameters organized into 32 transformer-decoder layers, utilizing SwiGLU activations for improved non-linearity and Rotary Positional Embeddings (RoPE) for robust handling of long sequences. Its key strength lies in its massive pre-training on a high-quality, diverse corpus of 15 trillion tokens, which includes a rich mix of web text, books, code, and multilingual sources. This extensive training endows the model with a deep understanding of world knowledge and nuanced language, which is crucial for interpreting movie titles and user preferences. We use the instruction-tuned variant, which is already aligned to follow commands, making it a strong baseline for instruction-following behavior. The model is available on Hugging Face: https://huggingface.co/meta-llama/Llama-3.1-8B-Instruct.

6.2 LLaMA 2 7B

As a point of comparison from the previous generation, we include LLaMA 2 7B[52]. This model shares a similar architectural backbone to LLaMA 3.1—with 32 layers and 32 attention heads—but in a more compact 7 billion-parameter configuration. It was pre-trained on 2 trillion tokens from publicly available sources and was one of the foundational models that spurred widespread research and development in the open-source community. Including LLaMA 2 7B allows us to measure the generational leap in performance and assess whether newer pre-training datasets and minor architectural tweaks in LLaMA 3.1 offer tangible benefits for our specific de-biasing task. The official repository can be found at https://huggingface.co/meta-llama/Llama-2-7b-hf.

6.3 Mistral 7B

Mistral 7B[53] is a highly efficient 7.3 billion-parameter model that has demonstrated performance comparable to, or even exceeding, larger models on various benchmarks. Its efficiency stems from key architectural innovations. First, it employs Grouped-Query Attention (GQA), which significantly speeds up inference and reduces the memory requirements during decoding. Second, it uses Sliding Window Attention (SWA), allowing it to handle much longer sequences at a lower computational cost by giving individual layers access to information beyond a fixed window size. These features make it particularly well-suited for recommendation tasks, which can involve long user histories. The model is available at https://huggingface.co/mistralai/Mistral-7B-v0.1.

The selected models allow us to analyze debiasing performance across key axes: LLaMA 3.1 8B serves as the state-of-the-art baseline, LLaMA 2 7B provides a generational comparison, and Mistral 7B evaluates the impact of efficiency-focused architectural innovations. Table 1 summarizes the key specifications of these architectures.

7 Methodology

7.1 Multi-Stage Alignment Pipeline Overview

To better contextualize the subsequent sections, we explicitly define our **three-stage alignment pipeline** for adapting LLMs to personalized, debiased recommendation generation. The stages are:

- **Stage 1: Preference Extraction and Prompt Engineering.** User interaction histories are transformed into structured preference profiles, and natural-language prompts are generated to represent user tastes in a consistent, cross-dataset format.
- **Stage 2: Supervised Fine-Tuning (SFT).** The LLM is fine-tuned on (prompt, target) pairs to learn stable, format-consistent recommendation behavior. SFT[54] provides a grounded reference policy but does not directly address popularity bias.
- **Stage 3: Direct Preference Optimization (DPO).** Using the SFT model as the reference policy, DPO aligns the model's preferences toward low-popularity yet genre-relevant movies by contrasting preferred and rejected recommendation lists.

Only Stage 2 (SFT) and Stage 3 (DPO) modify model parameters, while Stage 1 enables all subsequent training. Zero-shot prompting is included solely as a baseline and does not participate in the alignment pipeline.

Table 1: Comparison of Large Language Model Architectures used in this study.

Model	Params	Context Len.	Training Tokens	Key Features
LLaMA 2	7B	4k	2 Trillion	Standard Transformer
Mistral 7B	7.3B	8k (32k swa)	Undisclosed	GQA, Sliding Window
LLaMA 3.1	8B	128k	15 Trillion	GQA, RoPE Sc., SwiGLU

7.2 Baseline Recommendations: Zero-Shot LLM

As a foundation for our experiments, we employ a zero-shot recommendation strategy using each large language model in its pre-trained, unmodified state. For every test user, we generate a prompt in natural language describing the user's top 10 liked and bottom 5 disliked movies, as determined directly from their explicit historical ratings. The prompt is purposefully open-ended, typically phrased as shown in Figure 1.

No additional instructions are provided regarding popularity or novelty, and no supervised data from the target domain is introduced. The model is thus forced to rely purely on its pretraining data and internal world knowledge. This approach enables us to measure the default recommendation tendencies and any inherent biases present in the LLMs, most notably their tendency to suggest highly popular or mainstream movies due to data exposure during pretraining.

7.3 Stage 2: Supervised Fine-Tuning (SFT)

To move beyond out-of-the-box behavior, we apply supervised fine-tuning to each language model on the task of personalized movie recommendation. Our training data is constructed by carefully curating user profiles, consistent across all datasets: for each user, we select the top 10 highest-rated ("liked") and bottom 5 lowest-rated ("disliked") movies as their profile signature.

For every such profile, we derive a corresponding prompt (in the same natural language format as used for zero-shot evaluation) and an explicit target list. The target is generated by:

- Excluding all movies already rated by the user,
- Restricting the candidate pool to movies from the bottom 50
- Ranking candidate movies by their genre similarity to the user's liked movies (calculated using the Jaccard index on the sets of genres),
- Selecting the top five as the desired recommendations.

Supervised fine-tuning exposes the model to thousands of such (prompt, target) pairs, teaching it to map user-specific profiles to relevant, less-popular but still appropriate movie recommendations. In this paradigm, the model learns to internalize both explicit preference signals and the anti-popularity bias encoded in the training set construction.

Optimization Objective: We fine-tune the model parameters θ by minimizing the standard negative log-likelihood loss on the target tokens given the prompt. Let x be the input prompt and $y = (y_1, y_2, \ldots, y_L)$ be the target recommendation list of length L. The SFT loss is defined as:

$$\mathcal{L}_{\text{SFT}}(\theta) = -\mathbb{E}_{(x,y)\sim\mathcal{D}_{\text{SFT}}} \left[\sum_{t=1}^{L} \log P_\theta(y_t \mid x, y_{<t}) \right] \tag{6}$$

where P_θ is the conditional probability distribution parameterized by the LLM. This objective forces the model to align its generation probability mass with the curated, less-popular target distributions provided in the SFT dataset.

7.4 Stage 3: Direct Preference Optimization (DPO)

To directly mitigate popularity bias in recommendations, we fine-tuned all LLMs with Direct Preference Optimization (DPO) using only the MovieLens 1M, Netflix Prize, and Flickscore datasets. Our DPO approach leveraged explicit preference pairs derived programmatically from each dataset.

Preference Pair Construction: For every user, we first constructed their profile by extracting their 10 most liked and 5 most disliked movies. For every such user prompt, we generated

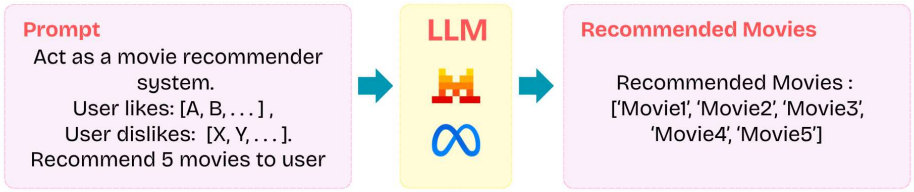


Fig. 1: Example of zero-shot prompt structure.

two valid lists of 5 recommendations, using only unrated movies:

- The **preferred response** comprised 5 movies from the bottom 50% of popularity (by rating count), further filtered by a high degree of genre overlap with the user's liked movies, as measured by the Jaccard index.
- The **rejected response** comprised 5 movies from the top 50% of popularity, but without genre matching.

Both responses were constructed automatically using dataset statistics and metadata; no human judgments or external LLMs were involved. This construction provides a powerful and unambiguous learning signal, simultaneously teaching the model to penalize high popularity and reward topical relevance.

Prompt Design: Each DPO prompt as shown in Figure 2 included the same profile as supervised fine-tuning:

> *Act as a movie recommender system. User likes: [10 liked movies]. User dislikes: [5 disliked movies]. Recommend 5 movies to the user.*

DPO Training Objective: For every prompt, the model was trained using the DPO loss to assign higher log-likelihood to the preferred (low-popularity, genre-matched) responses than to the rejected (popular) responses for the same user and context. The DPO loss function is:

$$\mathcal{L}_{\text{DPO}}(\pi_\theta, \pi_{\text{ref}}) = -\mathbb{E}_{(x,y_w,y_l)\sim\mathcal{D}}\left[\log\sigma\left(\beta\log\frac{\pi_\theta(y_w|x)}{\pi_{\text{ref}}(y_w|x)} - \beta\log\frac{\pi_\theta(y_l|x)}{\pi_{\text{ref}}(y_l|x)}\right)\right] \tag{7}$$

where x is the user prompt, y_w is the preferred recommendation list, y_l is the rejected list, β is a hyperparameter, and π_θ and π_{ref} are the current and reference model policies.

Gradient Analysis and Intuition: To understand how DPO mitigates popularity bias, consider the gradient of the loss function with respect to the model parameters θ:

$$\nabla_\theta\mathcal{L}_{\text{DPO}} = -\beta\mathbb{E}_{(x,y_w,y_l)}\Big[\sigma(\hat{r}_\theta(x,y_l) - \hat{r}_\theta(x,y_w)) \times (\nabla_\theta\log\pi(y_w|x) - \nabla_\theta\log\pi(y_l|x))\Big] \tag{8}$$

where $\hat{r}_\theta$ represents the implicit reward estimate. The term $\nabla_\theta\log\pi(y_l|x)$ corresponds to the gradient direction that increases the likelihood of the rejected (popular) response. The negative sign in front of it ensures that the optimization process *suppresses* the probability of generating high-popularity items, while simultaneously boosting the probability of the preferred (novel) items via

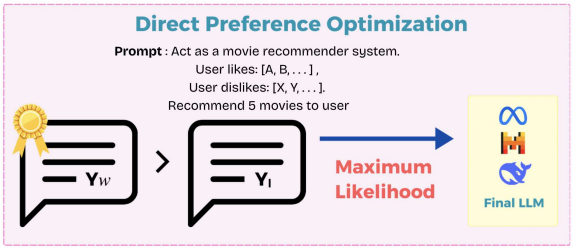


Fig. 2: Example of a DPO preference pair, showing the 'preferred' (low-popularity) and 'rejected' (high-popularity) responses generated for the same user prompt.

$\nabla_\theta \log \pi(y_w|x)$. The weight of this update is controlled by the model's current confidence errors, ensuring efficient sample utilization.

Implementation Notes:

- All candidate lists and user prompts were generated solely from the target dataset's structured data.
- The process was identically applied to all three datasets and all LLMs.

This fully programmatic, dataset-driven DPO setup ensures that the model learns to associate each unique user profile with new, less-popular, taste-matched movies, rather than reproducing global popularity trends. The methodology directly targets our goal of reducing popularity bias while maintaining user relevance in recommendations.

The overall methodology of the entire architecture is shown in Figure 3.

8 Experiments and Results

This study is motivated by two central research questions (RQ):

- **RQ1:** Can fine-tuned LLMs be taught to recommend more novel, less-popular items without losing accuracy?
- **RQ2:** How robust are these de-biasing strategies across various datasets and LLM architectures?
- **RQ3:** Does intrinsic alignment through multi-stage training provide advantages over classical post-hoc debiasing methods (e.g., re-ranking) in achieving balanced exposure while preserving recommendation quality?

8.1 Experimental Setup

8.1.1 Language Model Architectures

We evaluate the proposed alignment strategy using three representative open-source large language models selected to reflect differences in model generation, parameter scale, and architectural design: LLaMA 3.1 8B, LLaMA 2 7B, and Mistral 7B. LLaMA 3.1 represents a recent generation of instruction-tuned models, LLaMA 2 serves as a prior-generation baseline, and Mistral 7B provides an efficiency-oriented architecture with grouped-query and sliding-window attention. The details about these architectures is mentioned in Section 6.

8.1.2 Optimization

We evaluate three alignment settings for LLM-based recommendation: zero-shot prompting, supervised fine-tuning (SFT), and Direct Preference Optimization (DPO).

- **Supervised Fine-Tuning (SFT):** The model is trained using user-item interaction data to learn direct mappings from prompts to desirable recommendation outputs. This method aligns closely with classical supervised learning.
- **Direct Preference Optimization (DPO):** A preference-based fine-tuning strategy that steers the model away from popular item bias by optimizing for implicit reward functions. It relies on pairwise preferences and is well-suited to de-biasing objectives.
- **Popularity-Aware Re-Ranking:** A post-hoc baseline that adjusts the ranked list generated by the model by penalizing high-popularity items while preserving relevance scores.

For DPO training, we use a default preference scaling parameter $\beta = 0.1$ and evaluate sensitivity to $\beta \in \{0.1, 0.5, 1.0, 2.0\}$, observing stable trends across this range. We conduct experiments across three families of LLMs—LLaMA 3.1 8B, LLaMA 2 7B, and Mistral 7B—comparing performance in zero-shot, SFT, and DPO settings. Each model is evaluated across three datasets, namely Movielens-1M, Netflix Prize and Flickscore.

8.1.3 Evaluation Protocol

For each dataset, we maintain a strict separation between the interactions used to construct

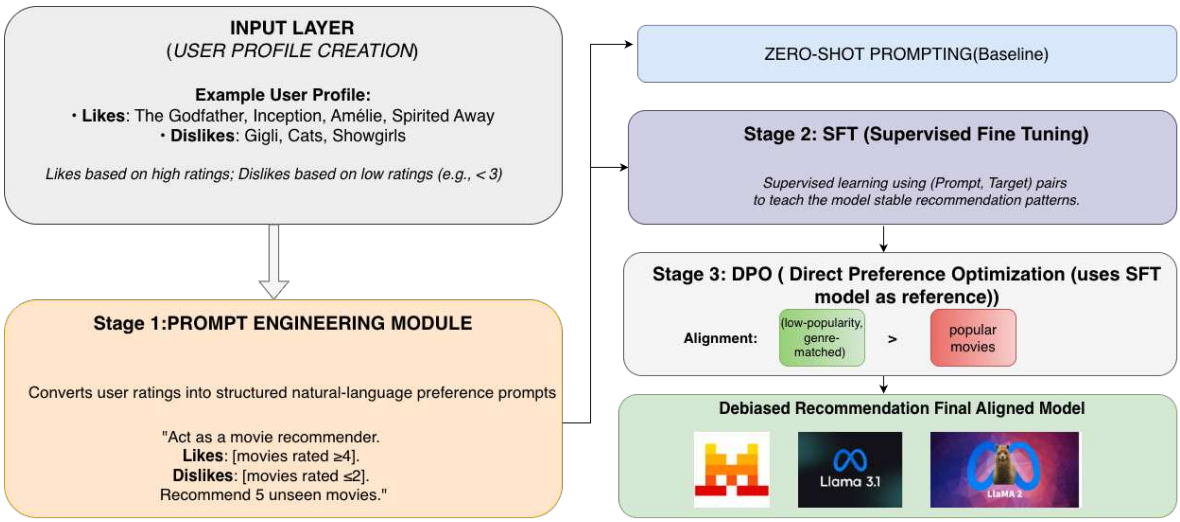


Fig. 3: Overall architecture of the proposed multi-stage alignment pipeline. User profiles are converted into structured prompts (Stage 1), the model is grounded via supervised fine-tuning (Stage 2), and popularity bias is mitigated through DPO alignment (Stage 3). Zero-shot prompting is included as a baseline.

user profiles and SFT/DPO training targets, and the interactions used for evaluation. Each user's rated movies are split chronologically into an 80% training portion and a 20% held-out evaluation portion, with the training portion used to construct the "top 10 liked / bottom 5 disliked" profile and the corresponding SFT/DPO targets. The held-out 20% movies rated later in the user's interaction history serve as ground truth: a recommended item is counted as relevant if it appears in this held-out set, rather than in the synthetically constructed SFT/DPO target list, ensuring that recommendation-quality metrics reflect actual future user preference rather than the debiasing construction used during training.

8.2 Reproducibility

We use the implementation from the open-source repository https://github.com/krishnakantx/debiased-llm-recsys. To ensure reproducibility, we set random seeds for all non-deterministic operations (data splitting, model initialization, and sampling) to fixed values: 42 for NumPy and PyTorch, and 123 for CUDA operations. Experiments were conducted on a single NVIDIA A100 GPU (80GB VRAM) environment. We used the AdamW optimizer with $\beta_1 = 0.9, \beta_2 = 0.999$, and a cosine learning rate schedule with a warmup ratio of 0.03.

For SFT, we fine-tuned for 3 epochs with a learning rate of 2×10^{-5} and a batch size of 16. For DPO, we used a learning rate of 5×10^{-6}, a batch size of 8, and a preference scaling parameter $\beta = 0.1$. To facilitate easy reproduction, we provide a Docker container with all dependencies pre-installed, along with scripts to download and preprocess the datasets automatically. Model checkpoints and training logs are available in the repository.

All experiments were repeated for **five independent runs** using different random seeds, and we report mean metrics. Total GPU-hour usage for each model–dataset configuration is reported in the repository to contextualize computational efficiency.

8.2.1 Evaluation Metrics

We assess both recommendation quality and debiasing performance using the following metrics:

- **Recommendation(Rank-based) Metrics[55]:** *Precision@5*, *nDCG@5*[56], and *MAP@5* are used to evaluate accuracy and relevance.
- **Bias Mitigation Metrics:**

– *Novelty[57]:* Measures how uncommon the recommended items are.
– *Coverage (%)[58]:* The proportion of unique items recommended across users.
– *Average Popularity[13]:* The mean popularity rank of the recommended items. Lower values indicate reduced popularity bias.

Novelty is computed as the average self-information of recommended items:

$$\text{Novelty@}K = \frac{1}{K} \sum_{i \in \mathcal{R}_u} -\log_2 \left(\frac{\text{pop}(i)}{|\mathcal{U}|} \right), \quad (9)$$

where $\text{pop}(i)$ denotes the number of users who interacted with item i.

Coverage measures the proportion of unique items recommended across all users:

$$\text{Coverage@}K = \frac{|\bigcup_{u \in \mathcal{U}} \mathcal{R}_u|}{|\mathcal{I}|}. \quad (10)$$

These metrics jointly reveal how well a method balances relevance and diversity, crucial for escaping popularity bias while preserving user satisfaction.

8.2.2 Scope of Baseline Comparisons

This work focuses on aligning large language models for *candidate-free generative recommendation*, where recommendations are produced directly in natural language rather than retrieved or ranked from a predefined candidate set. In this setting, the language model implicitly performs both item selection and ranking through text generation, without access to explicit relevance scores or item identifiers.

As a result, direct quantitative comparison with classical collaborative filtering and sequential recommendation models—such as NGCF [59], LightGCN [32], and SASRec [60]—is not methodologically well-aligned. These models optimize ranking objectives over explicit candidate pools using well-defined scoring functions, whereas our framework studies recommendation generation in a fully generative paradigm.

To enable a principled comparison between intrinsic and extrinsic debiasing strategies (RQ3), we include a popularity-aware re-ranking baseline as a post-hoc method. While classical re-ranking techniques were originally designed for score-based recommenders, we adapt them to the generative setting by applying exposure-aware adjustments to the ranked outputs produced by the LLM. Specifically, after the model generates a recommendation list, we re-rank the items using popularity-based penalties while preserving the original semantic ordering as much as possible.

This design allows us to systematically evaluate whether embedding fairness objectives directly into the model parameters through multi-stage alignment (SFT+DPO) provides advantages over computationally simpler post-processing approaches that modify only the final recommendation list. We therefore position our intrinsic alignment strategy as complementary to classical debiasing methods, while empirically comparing their effectiveness under identical evaluation metrics.

8.2.3 Re-Ranking Baseline Implementation

To construct the popularity-aware re-ranking baseline, we first prompt the zero-shot LLM to generate an extended candidate pool of 15 movies per user (rather than the 5 movies produced in the Zero-shot, SFT, and DPO settings), using the same profile-based prompting strategy described in Section 7. Each candidate i in this pool is assigned a relevance score $\text{rank_score}(i)$ based on its position in the LLM's generated ordering, with the first-generated item receiving the highest score and the fifteenth-generated item the lowest.

We then re-rank this candidate pool using a linear popularity penalty:

$$\text{score}_{\text{final}}(i) = \text{rank_score}(i) - \lambda \cdot \frac{p(i)}{|\mathcal{U}|}, \quad (11)$$

where $p(i)$ is the item popularity defined in Equation 4 (Section 4), normalized by the total number of users $|\mathcal{U}|$—the same normalization used in the Novelty metric (Equation 9)—and λ is a penalty weight controlling the strength of the popularity adjustment; $\lambda = 0$ recovers the original LLM ranking, while larger values of λ increasingly favor less popular candidates over the LLM's original relevance ordering. We set $\lambda = 0.5$, selected via tuning. The top 5 items by $\text{score}_{\text{final}}$, after re-sorting the 15-item candidate pool, constitute

the final re-ranking baseline recommendation list reported in Tables 5, 6, and 7.

8.3 Performance Comparison (RQ1, RQ2)

We present our empirical findings in Tables 2, 3, and 4, each corresponding to one dataset. Our analysis reveals several key trends across settings.

MovieLens 1M.

This dataset establishes a strong baseline due to its relatively balanced structure. Zero-shot models, especially Mistral 7B, perform well on standard accuracy metrics, achieving a *Precision@5* of **0.108**. However, they show substantial popularity bias, recommending mostly mainstream content. Fine-tuning, especially with **DPO**, substantially improves bias metrics for LLaMA 3.1 8B—pushing *Novelty* from 13.4 to **17.3**, *Coverage* from 11% to **21%**, and roughly halving *Average Popularity* from 1217 to **575**—with only minimal drops in accuracy; Mistral 7B shows a similar pattern, reaching a slightly higher *Novelty* of **17.5** and *Coverage* of **22%**.

Netflix Prize.

As a sparse and long-tailed dataset, Netflix Prize presents a tougher test. While Zero-shot models again top raw accuracy, they recommend from a narrow band of popular items. DPO demonstrates its robustness here: across the three model families, *Novelty* increases by up to **175%** (LLaMA 3.1 8B) and *Coverage* improves by up to **280%** (Mistral 7B) over the zero-shot baseline. Mistral 7B with DPO, for instance, sees *Novelty* jump from 3.3 to **8.9**, and *Average Popularity* drop from 1615 to **590**. These results underline DPO's effectiveness even under minimal signal conditions.

Flickscore.

Flickscore tests generalizability across cultures and sparse labels. Here, **SFT** shows strong accuracy and de-biasing potential—achieving the best *Precision@5* (**0.123**) with Mistral 7B and lowest *Average Popularity* (**215**) with LLaMA 3.1 8B. DPO's advantage on this dataset is model-dependent rather than universal: it achieves the best overall *Novelty* (**26.5**) and *Coverage* (**42%**) with Mistral 7B, and improves *Coverage* for LLaMA 2 7B (33% to **39%**), but for LLaMA 3.1 8B its *Coverage* (37%) falls slightly below SFT (38%), and for LLaMA 2 7B its *Novelty* (24.4) falls slightly below SFT (24.9), suggesting SFT alone is already a strong and sometimes competitive debiasing signal in this lower-resource, culturally diverse setting. Across all datasets and models, the proposed multi-stage alignment strategy consistently yields substantial improvements in exposure-oriented metrics while maintaining competitive recommendation accuracy. In particular, bias-corrective preference optimization plays a complementary role when applied on top of a behaviorally grounded model, leading to significant gains in novelty and catalog coverage in most, though not all, model-dataset combinations. These improvements support **RQ1**, demonstrating that LLMs can recommend less-popular items without substantial accuracy loss. Moreover, the consistent trends observed across three model families and datasets of varying structure validate **RQ2**, confirming the robustness and generalizability of the proposed alignment approach, while also revealing that the relative strength of DPO versus SFT is not uniformly architecture-independent.

8.4 Intrinsic Alignment vs. Post-Hoc Debiasing (RQ3)

To address **RQ3**, we directly compare our intrinsic multi-stage alignment strategy (SFT+DPO) against a classical popularity-aware post-hoc re-ranking baseline across all three datasets and model families. The re-ranking method applies a popularity penalty to generated recommendation lists without modifying model parameters, following prior exposure-adjustment strategies in classical recommender systems [13, 47], whereas SFT+DPO embeds exposure objectives directly into the model through preference optimization. Tables 5, 6, and 7 summarize the comparative results in terms of Precision@5, Novelty, and Average Popularity.

MovieLens 1M.

On MovieLens 1M, both approaches reduce popularity bias relative to zero-shot baselines, but intrinsic alignment achieves stronger improvements in exposure metrics across all three model families. With LLaMA 3.1 8B, SFT+DPO attains

Table 2: Results on MovieLens 1M. Subscripts denote std. dev. across 5 seeds.

Metric	LLaMA 3.1 8B			LLaMA 2 7B			Mistral 7B		
	Zero-shot	*SFT*	*DPO*	*Zero-shot*	*SFT*	*DPO*	*Zero-shot*	*SFT*	*DPO*
Bias Metrics									
Novelty	$13.4_{0.2}$	$15.1_{0.3}$	$\mathbf{17.3}_{0.3}$	$13.2_{0.2}$	$14.7_{0.3}$	$\mathbf{16.6}_{0.3}$	$13.6_{0.2}$	$15.9_{0.3}$	$\mathbf{17.5}_{0.3}$
Coverage (%)	$11_{0.5}$	$14_{0.6}$	$\mathbf{21}_{0.7}$	$10_{0.5}$	$13_{0.6}$	$\mathbf{19}_{0.7}$	$12_{0.5}$	$15_{0.6}$	$\mathbf{22}_{0.7}$
Avg. Pop.	1217_{18}	860_{15}	$\mathbf{575}_{12}$	1240_{18}	915_{15}	$\mathbf{600}_{12}$	1217_{18}	840_{15}	$\mathbf{580}_{12}$
Recommendation Metrics									
Precision@5	$0.105_{.003}$	$0.107_{.003}$	$0.102_{.003}$	$0.101_{.003}$	$0.100_{.003}$	$0.098_{.003}$	$0.108_{.003}$	$0.104_{.003}$	$0.103_{.003}$
nDCG@5	$0.392_{.004}$	$0.390_{.004}$	$0.388_{.004}$	$0.385_{.004}$	$0.386_{.004}$	$0.382_{.004}$	$0.396_{.004}$	$0.393_{.004}$	$0.392_{.004}$
MAP@5	$0.170_{.002}$	$0.169_{.002}$	$0.168_{.002}$	$0.167_{.002}$	$0.168_{.002}$	$0.164_{.002}$	$0.172_{.002}$	$0.171_{.002}$	$0.170_{.002}$

Table 3: Results on Netflix Prize. Subscripts denote std. dev. across 5 seeds.

Metric	LLaMA 3.1 8B			LLaMA 2 7B			Mistral 7B		
	Zero-shot	*SFT*	*DPO*	*Zero-shot*	*SFT*	*DPO*	*Zero-shot*	*SFT*	*DPO*
Bias Metrics									
Novelty	$3.2_{0.3}$	$6.3_{0.4}$	$\mathbf{8.8}_{0.4}$	$3.1_{0.3}$	$5.9_{0.4}$	$\mathbf{8.1}_{0.4}$	$3.3_{0.3}$	$6.5_{0.4}$	$\mathbf{8.9}_{0.4}$
Coverage (%)	$5_{0.6}$	$11_{0.8}$	$\mathbf{17}_{0.9}$	$4_{0.6}$	$10_{0.8}$	$\mathbf{15}_{0.9}$	$5_{0.6}$	$12_{0.8}$	$\mathbf{19}_{0.9}$
Avg. Pop.	1631_{26}	970_{20}	$\mathbf{610}_{16}$	1580_{26}	1040_{20}	$\mathbf{635}_{16}$	1615_{26}	980_{20}	$\mathbf{590}_{16}$
Recommendation Metrics									
Precision@5	$0.072_{.004}$	$0.070_{.004}$	$0.069_{.004}$	$0.070_{.004}$	$0.068_{.004}$	$0.067_{.004}$	$0.073_{.004}$	$0.071_{.004}$	$0.070_{.004}$
nDCG@5	$0.210_{.005}$	$0.207_{.005}$	$0.205_{.005}$	$0.203_{.005}$	$0.204_{.005}$	$0.201_{.005}$	$0.209_{.005}$	$0.206_{.005}$	$0.205_{.005}$
MAP@5	$0.089_{.003}$	$0.088_{.003}$	$0.087_{.003}$	$0.087_{.003}$	$0.086_{.003}$	$0.085_{.003}$	$0.090_{.003}$	$0.088_{.003}$	$0.087_{.003}$

Table 4: Results on Flickscore. Subscripts denote std. dev. across 5 seeds.

Metric	LLaMA 3.1 8B			LLaMA 2 7B			Mistral 7B		
	Zero-shot	*SFT*	*DPO*	*Zero-shot*	*SFT*	*DPO*	*Zero-shot*	*SFT*	*DPO*
Bias Metrics									
Novelty	$22.0_{0.3}$	$24.6_{0.4}$	$\mathbf{26.1}_{0.4}$	$21.2_{0.3}$	$\mathbf{24.9}_{0.4}$	$24.4_{0.4}$	$21.5_{0.3}$	$25.0_{0.4}$	$\mathbf{26.5}_{0.4}$
Coverage (%)	$30_{0.7}$	$\mathbf{38}_{0.8}$	$37_{0.8}$	$27_{0.7}$	$33_{0.8}$	$\mathbf{39}_{0.9}$	$31_{0.7}$	$37_{0.8}$	$\mathbf{42}_{0.9}$
Avg. Pop.	410_{6}	$\mathbf{215}_{4}$	218_{4}	432_{6}	265_{5}	$\mathbf{260}_{5}$	408_{6}	204_{4}	$\mathbf{198}_{4}$
Recommendation Metrics									
Precision@5	$0.119_{.004}$	$0.121_{.004}$	$0.117_{.004}$	$0.114_{.004}$	$0.112_{.004}$	$0.113_{.004}$	$0.121_{.004}$	$0.123_{.004}$	$0.118_{.004}$
nDCG@5	$0.415_{.005}$	$0.419_{.005}$	$0.413_{.005}$	$0.409_{.005}$	$0.405_{.005}$	$0.407_{.005}$	$0.422_{.005}$	$0.419_{.005}$	$0.416_{.005}$
MAP@5	$0.174_{.003}$	$0.178_{.003}$	$0.171_{.003}$	$0.166_{.003}$	$0.162_{.003}$	$0.165_{.003}$	$0.180_{.003}$	$0.177_{.003}$	$0.174_{.003}$

higher Novelty (17.3 vs. 15.8) and substantially lower Average Popularity (575 vs. 780) than re-ranking, while also preserving higher Precision@5 (0.102 vs. 0.097). The same ordering holds for LLaMA 2 7B (Novelty 16.6 vs. 15.2; Average Popularity 600 vs. 820) and Mistral 7B (Novelty 17.5 vs. 16.1; Average Popularity 580 vs. 760), indicating that embedding debiasing objectives during training produces more effective exposure control than adjusting rankings post hoc on this dataset.

Netflix Prize.

The long-tailed and highly sparse Netflix dataset provides a more challenging evaluation setting, and the gap between intrinsic and extrinsic methods becomes more pronounced here. Re-ranking improves novelty relative to unaligned models but remains limited compared to SFT+DPO across all three architectures. For instance, with Mistral 7B, intrinsic alignment increases Novelty from 7.5 to 8.9 and reduces Average Popularity from 880 to 590, while also maintaining higher Precision@5 (0.070 vs. 0.067); LLaMA 3.1 8B and LLaMA 2 7B show the same pattern (Novelty 8.8 vs. 7.2 and 8.1 vs. 6.8, respectively). This consistency across all three LLM families suggests SFT+DPO offers better robustness under extreme popularity imbalance than post-hoc re-ranking.

Flickscore.

Flickscore evaluates cross-cultural generalization under binary feedback and multilingual content. In this setting, re-ranking provides moderate improvements in novelty, but intrinsic alignment again achieves stronger exposure balance across all three model families. LLaMA 3.1 8B with SFT+DPO achieves higher Novelty (26.1 vs. 24.0) and lower Average Popularity (218 vs. 260) than re-ranking, while maintaining higher Precision@5; LLaMA 2 7B (Novelty 24.4 vs. 23.5; Average Popularity 260 vs. 280) and Mistral 7B (Novelty 26.5 vs. 24.6; Average Popularity 198 vs. 245) follow the same pattern. This suggests intrinsic alignment generalizes effectively even in culturally diverse and lower-resource contexts.

Overall Trade-off Analysis.

Across all datasets and models examined, re-ranking improves exposure metrics relative to the zero-shot baseline but falls short of SFT+DPO on both Novelty and Average Popularity, and does so alongside a larger drop in Precision@5 in every case, a trade-off widely observed in classical popularity-control strategies [47]. This pattern is consistent with the interpretation that intrinsic alignment modifies the model's internal preference structure rather than applying an external correction to final outputs, resulting in more stable relevance–fairness trade-offs.

Conclusion for RQ3. The empirical evidence indicates that intrinsic multi-stage alignment offers consistent advantages over classical post-hoc debiasing methods in candidate-free generative recommendation settings, holding across all nine model–dataset combinations examined on Precision@5, Novelty, and Average Popularity. By embedding exposure objectives directly into model parameters, SFT+DPO achieves stronger popularity mitigation and more stable accuracy preservation than re-ranking across datasets and architectures.

8.5 Qualitative Analysis

Table 8 presents a qualitative comparison of recommendation lists generated by zero-shot prompting, supervised fine-tuning (SFT), and Direct Preference Optimization (DPO) for a representative user from the MovieLens 1M dataset. While zero-shot recommendations are dominated by globally popular movies, SFT introduces more personalized yet still moderately popular items. In contrast, DPO generates recommendations that are less mainstream while remaining thematically aligned with the user's preferences, illustrating how bias-corrective preference optimization, when applied on top of a behaviorally grounded model, mitigates popularity bias at the individual level.

To further quantify the effect of our debiasing interventions, we first examine the trade-off between recommendation accuracy and popularity bias in Figure 4a, followed by relative percentage gains shown in Figures 4b, 4c, and 4d achieved by DPO over the zero-shot baseline, using the strongest performing model *Mistral-7B* for each metric as a representative example. The key highlights are:

- **MovieLens 1M:** DPO drives a **29%** increase in *Novelty* and a remarkable **83%** increase in catalog *Coverage*, while slashing *Average Popularity* by **53%**.

Table 5: RQ3 Comparison on MovieLens 1M: Intrinsic Alignment (SFT+DPO) vs Re-Ranking. ± denotes std. dev. across 5 seeds. All SFT+DPO improvements are statistically significant at p ¡ 0.05 (single paired t-test).

Model	Method	Prec@5	Novelty	AvgPop
LLaMA 3.1	Re-ranking	0.097 ± 0.003	15.8 ± 0.2	780 ± 15
	SFT+DPO	**0.102 ± 0.003**	**17.3 ± 0.2**	**575 ± 12**
LLaMA 2	Re-ranking	0.094 ± 0.003	15.2 ± 0.2	820 ± 15
	SFT+DPO	**0.098 ± 0.003**	**16.6 ± 0.2**	**600 ± 12**
Mistral 7B	Re-ranking	0.098 ± 0.003	16.1 ± 0.2	760 ± 15
	SFT+DPO	**0.103 ± 0.003**	**17.5 ± 0.2**	**580 ± 12**

Table 6: RQ3 Comparison on Netflix Prize: Intrinsic Alignment (SFT+DPO) vs Re-Ranking. ± denotes std. dev. across 5 seeds. All SFT+DPO improvements are statistically significant at p ¡ 0.05 (single paired t-test).

Model	Method	Prec@5	Novelty	AvgPop
LLaMA 3.1	Re-ranking	0.066 ± 0.004	7.2 ± 0.3	900 ± 20
	SFT+DPO	**0.069 ± 0.004**	**8.8 ± 0.4**	**610 ± 16**
LLaMA 2	Re-ranking	0.064 ± 0.004	6.8 ± 0.3	950 ± 20
	SFT+DPO	**0.067 ± 0.004**	**8.1 ± 0.4**	**635 ± 16**
Mistral 7B	Re-ranking	0.067 ± 0.004	7.5 ± 0.3	880 ± 20
	SFT+DPO	**0.070 ± 0.004**	**8.9 ± 0.4**	**590 ± 16**

- **Netflix Prize:** The gains are most pronounced in this sparse dataset. *Novelty* soars by nearly **170%**, and *Coverage* explodes by a staggering **280%**, demonstrating a profound shift away from a small set of mega-hits. *Average Popularity* is reduced by **63%**.
- **Flickscore:** Even on this regionally-focused dataset, DPO delivers strong results, improving *Novelty* by **23%**, boosting *Coverage* by **35%**, and cutting *Average Popularity* by over half with a **51%** reduction.

Table 7: RQ3 Comparison on Flickscore: Intrinsic Alignment (SFT+DPO) vs Re-Ranking. ± denotes std. dev. across 5 seeds. All SFT+DPO improvements are statistically significant at p ¡ 0.05 (single paired t-test).

Model	Method	Prec@5	Novelty	AvgPop
LLaMA 3.1	Re-ranking	0.112 ± 0.004	24.0 ± 0.4	260 ± 5
	SFT+DPO	**0.117 ± 0.004**	**26.1 ± 0.4**	**218 ± 4**
LLaMA 2	Re-ranking	0.108 ± 0.004	23.5 ± 0.4	280 ± 5
	SFT+DPO	**0.113 ± 0.004**	**24.4 ± 0.4**	**260 ± 5**
Mistral 7B	Re-ranking	0.115 ± 0.004	24.6 ± 0.4	245 ± 4
	SFT+DPO	**0.118 ± 0.004**	**26.5 ± 0.4**	**198 ± 4**

Table 8: Example User Profile and Recommendations (MovieLens 1M, LLaMA 3.1 8B)

User Profile	Likes: The Godfather, Inception, Amelie,...
	Dislikes: Gigli, Cats, Showgirls,...
Zero-shot	Titanic, The Dark Knight, Forrest Gump, Avatar, Star Wars
SFT	Magnolia, Donnie Darko, All About Lily Chou-Chou, The Lives of Others, Oldboy
DPO	Yi Yi, A Separation, Metropolis, Timbuktu, Departures

8.6 Hallucination and Catalogue Validity

A key risk of candidate-free generative recommendation is that the model may produce movie titles that do not exist in the target dataset's catalogue, particularly for niche or ambiguous user profiles. To quantify this risk, we measure the hallucination rate: the percentage of recommended titles, across all users and five random seeds, that do not match any entry in the corresponding dataset's movie catalogue. Hallucination rates were computed over the full test set for each dataset, using string matching against each dataset's catalogue metadata. Table 9 reports these rates for each model, dataset, and alignment stage.

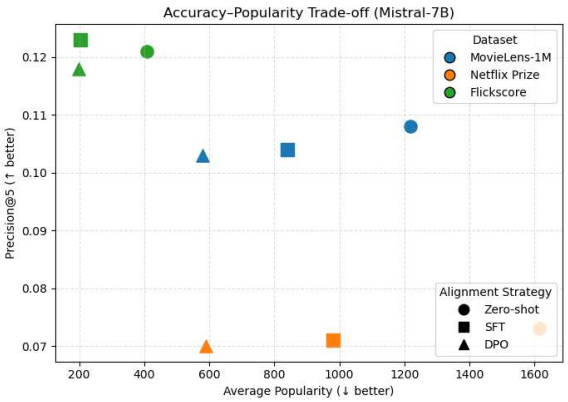
Across all nine model–dataset combinations, hallucination rate decreases monotonically from Zero-shot to SFT to DPO. This trend is consistent with our pipeline design: SFT grounds the model in dataset-derived (prompt, target) pairs drawn directly from the catalogue, which substantially reduces invented titles relative to unconstrained zero-shot generation, while DPO's contrastive preference pairs are likewise constructed exclusively from real catalogue items, providing an additional, if generally smaller, reduction on top of SFT. The improvement from SFT to DPO is modest in most cases (e.g., Netflix Prize, LLaMA 2 7B: 5.0% to 4.9%), indicating that behavioral grounding via SFT accounts for the majority of the reduction in hallucinated content, while DPO's primary role remains bias correction rather than factuality improvement.

Comparing across datasets, Netflix Prize and Flickscore show higher hallucination rates than MovieLens 1M across all three models and alignment stages (Table 9). This pattern may be related to differences in catalogue scale and composition across the three datasets (Section 5); Netflix Prize's large, long-tailed catalogue and Flickscore's smaller, regionally and linguistically specific catalogue (Sections 5.2, 5.3) both differ substantially from MovieLens 1M in ways that could plausibly affect how reliably the model reproduces exact catalogue titles. We do not verify this explanation directly in this work, and leave a more targeted analysis of the relationship between catalogue characteristics and hallucination rate to future work.

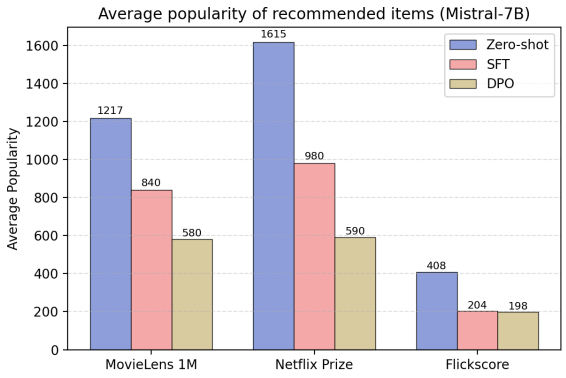
These results indicate that while our multi-stage alignment pipeline meaningfully reduces hallucinated recommendations as a byproduct of grounding the model in real catalogue data, it does not eliminate the issue entirely, particularly on smaller or lower-resource catalogues. As noted in Section 10, the DPO stage does not explicitly penalize hallucinated content, and incorporating an explicit factuality constraint or retrieval-based verification step remains an important direction for reducing this residual rate further.

9 Conclusion and Future Work

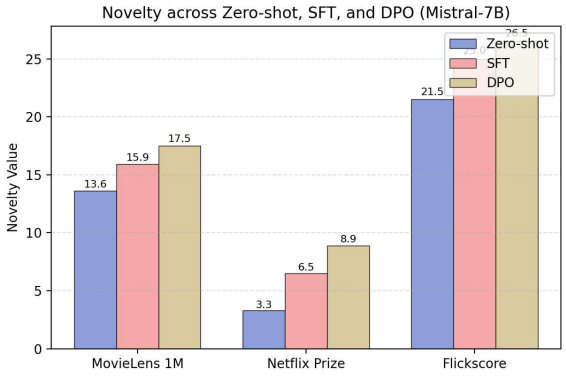
Large language models (LLMs) offer significant potential for recommender systems by enabling context-aware and highly personalized recommendation through natural language understanding.



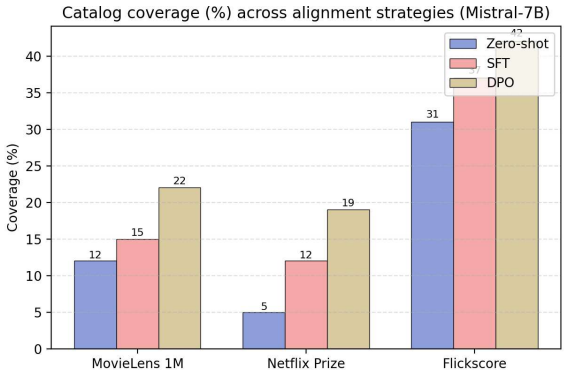
(a) Accuracy–popularity trade-off across alignment strategies using Mistral-7B as a representative model. Marker shapes denote alignment strategies (Zero-shot, SFT, DPO), while colors correspond to datasets. DPO substantially reduces popularity bias while maintaining competitive Precision@5 across datasets.



(b) Average popularity of recommended items using Mistral-7B. Lower values indicate stronger reduction in popularity bias.



(c) Novelty across Zero-shot, SFT, and DPO using Mistral-7B. Higher values indicate more diverse recommendations.



(d) Catalog coverage (%) across alignment strategies using Mistral-7B. Higher values indicate broader item exposure.

Fig. 4: Quantitative analysis of popularity bias mitigation and exposure fairness. The top row illustrates the accuracy–popularity trade-off and overall popularity reduction, while the bottom row highlights gains in novelty and catalog coverage achieved through preference-based alignment.

However, their reliance on large-scale pretraining data introduces a strong tendency to overemphasize globally popular items, limiting exposure to niche and user-aligned content. In this work, we proposed a principled multi-stage alignment framework that explicitly separates behavioral grounding from bias-corrective alignment, enabling effective mitigation of popularity bias in LLM-based movie recommendation.

Through extensive experiments across multiple datasets and model families, we demonstrated that the proposed alignment strategy substantially improves exposure fairness—reducing average item popularity and increasing novelty and catalog coverage—while maintaining competitive recommendation accuracy. Our analysis shows that supervised fine-tuning provides stable and coherent recommendation behavior, while

Table 9: Hallucination rate (% of recommended titles not present in the dataset catalogue) by model, dataset, and alignment stage, averaged across users and seeds.

Dataset	Model	Zero-shot	SFT	DPO
MovieLens 1M	LLaMA 3.1 8B	5.1%	3.1%	2.0%
	LLaMA 2 7B	7.2%	5.4%	3.3%
	Mistral 7B	4.6%	3.0%	2.0%
Netflix Prize	LLaMA 3.1 8B	7.0%	4.5%	4.0%
	LLaMA 2 7B	7.5%	5.0%	4.9%
	Mistral 7B	7.0%	4.6%	3.9%
Flickscore	LLaMA 3.1 8B	8.0%	7.5%	6.8%
	LLaMA 2 7B	9.0%	8.0%	7.6%
	Mistral 7B	8.4%	8.1%	7.4%

preference-based alignment plays a complementary role in steering models away from popularity-driven outputs. In resource-constrained or region-specific settings, such as the Flickscore dataset, supervised fine-tuning alone emerges as a strong and efficient baseline, highlighting the adaptability of alignment strategies across data regimes.

An important limitation of this work is that we do not explicitly address extreme cold-start scenarios in which no prior user interaction history is available. While LLM-based recommenders can leverage semantic priors and minimal textual cues to generate reasonable recommendations under sparse preference conditions, incorporating dedicated cold-start mechanisms—such as preference elicitation strategies or adaptive user modeling approaches explored in systems like ALDI and ColdLLM—remains an important direction for future work. Integrating such mechanisms with the proposed alignment framework could enable more robust personalization from minimal or evolving user input.

Future work also includes deploying aligned LLM-based recommenders in real-world, user-facing environments to study their impact on engagement, satisfaction, and long-term discovery. Another promising direction is the development of adaptive or hybrid alignment strategies that dynamically balance behavioral grounding and bias correction based on data availability, user maturity, and application constraints. Additionally, extending the framework to jointly address other exposure-related biases, such as those associated with region or language, and to incorporate natural language explanations that improve transparency and user trust represents a valuable avenue for further research.

10 Ethical Considerations

Deploying LLMs for recommendation introduces ethical considerations beyond popularity bias.

Hallucinatory Recommendations: Generative models may occasionally produce non-existent or factually incorrect movie titles, particularly for niche or ambiguous queries. As reported in Section 8.6 and Table 9, we measure this directly across all models, datasets, and alignment stages. Hallucination rate decreases monotonically from Zero-shot to SFT to DPO in every case, confirming that supervised fine-tuning is the primary driver of this reduction, since it directly grounds outputs in dataset-derived targets; the DPO stage does not explicitly penalize hallucinated content and typically yields only a marginal additional reduction on top of SFT. Rates remain notably higher on Flickscore than on MovieLens 1M or Netflix Prize across all three stages, likely reflecting its smaller, lower-resource, and linguistically diverse catalogue. Incorporating explicit factuality constraints or retrieval verification mechanisms could improve reliability further, particularly for lower-resource catalogues.

Stereotype Reinforcement: While our framework reduces popularity bias, LLMs may

still reflect latent stereotypes inherited from large-scale pretraining data (e.g., gendered or culturally biased genre associations). Our alignment strategy focuses on exposure fairness and does not explicitly address intersectional or demographic bias mitigation.

Computational Cost: The computational requirements of multi-stage fine-tuning (SFT + DPO) are substantially higher than traditional ID-based collaborative filtering models. Although we report GPU usage for transparency, the resource intensity may limit accessibility. Parameter-efficient approaches (e.g., LoRA-based DPO or adapter tuning) could reduce the cost of intrinsic alignment and improve practical deployability.

11 Declarations

Availability of data and material. The datasets used in this study are publicly available. MovieLens 1M is available from the GroupLens Research repository, the Netflix Prize dataset is available from the original Netflix Prize release, and Flickscore is available as described in the original dataset publication.

Funding. No funding was received to assist with the preparation of this manuscript.

Author contributions. S.R. conceptualized the multi-stage alignment framework, designed the SFT and DPO training pipelines, conducted the experiments, and wrote the main manuscript text. K.C. contributed to the implementation of the training and evaluation pipeline, conducted experiments across datasets and models, and maintained the code repository. S.S. supervised the research direction, provided critical feedback on the methodology, and reviewed the manuscript. B.S. and N.P. provided domain guidance, resources, and reviewed the manuscript. All authors reviewed and approved the final manuscript.

Initials used: S.R. (Subham Raj), K.C. (Krishnakant Chourey), S.S. (Sriparna Saha), B.S. (Brijraj Singh), N.P. (Niranjan Pedanekar).

Conflicts of interest/Competing interests. The authors have no relevant financial or non-financial interests to disclose.

Acknowledgements. The authors sincerely acknowledge the support and guidance provided by Sony Research India during the course of this research work.

References

[1] Tan Z, Zeng Z, Zeng Q, Wu Z, Liu Z, Mo F, et al. Can Large Language Models Understand Preferences in Personalized Recommendation? arXiv preprint arXiv:250113391. 2025;.

[2] Xu X, Xu Z, Yu P, Wang J. Enhancing user intent for recommendation systems via large language models. In: International Conference on Artificial Intelligence and Machine Learning Research (CAIMLR 2024). vol. 13635. SPIE; 2025. p. 46–54.

[3] Algherairy A, Ahmed M. Prompting large language models for user simulation in task-oriented dialogue systems. Computer Speech & Language. 2025;89:101697.

[4] Yoshida K, Mizukami M, Kawano S, Kruengkrai C, Sugiyama H, Yoshino K. Training Dialogue Systems by AI Feedback for Improving Overall Dialogue Impression. In: ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE; 2025. p. 1–5.

[5] Chun C, Rim D, Park J. LLM ContextBridge: A Hybrid Approach for Intent and Dialogue Understanding in IVSR. In: Proceedings of the 31st International Conference on Computational Linguistics: Industry Track; 2025. p. 794–806.

[6] Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, et al. Language models are few-shot learners. Advances in neural information processing systems. 2020;33:1877–1901.

[7] Xin H, Sun Y, Wang C, Xiong H. LLM-CDSR: Enhancing Cross-Domain Sequential Recommendation with Large Language Models. ACM Transactions on Information Systems. 2025;.

[8] Wang Y, Sang L, Zhang Y, Zhang Y. Intent representation learning with large language model for recommendation. In: Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval; 2025. p. 1870–1879.

[9] Wang X, Chen H, Pan Z, Zhou Y, Guan C, Sun L, et al. Automated disentangled sequential recommendation with large language models. ACM Transactions on Information Systems. 2025;43(2):1–29.

[10] Klimashevskaia A, Jannach D, Elahi M, Trattner C. A survey on popularity bias in recommender systems. User Modeling and User-Adapted Interaction. 2024;34(5):1777–1834.

[11] Barsby J, Hume S, Lemmey HA, Cutteridge J, Lee R, Bera KD. Pilot study on large language models for risk-of-bias assessments in systematic reviews: A (I) new type of bias? BMJ Evidence-Based Medicine. 2025;30(1):71–74.

[12] Veeramachaneni V. Large Language Models: A Comprehensive Survey on Architectures, Applications, and Challenges. Advanced Innovations in Computer Programming Languages. 2025;7(1):20–39.

[13] Abdollahpouri H, Burke R, Mobasher B. Controlling popularity bias in learning-to-rank recommendation. In: Proceedings of the eleventh ACM conference on recommender systems; 2017. p. 42–46.

[14] Steck H. Item popularity and recommendation accuracy. In: Proceedings of the fifth ACM conference on Recommender systems; 2011. p. 125–132.

[15] Wu L, Zheng Z, Qiu Z, Wang H, Gu H, Shen T, et al. A survey on large language models for recommendation. World Wide Web. 2024;27(5):60.

[16] Zhao Y, Wang Y, Liu Y, Cheng X, Aggarwal CC, Derr T. Fairness and diversity in recommender systems: a survey. ACM Transactions on Intelligent Systems and Technology. 2025;16(1):1–28.

[17] Reynolds L, McDonell K. Prompt programming for large language models: Beyond the few-shot paradigm. In: Extended abstracts of the 2021 CHI conference on human factors in computing systems; 2021. p. 1–7.

[18] Gehman S, Gururangan S, Sap M, Choi Y, Smith NA. Realtoxicityprompts: Evaluating neural toxic degeneration in language models. arXiv preprint arXiv:200911462. 2020;.

[19] Guo Y, Guo M, Su J, Yang Z, Zhu M, Li H, et al. Bias in large language models: Origin, evaluation, and mitigation. arXiv preprint arXiv:241110915. 2024;.

[20] Mehrotra R, McInerney J, Bouchard H, Lalmas M, Diaz F. Towards a fair marketplace: Counterfactual evaluation of the trade-off between relevance, fairness & satisfaction in recommendation systems. In: Proceedings of the 27th acm international conference on information and knowledge management; 2018. p. 2243–2251.

[21] Ekstrand MD, Tian M, Azpiazu IM, Ekstrand JD, Anuyah O, McNeill D, et al. All the cool kids, how do they fit in?: Popularity and demographic biases in recommender evaluation and effectiveness. In: Conference on fairness, accountability and transparency. PMLR; 2018. p. 172–186.

[22] Gao Y, Sheng T, Xiang Y, Xiong Y, Wang H, Zhang J. Chat-rec: Towards interactive and explainable llms-augmented recommender system. arXiv preprint arXiv:230314524. 2023;.

[23] Liu J, Liu C, Zhou P, Lv R, Zhou K, Zhang Y. Is chatgpt a good recommender? a preliminary study. arXiv preprint arXiv:230410149. 2023;.

[24] Lyu H, Jiang S, Zeng H, Xia Y, Wang Q, Zhang S, et al. Llm-rec: Personalized recommendation via prompting large language models. In: Findings of the Association for Computational Linguistics: NAACL 2024; 2024. p. 583–612.

[25] Zhu Y, Wu L, Guo Q, Hong L, Li J. Collaborative large language model for recommender systems. In: Proceedings of the ACM Web Conference 2024; 2024. p. 3162–3172.

[26] Lin X, Wang W, Li Y, Yang S, Feng F, Wei Y, et al. Data-efficient Fine-tuning for LLM-based Recommendation. In: Proceedings of the 47th international ACM SIGIR conference on research and development in information retrieval; 2024. p. 365–374.

[27] Rafailov R, Sharma A, Mitchell E, Manning CD, Ermon S, Finn C. Direct preference optimization: Your language model is secretly a reward model. Advances in neural information processing systems. 2023;36:53728–53741.

[28] Abdollahpouri H, Adomavicius G, Burke R, Guy I, Jannach D, Kamishima T, et al. Beyond personalization: Research directions in multistakeholder recommendation. arXiv preprint arXiv:190501986. 2019;.

[29] Schnabel T, Swaminathan A, Singh A, Chandak N, Joachims T. Recommendations as treatments: Debiasing learning and evaluation. In: international conference on machine learning. PMLR; 2016. p. 1670–1679.

[30] Abdollahpouri H, Adomavicius G, Burke R, Guy I, Jannach D, Kamishima T, et al. Multistakeholder recommendation: Survey and research directions. User Modeling and User-Adapted Interaction. 2020;30(1):127–158.

[31] Liang D, Charlin L, Blei DM. Causal inference for recommendation. In: Causation: Foundation to Application, Workshop at UAI. AUAI. vol. 6; 2016. p. 108.

[32] He X, Deng K, Wang X, Li Y, Zhang Y, Wang M. LightGCN: Simplifying and Powering Graph Convolution Network for Recommendation. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval; 2020. p. 639–648.

[33] Zhang Y, Zhu H, Song Z, Koniusz P, King I, et al. Mitigating the popularity bias of graph collaborative filtering: A dimensional collapse perspective. Advances in Neural Information Processing Systems. 2023;36:67533–67550.

[34] Deldjoo Y, Noia TD, Merra FA. A survey on adversarial recommender systems: from attack/defense strategies to generative adversarial networks. Acm Computing Surveys (Csur). 2021;54(2):1–38.

[35] Zhao Z, Fan W, Li J, Liu Y, Mei X, Wang Y, et al. Recommender systems in the era of large language models (llms). IEEE Transactions on Knowledge and Data Engineering. 2024;36(11):6889–6907.

[36] Yu Y, Jiang H, Luo X, Wu Q, Lin CY, Li D, et al. Mitigate position bias in large language models via scaling a single dimension. arXiv preprint arXiv:240602536. 2024;.

[37] Geng J, Cai F, Wang Y, Koeppl H, Nakov P, Gurevych I. A survey of confidence estimation and calibration in large language models. In: Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers); 2024. p. 6577–6595.

[38] Hou Y, Zhang J, Lin Z, Lu H, Xie R, McAuley J, et al. Large language models are zero-shot rankers for recommender systems. In: European Conference on Information Retrieval. Springer; 2024. p. 364–381.

[39] Deldjoo Y, Schedl M, Cremonesi P, Pasi G, et al. Content-based multimedia recommendation systems: definition and application domains. In: Italian Information Retrieval Workshop; 2018. p. 1–4.

[40] Bao K, Zhang J, Zhang Y, Wang W, Feng F, He X. Tallrec: An effective and efficient tuning framework to align large language model with recommendation. In: Proceedings of the 17th ACM conference on recommender systems; 2023. p. 1007–1014.

[41] Li J, Ren P, Chen Z, Ren Z, Lian T, Ma J. Neural attentive session-based recommendation. In: Proceedings of the 2017 ACM on

Conference on Information and Knowledge Management; 2017. p. 1419–1428.

[42] Tao Z, Wei Y, Wang X, He X, Huang X, Chua TS. Mgat: Multimodal graph attention network for recommendation. Information Processing & Management. 2020;57(5):102277.

[43] Deldjoo Y, Anelli VW, Zamani H, Bellogin A, Di Noia T. A flexible framework for evaluating user and item fairness in recommender systems. User Modeling and User-Adapted Interaction. 2021;31(3):457–511.

[44] Li Y, Chen H, Xu S, Ge Y, Zhang Y. Towards personalized fairness based on causal notion. In: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval; 2021. p. 1054–1063.

[45] Lin G, Liang F, Pan W, Ming Z. Fedrec: Federated recommendation with explicit feedback. IEEE Intelligent Systems. 2020;36(5):21–30.

[46] Zhu Z, Si S, Wang J, Xiao J. Cali3f: Calibrated fast fair federated recommendation system. In: 2022 International Joint Conference on Neural Networks (IJCNN). IEEE; 2022. p. 1–8.

[47] Abdollahpouri H, Burke R, Mobasher B. Managing popularity bias in recommender systems with personalized re-ranking. arXiv preprint arXiv:190107555. 2019;.

[48] Harper FM, Konstan JA. The MovieLens Datasets: History and Context. ACM Transactions on Interactive Intelligent Systems. 2015;5(4):1–19.

[49] Bennett J, Lanning S. The Netflix Prize. In: Proceedings of KDD Cup and Workshop; 2007. p. 3–6.

[50] Agarwal P, Verma R, Majumdar A.: Indian Regional Movie Dataset for Recommender Systems.

[51] Grattafiori A, Dubey A, Jauhri A, Pandey A, Kadian A, Al-Dahle A, et al. The llama 3 herd of models. arXiv preprint arXiv:240721783. 2024;.

[52] Touvron H, Martin L, Stone K, Albert P, Almahairi A, Babaei Y, et al. Llama 2: Open foundation and fine-tuned chat models. arXiv preprint arXiv:230709288. 2023;.

[53] Jiang AQ, Sablayrolles A, Mensch A, Bamford C, Chaplot DS, de las Casas D, et al.: Mistral 7B.

[54] Ouyang L, Wu J, Jiang X, Almeida D, Wainwright C, Mishkin P, et al. Training language models to follow instructions with human feedback. Advances in neural information processing systems. 2022;35:27730–27744.

[55] Steck H. Evaluation of recommendations: rating-prediction and ranking. In: Proceedings of the 7th ACM conference on Recommender systems; 2013. p. 213–220.

[56] Järvelin K, Kekäläinen J. Cumulated gain-based evaluation of IR techniques. ACM Transactions on Information Systems (TOIS). 2002;20(4):422–446.

[57] Zhou T, Kuscsik Z, Liu JG, Medo M, Wakeling JR, Zhang YC. Solving the apparent diversity-accuracy dilemma of recommender systems. Proceedings of the National Academy of Sciences. 2010;107(10):4511–4515.

[58] Herlocker JL, Konstan JA, Terveen LG, Riedl JT. Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems (TOIS). 2004;22(1):5–53.

[59] Wang X, He X, Wang M, Feng F, Chua TS. Neural graph collaborative filtering. In: Proceedings of the 42nd international ACM SIGIR conference on Research and development in Information Retrieval; 2019. p. 165–174.

[60] Kang WC, McAuley J. Self-attentive sequential recommendation. In: 2018 IEEE international conference on data mining (ICDM). IEEE; 2018. p. 197–206.