FairDiff: Enhancing User-Oriented Fairness via Curriculum-Scheduled Ambient Diffusion

Chenyang Qin
DCST, Tsinghua University
Beijing, China
qcy22@mails.tsinghua.edu.cn

Xin Wang[✉*]
DCST, BNRist, Tsinghua University
Beijing, China
xin_wang@tsinghua.edu.cn

Hantao Shu[✉]
Kuaishou Technology
Beijing, China
shuhantao@kuaishou.com

Simiao Qin
Kuaishou Technology
Beijing, China
qinsimiao@kuaishou.com

Cheng Ling
Kuaishou Technology
Beijing, China
lingcheng@kuaishou.com

Xiaoshuang Chen
Kuaishou Technology
Beijing, China
chenxiaoshuang@kuaishou.com

Kaiqiao Zhan
Kuaishou Technology
Beijing, China
zhankaiqiao@kuaishou.com

Wenwu Zhu[✉]
DCST, BNRist, Tsinghua University
Beijing, China
wwzhu@tsinghua.edu.cn

Abstract

Addressing the User-Oriented Fairness (UOF) issue is a critical challenge in recommender systems (e.g., short-video platforms), as models tend to generate lower-fidelity embeddings for users with sparse signals. Many existing approaches simplify this issue using a binary advantaged/disadvantaged dichotomy. In terms of evaluation, such group-level metrics obscure intra-group unfairness. Methodologically, overlooking the continuous spectrum of embedding degradation homogenizes disadvantaged users, which discards the residual signals essential for global interest modeling. Consequently, existing knowledge transfer methods forcibly align their degraded embeddings with limited advantaged prototypes, ultimately causing interest distribution collapse. To address these problems, we first introduce a fine-grained fairness evaluation metric based on Optimal Transport to capture the distributional unfairness across the entire population. Furthermore, we propose FairDiff, a novel framework that employs Curriculum-Scheduled Ambient Diffusion to adaptively restore degraded embeddings by modeling high-fidelity interest distribution. Instead of indiscriminately discarding information from degraded embeddings, we incorporate the entire spectrum of user data into training by explicitly modeling their varying degradation levels. Consequently, this enriches the learned interest distribution without sacrificing its high fidelity, ultimately enhancing both overall utility and fairness. Extensive evaluations in both ID-based and multimodal scenarios, coupled with online A/B testing on a real-world short-video platform, demonstrate the effectiveness of our method.

[*]DCST is the abbreviation of Department of Computer Science and Technology. BNRist is the abbreviation of Beijing National Research Center for Information Science and Technology.

MM '26, Rio de Janeiro, Brazil

ACM ISBN 979-8-4007-2213-4/2026/11
https://doi.org/10.1145/3767308.3835412

CCS Concepts

• **Information systems** → **Recommender systems**.

Keywords

Recommender Systems, User Fairness, Diffusion Models, Ambient Diffusion

ACM Reference Format:
Chenyang Qin, Xin Wang, Hantao Shu, Simiao Qin, Cheng Ling, Xiaoshuang Chen, Kaiqiao Zhan, and Wenwu Zhu. 2026. FairDiff: Enhancing User-Oriented Fairness via Curriculum-Scheduled Ambient Diffusion. In *Proceedings of the 34th ACM International Conference on Multimedia (MM '26), November 10–14, 2026, Rio de Janeiro, Brazil.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3767308.3835412

1 Introduction

Recommender systems have become fundamental tools for information delivery in applications such as short-video platforms. Leveraging user interaction data, existing approaches typically seek to maximize the overall utility of interest prediction. However, this global focus often neglects User-Oriented Fairness (UOF) [10, 11, 19]. Specifically, users with sparse interactions fail to provide sufficient signals for accurate inference and are dominated by the data-rich population during model optimization [39]. This dual challenge leads to embedding degradation, where learned representations deviate from the optimal embeddings aligned with true interests. Consequently, the model delivers superior performance to data-rich users while underserving the long tail, resulting in unfairness.

To mitigate this issue, existing approaches [10, 11, 19, 41] typically treat unfairness induced by user activity levels analogously to that associated with binary sensitive attributes (e.g., gender). Consequently, they dichotomize users into two discrete groups: a data-rich, advantaged minority with high-fidelity embeddings and a data-sparse, disadvantaged majority with degraded embeddings. Guided by this dichotomy, previous studies use re-ranking [19], distributionally robust optimization [41], or knowledge transfer [10, 11] to bridge the performance gap between these distinct groups. Among them, the latter achieves superior performance by aligning the

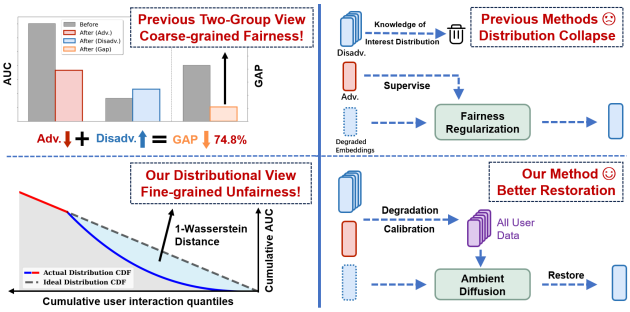

Figure 1: Comparison of methods and metrics. The left contrasts the coarse-grained metric with our fine-grained distributional perspective. The right compares previous methods (prone to distribution collapse) against our Ambient Diffusion framework for effective embedding restoration.

degraded representations of disadvantaged users with the embeddings of similar advantaged neighbors. This alignment transfers knowledge of high-fidelity interest distributions, compensating for the interaction sparsity of the disadvantaged group.

However, unlike inherently discrete sensitive attributes, user activity levels is fundamentally a continuous variable. Artificially dichotomizing users overlooks this continuity, failing to capture that embedding degradation and the resulting model utility vary along a continuous spectrum. In practice, substantial heterogeneity can exist within the large disadvantaged group, while the actual difference between users situated just across the classification boundary may be negligible. Therefore, treating the disadvantaged majority as a homogeneous entity introduces critical limitations. *In terms of evaluation*, coarse-grained group-level metrics obscure intra-group disparities arising from varying levels of embedding degradation (Fig. 1 Left). *In terms of methodology*, lacking explicit degradation modeling, knowledge transfer methods fail to exploit the residual informative signals preserved in degraded embeddings to refine the estimation of users' interest distributions (Fig. 1 Right). Instead, they enforce alignment between disadvantaged users and limited advantaged prototypes, causing interest distributions to collapse toward minority preferences and constraining performance gains. Furthermore, by treating disadvantaged users uniformly, these methods fail to prioritize embeddings with higher degradation levels, leaving intra-group disparities unaddressed.

To overcome the evaluation limitation, we advocate for a fine-grained distribution-based perspective. To this end, we introduce a novel fairness metric grounded in Optimal Transport [35]. By utilizing the Wasserstein distance to quantify the discrepancy between the actual and ideal utility distributions, our metric effectively captures the distributional unfairness across the entire population.

To address the methodological shortcomings, we propose FairDiff, a novel framework employing Ambient Diffusion [6, 7, 15, 33] for their superior capacity to model high-fidelity distributions using degraded data. By explicitly modeling varying levels of embedding degradation, FairDiff captures the high-fidelity interest distribution leveraging all embeddings and adaptively restores the degraded embeddings, achieved through three key optimizations. First, to explicitly model varying embedding degradation levels, we propose a Time-Conditioned Discriminator that quantifies the deviation of each embedding from the high-fidelity interest distribution. Second, to enrich the knowledge of interest distributions, we combine an Ambient Diffusion Loss with Timestep-Guided Curriculum Learning, thus integrating all embeddings with heterogeneous degradation levels into training without degrading the interest distribution learning. Third, to prioritize embeddings with higher degradation levels during restoration, we propose Calibrated Reverse Initialization, which initiates the reverse process based on the specific degradation level of each embedding. By adaptively intensifying restoration for more severely degraded embeddings, our method can effectively mitigate intra-group unfairness.

In summary, our contributions are as follows:

- We introduce a novel fairness metric to capture the distributional unfairness across the entire population.
- We propose FairDiff, a novel Ambient Diffusion framework that adaptively restores degraded user representations by modeling high-fidelity interest distributions using all embeddings with varying degradation levels.
- Extensive evaluations in ID-based and multimodal scenarios, coupled with online A/B testing on a real-world short-video platform, demonstrate the effectiveness of our framework in improving overall performance and fairness.

2 Preliminaries

2.1 Problem Formulation

We consider a recommendation scenario with user set $\mathcal{U}$, item set $\mathcal{I}$, and interaction matrix $\mathbf{R} \in \{0,1\}^{|\mathcal{U}|\times|\mathcal{I}|}$. Each user $u \in \mathcal{U}$ is associated with a representation $\mathbf{x}_u \in \mathbb{R}^d$, and each item $i \in \mathcal{I}$ with an embedding $\mathbf{v}_i \in \mathbb{R}^d$. In our framework, $\mathbf{x}_u$ serves as the varying-quality user representation to be refined. The predicted preference score is formalized as $\hat{y}_{ui} = f(\mathbf{x}_u, \mathbf{v}_i)$, where $f(\cdot)$ is a scoring function.

Utility Metric. We adopt the Area Under the Curve (AUC) to quantify individual ranking quality:

$$\mathrm{AUC}(u) = \frac{1}{|\mathcal{I}_u^+||\mathcal{I}_u^-|} \sum_{i\in\mathcal{I}_u^+} \sum_{j\in\mathcal{I}_u^-} \mathbb{I}(\hat{y}_{ui} > \hat{y}_{uj}), \tag{1}$$

where $\mathcal{I}_u^+$ and $\mathcal{I}_u^-$ denote the positive and negative item sets for user u, respectively. For a given user group $\mathcal{G}$, we denote its average performance as $\overline{\mathrm{AUC}}(\mathcal{G}) = \frac{1}{|\mathcal{G}|}\sum_{u\in\mathcal{G}} \mathrm{AUC}(u)$. The overall system utility is then defined as the average performance across all users, namely $\mathcal{J}_{\text{overall}} = \overline{\mathrm{AUC}}(\mathcal{U})$.

2.2 Fairness Metrics

The Binary Gap Metric. Conventional approaches [10, 11, 19] partition users into disjoint Advantaged ($\mathcal{G}_{adv}$) and Disadvantaged ($\mathcal{G}_{dis}$) groups based on activity levels (or interaction counts in public datasets). Denoting their respective performance as $\mathcal{J}_{adv} = \overline{\mathrm{AUC}}(\mathcal{G}_{adv})$ and $\mathcal{J}_{dis} = \overline{\mathrm{AUC}}(\mathcal{G}_{dis})$, fairness is typically measured by the inter-group performance disparity, i.e., $\mathcal{J}_{\text{gap}} = |\mathcal{J}_{adv} - \mathcal{J}_{dis}|$. However, this coarse-grained metric relies on an oversimplified binary split. It fails to capture the continuous nature of user activity, thereby masking significant intra-group unfairness hidden within the vast disadvantaged population.

Distributional Utility Disparity (DUD). To overcome the limitations of binary metrics, we propose a continuous metric rooted in Optimal Transport theory [35]. Fundamentally, this formulation is mathematically equivalent to the 1-Wasserstein distance between the *observed* and *ideal* utility distributions.

Specifically, we partition users into M disjoint groups based on unique activity levels $v_1 < \cdots < v_M$, where group k contains n_k users. Let U_k denote the aggregated AUC observed in group k. The observed cumulative AUC (S_k), total utility (S_{total}), and ideal cumulative AUC ($\hat{S}_k$) are defined as:

$$S_k = \sum_{j=1}^{k} U_j, \quad S_{\text{total}} = \sum_{j=1}^{M} U_j, \quad \hat{S}_k = \frac{\sum_{j=1}^{k} n_j}{N} \cdot S_{\text{total}}. \tag{2}$$

$\mathcal{J}_{\text{DUD}}$ is then defined as:

$$\mathcal{J}_{\text{DUD}} = \frac{\sum_{k=1}^{M} n_k \left| S_k - \hat{S}_k \right|}{\sum_{k=1}^{M} n_k \hat{S}_k}. \tag{3}$$

By treating the components as Riemann sum approximations, we aim to capture systematic bias across the spectrum of activity levels while tolerating stochastic noise within groups where users share the same activity level.

2.3 Ambient Diffusion

Diffusion models [15] reverse a Gaussian noising process $\mathbf{x}_t = \mathbf{x}_0 + \sigma_t \epsilon$ by optimizing a denoiser $\mathbf{h}_\theta(\mathbf{x}_t, t)$ via the Denoising Score Matching (DSM) objective:

$$J_{\text{DSM}}(\theta) = \mathbb{E}_{\mathbf{x}_0, t, \mathbf{x}_t} \left\| \mathbf{h}_\theta(\mathbf{x}_t, t) - \mathbf{x}_0 \right\|^2. \tag{4}$$

The optimal denoiser satisfies $\mathbf{h}^*_\theta(\mathbf{x}_t, t) = \mathbb{E}[\mathbf{x}_0 \mid \mathbf{x}_t, t]$, allowing us to estimate the score function $\nabla \log p_t(\mathbf{x}_t)$ via Tweedie's formula to solve the reverse-time SDE.

Ambient Diffusion via Corrupted Data. When clean data is scarce, models have to learn from corrupted distributions p_{corr}. While distinct from clean prototypes at $t = 0$, corrupted samples become statistically indistinguishable at higher timesteps ($t > t_n$). Ambient Diffusion Omni [7] leverages this to train diffusion models on further corrupted states $\mathbf{x}_t$ derived from the valid noisy observation $\mathbf{x}_{t_n}$. The objective is defined as:

$$J_{\text{Amb}}(\theta) = \mathbb{E}_{\mathbf{x}_{t_n}, t, \mathbf{x}_t} \left\| \frac{\sigma_t^2 - \sigma_{t_n}^2}{\sigma_t^2} \boldsymbol{h}_\theta(\boldsymbol{x}_t, t) + \frac{\sigma_{t_n}^2}{\sigma_t^2} \boldsymbol{x}_t - \boldsymbol{x}_{t_n} \right\|^2. \tag{5}$$

Minimizing J_{Amb} is theoretically proven to be equivalent to learning $\mathbb{E}[\mathbf{x}_0 \mid \mathbf{x}_t, t]$ when $t > t_n$, enabling the learning of the clean data distribution by leveraging corrupted observations that would otherwise be discarded.

3 Methodology

3.1 Overview

In this section, we present **FairDiff**, an Ambient Diffusion framework designed to enhance UOF by adaptively restoring user embeddings through modeling the global high-fidelity interest distribution. After pre-training the backbone, we freeze the resulting item embeddings $\mathbf{v}_i$ to serve as semantic anchors. User embeddings $\mathbf{x}_u$, also serving as the initial diffusion states $\mathbf{x}_0$, are treated as a spectrum of degraded observations (p_{corr} in Sec.2.3) derived from high-fidelity prototypes (p_{clean}). Rather than indiscriminately discarding the degraded embeddings, we integrate all embeddings with heterogeneous degradation levels into training, enriching the interest distribution learning without degrading its high fidelity.

The framework operates in three coordinated stages: Initially, we implement Adaptive Degradation Level Calibration (Sec. 3.2 and Stage 1 in Fig. 2), where a discriminator assesses the degradation of each embedding to determine the minimum timestep t_n for diffusion training. To ensure calibration stability, we employ Multi-trial Augmentation to generate a robust set of pairs $\{(\mathbf{x}_{t_n}, t_n)\}$. This annotated dataset then serves as the foundation for Ambient Diffusion Training (Sec. 3.3 and Stage 2 in Fig. 2), where we learn the global user interest distributions via Timestep-Guided Curriculum Learning. Finally, the optimized denoiser is deployed for High-Fidelity Prototype Restoration (Sec. 3.4 and Stage 3 in Fig. 2). By synergizing Calibrated Reverse Initialization to set an optimal starting point with Interaction-Anchored Semantic Guidance using historical interactions as conditioning, we effectively restore high-fidelity prototypes from the calibrated states $\mathbf{x}_{t_n}$.

3.2 Adaptive Degradation Level Calibration via Multi-trial Augmentation

This phase focuses on Adaptive Degradation Level Calibration, aiming to construct a reliable training dataset by identifying the minimum timestep t_n for each embedding, where the noisy version of the degraded observation becomes statistically indistinguishable from the high-fidelity prototype distribution of advantaged users. Intuitively, since embeddings with higher degradation exhibit greater divergence from the high-fidelity distribution, they tend to yield larger t_n values. Thus, we utilize t_n as a direct proxy to quantify the severity of embedding degradation. To achieve this, we first train a time-conditioned discriminator to distinguish between high-fidelity and degraded embeddings across varying diffusion stages. We then implement Multi-trial Augmentation to mitigate individual estimation bias and ensure calibration stability.

3.2.1 Training the Time-Conditioned Discriminator. To provide supervision for the calibration process, we treat the high-fidelity embeddings of the top $n\%$ users with the highest activity levels as the clean distribution (p_{clean}), while the remaining degraded embeddings are regarded as the corrupted observation distribution (p_{corr}). Based on this categorization, we train a Time-Conditioned Discriminator, $D_\phi(\mathbf{x}_t, t)$ following [7]. Here, $\mathbf{x}_t$ denotes the noisy state of embeddings from either p_{clean} or p_{corr}, perturbed via the predefined forward noise schedule σ_t. The discriminator's objective is to assess the evolving divergence between these two distributions across continuous timesteps. During training, we sample $t \sim \mathcal{U}(0, 1)$ and optimize ϕ to minimize the binary cross-entropy loss:

$$\begin{aligned} \mathcal{L}_{\text{D}}(\phi) = \mathbb{E}_{t,\epsilon}[\mathbb{E}_{\mathbf{x}_0 \sim p_{\text{clean}}}[-\log D_\phi(\mathbf{x}_t, t)] \\ + \mathbb{E}_{\mathbf{x}_0 \sim p_{\text{corr}}}[-\log(1 - D_\phi(\mathbf{x}_t, t))]]. \end{aligned} \tag{6}$$

As t increases, the injected noise gradually dominates the signal, causing the discrepancy between the two distributions to vanish. Consequently, the optimal discriminator output $D_\phi(\mathbf{x}_t, t)$ asymptotically approaches 0.5. This convergence provides the theoretical basis for identifying the minimum timestep t_n, defined as the point where the patterns of degraded observations become statistically indistinguishable from the high-fidelity embedding distribution.

3.2.2 Inference of t_n via Multi-trial Augmentation. We perform inference with the trained Time-Conditioned Discriminator to find

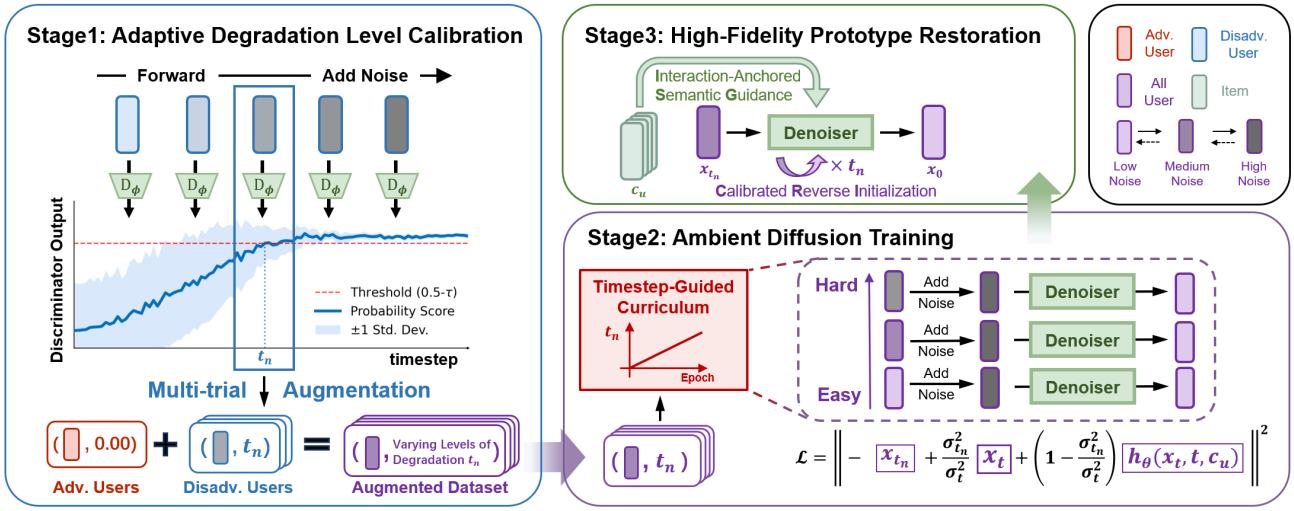

Figure 2: Illustration for our framework.

the minimum timestep t_n for each degraded embedding. At t_n, the embedding achieves distributional indistinguishability, indicated by the discriminator's output stabilizing around 0.5 within a tolerance τ. Formally, for a given realization of the forward process, t_n is defined as the earliest timestep satisfying:

$$t_n = \inf\{t \in [0,1] \mid |D_\phi(\mathbf{x}_t, t) - 0.5| \le \tau\}, \tag{7}$$

For any $t \ge t_n$, the further noise-perturbed states $\mathbf{x}_t$ of $\mathbf{x}_{t_n}$ are considered statistically indistinguishable from p_{clean}, ensuring their validity for the subsequent diffusion training.

Acknowledging that a single realization of the forward process may introduce stochastic bias, we implement the Multi-trial Augmentation strategy to ensure calibration stability. For each embedding, we execute N independent inference trials using distinct forward process realizations, yielding a robust distribution of calibrated pairs $\{(\mathbf{x}_{t_n}^{(i)}, t_n^{(i)})\}_{i=1}^N$. This process marginalizes the estimation over the noise space, preventing the model from overfitting to isolated perturbations. The final training set $\mathcal{D}_{\text{train}}$ is constructed by consolidating these calibrated pairs using Eq. 8, enabling the subsequent training to leverage the full spectrum of user data.

$$\mathcal{D}_{\text{train}} = \{(\mathbf{x}_{t_n}^{(i)}, t_n^{(i)})\}_{\mathbf{x}\sim p_{\text{corr}}, i\in[1,N]} \cup \{(\mathbf{x}, 0)\}_{\mathbf{x}\sim p_{\text{clean}}}. \tag{8}$$

3.3 Ambient Diffusion Training via Timestep-Guided Curriculum Learning

Leveraging the calibrated dataset $\mathcal{D}_{\text{train}}$, we learn the generative restoration process primarily through a conditional ambient diffusion objective. This objective is designed to reconstruct high-fidelity interest prototypes from intermediate noisy observations without the need for clean representations for the entire population. We refine this learning process by incorporating interaction-anchored guidance for semantic stabilization alongside a timestep-guided curriculum that progressively schedules training according to the calibrated difficulty spectrum.

3.3.1 Interaction-Anchored Semantic Guidance. To prevent the model from collapsing into generic patterns, we introduce a preference condition $\mathbf{c}_u$. We posit that a user's historical interactions provide the necessary semantic anchors to constrain the restoration trajectory. For each user u, the condition $\mathbf{c}_u$ is computed by aggregating the embeddings of interacted items $\mathbf{v}_i \in \mathcal{I}_u^+$, with the weights α_{ui} derived from the preference scores $\hat{y}_{ui}$ (as defined in Section 2.1) between the user and each item:

$$\mathbf{c}_u = \sum_{i\in\mathcal{I}_u^+} \alpha_{ui}\mathbf{v}_i, \quad \alpha_{ui} = \frac{\hat{y}_{ui}}{\sum_{j\in\mathcal{I}_u^+}\hat{y}_{uj}}. \tag{9}$$

This guidance ensures the model recovers personalized preferences within the bounds of the user's established interaction manifold.

3.3.2 The Conditional Ambient Objective. Initially, we sample a calibrated state $\mathbf{x}_{t_n}$ from $\mathcal{D}_{\text{train}}$ and further perturb it to $\mathbf{x}_t$ at a future time $t > t_n$ through $\mathbf{x}_t = \mathbf{x}_{t_n} + (\sigma_t^2 - \sigma_{t_n}^2)^{1/2}\boldsymbol{\epsilon}$. The network h_θ is optimized by minimizing:

$$\mathcal{L}(\theta) = \mathbb{E}_{t,\boldsymbol{\epsilon}}\left[\left\|\left(1 - \frac{\sigma_{t_n}^2}{\sigma_t^2}\right)h_\theta(\mathbf{x}_t, t, \mathbf{c}_u) + \frac{\sigma_{t_n}^2}{\sigma_t^2}\mathbf{x}_t - \mathbf{x}_{t_n}\right\|^2\right] \tag{10}$$

Since a sample with a timestep t_n is valid for training h_θ at any $t > t_n$, encompassing samples with heterogeneous t_n into $\mathcal{D}_{\text{train}}$ ensures the model learns the optimal conditional expectation $h_{\theta^*}(\mathbf{x}_t, t, \mathbf{c}_u) = \mathbb{E}[\mathbf{x}_0 \mid \mathbf{x}_t, t, \mathbf{c}_u]$ across all $t \in [0, T]$. Compared to relying solely on a scarce subset of clean data, this approach better approximates the true data distribution; simultaneously, compared to indiscriminately utilizing all samples, it effectively preserves the fidelity of the learned distribution. Consequently, this all-scale learning capability empowers the model to perform precise denoising for any calibrated state $\mathbf{x}_{t_n}$, thereby enhancing the quality of restored user embeddings and boosting the overall utility.

3.3.3 Timestep-Guided Curriculum. To stabilize the optimization landscape, we implement a Timestep-Guided Curriculum following an easy-to-hard paradigm. Within $\mathcal{D}_{\text{train}}$, training difficulty is naturally indexed by the calibrated timestep t_n. A larger t_n indicates that a higher intensity of Gaussian noise was injected during the Adaptive Degradation Level Calibration phase, which amplifies optimization uncertainty. Consequently, samples with small t_n offer

reliable gradient guidance crucial for establishing a stable base distribution, whereas those with large t_n introduce severe stochastic perturbations that can easily destabilize the early training phase.

Motivated by this insight, we formulate our Timestep-Guided Curriculum to structure the training process along a progressive difficulty gradient. We first sort all training instances into a sequence $\mathcal{S}_{\text{sorted}}$ in ascending order of their t_n values. During training epoch k, the effective sampling scope is dynamically constrained to:

$$\mathcal{S}_k = \{(\mathbf{x}_{t_n}, t_n) \in \mathcal{S}_{\text{sorted}} \mid \text{rank}(\mathbf{x}_{t_n}) \leq \gamma(k) \cdot |\mathcal{D}_{\text{train}}|\}, \tag{11}$$

where $\gamma(k) \in (0, 1]$ represents a monotonically increasing pacing function. This strategy ensures the model initially prioritizes learning a robust representation of the clean prototype distribution from high-fidelity data. As the optimization matures, the curriculum gradually expands its scope to incorporate harder samples with higher noise intensities. Ultimately, this systematic progression prevents early-stage divergence and facilitates smooth generalization across the entire spectrum of user data.

3.4 High-Fidelity Prototype Restoration via Adaptive Conditional Guidance

The final phase of FairDiff restores high-fidelity prototypes for downstream recommendation through iterative diffusion denoising. To ensure the restored representations strictly align with actual user interests, we employ Adaptive Conditional Guidance (ACG), which integrates two key mechanisms: Calibrated Reverse Initialization (CRI) sets an optimal starting point to retain valid information from the original embedding, while Interaction-Anchored Semantic Guidance (ISG) actively steers the generative trajectory using the preference condition $\mathbf{c}_u$ defined in Section 3.3.1.

3.4.1 Calibrated Reverse Initialization. Drawing inspiration from SDEdit [26], we recognize that the backbone embeddings already encode valuable latent semantics related to user interests. Therefore, rather than initiating generation from pure Gaussian noise, we utilize a noisy version of the base embedding, specifically the calibrated state $\mathbf{x}_{t_n}$ (obtained in Section 3.2.2), as the starting point for the reverse restoration trajectory.

This approach not only reduces restoring steps but also acts as an adaptive filter that balances fidelity and reconstruction. By launching restoration from t_n, we ensure that users with reliable history (small t_n) undergo a short reverse trajectory, thereby retaining the majority of their useful original information. Conversely, for users with severe embedding degradation (large t_n), the process engages in a deeper generative refinement to effectively reconstruct latent interests from the global manifold.

3.4.2 Deterministic Sampling with Classifier-free Guidance. We resolve the restoration trajectory via the Probability Flow ODE [33], where classifier-free guidance (CFG) is used to incorporate the condition information $\mathbf{c}_u$, establishing a deterministic mapping from the calibrated starting point $\mathbf{x}_{t_n}$ back to the high-fidelity interest distributions. The trajectory is solved via Euler discretization with the following update rule:

$$\mathbf{x}_{i-1} = \mathbf{x}_i + (\sigma_{i-1} - \sigma_i) \cdot \frac{\mathbf{x}_i - \tilde{h}_\theta(\mathbf{x}_i, i, \mathbf{c}_u)}{\sigma_i} \tag{12}$$

$$\tilde{h}_\theta(\mathbf{x}_i, i, \mathbf{c}_u) = (1 + w)h_\theta(\mathbf{x}_i, i, \mathbf{c}_u) - wh_\theta(\mathbf{x}_i, i, \emptyset) \tag{13}$$

where $\emptyset$ denotes the null condition and w is the guidance scale. The resulting high-fidelity prototype $\hat{\mathbf{x}}_0$ is then deployed for downstream recommendation.

3.4.3 Discussion. Beyond denoising, the reverse process can be fundamentally viewed as a hierarchical knowledge distillation mechanism. While samples calibrated at a specific timestep t_n supervise the training for $t > t_n$, their own restoration relies on the trajectory from t_n to 0. This dictates that noisier samples distill collaborative denoising knowledge from clearer embeddings with smaller t_n values. When applied to user representations, those with lower activity levels exhibit higher degradation and are thus assigned correspondingly larger t_n values. Consequently, the more disadvantaged a user is, the broader the collaborative reference signals their extended trajectory harnesses, and the greater the relative performance gains they achieve. By effectively bridging the representation gap in this manner, this mechanism naturally promotes distributional fairness.

4 Experiments

To comprehensively evaluate the proposed framework, we conduct extensive experiments on real-world datasets and an online video platform. We aim to answer the following research questions:

- **RQ1:** How does FairDiff compare against state-of-the-art methods in terms of improving overall utility and fairness?
- **RQ2:** Can FairDiff be generalized to multimodal recommendation scenarios to demonstrate its broad applicability?
- **RQ3:** What are the specific contributions of Ambient Diffusion and the three components of the framework?
- **RQ4:** Does FairDiff demonstrate tangible performance improvements in the online system?

4.1 Datasets and Settings

4.1.1 Datasets. We evaluate our proposed method on three real-world datasets: MovieLens [12], Epinions [25], and Gowalla [24]. These datasets vary in terms of interaction scales, sparsity levels, and application domains, allowing us to assess the robustness of our framework across diverse recommendation settings. The detailed statistics of the datasets are summarized in Table 1.

Table 1: Dataset Statistics

Dataset	Users	Items	Interactions	Sparsity	Domain
MovieLens	3,695	3,260	893,584	92.58%	Movie
Epinions	4,481	8,937	203,032	99.49%	Opinion
Gowalla	2,133	11,001	194,444	99.17%	POI

To prepare the experimental data, we convert explicit user ratings into binary implicit feedback, establishing positive and negative samples within the exposure space. To test the generalizability of our method under the multimodal setting (RQ2), we extract visual and textual features of MovieLens items as described in [34], and remove items with incomplete feature sets. We then apply 10-core pre-processing [30, 39] to ensure data quality.

To prevent highly active users from dominating evaluation metrics, we replace conventional random splitting with a user-stratified approach. Specifically, this strategy randomly reserves a fixed number of positive and negative instances per user for validation (P_{val} =

N_{val}) and testing ($P_{\text{test}} = N_{\text{test}}$). These values are set to 5 and 10, respectively, for MovieLens, and 3 and 6 for other datasets, thereby ensuring fair evaluation across all users.

4.1.2 Backbone Models and Baselines. To evaluate the generality of our approach, we conduct experiments on three representative ID-based recommendation backbones, including PMF [32], NGCF [38], and LightGCN [14]. We further extend the evaluation to the multimodal setting with five additional backbones: VBPR [13], MMGCN [40], LATTICE [45], FREEDOM [48], and PGL [44]. Based on these backbones, we compare FairDiff with representative user-oriented fairness baselines, including Vanilla, UFR [19], S-DRO [41], In-UCDS [10], and FairDGCL [4].

4.1.3 Evaluation Protocols. We utilize $\mathcal{J}_{\text{overall}}$ to evaluate the global recommendation utility. To provide a multi-dimensional fairness assessment, we first evaluate coarse-grained group-based fairness following prior work [10, 30] by partitioning the user population into *advantaged* (top 5% by activity) and *disadvantaged* (remaining 95%) groups. We report the average AUC for each group, denoted as $\mathcal{J}_{\text{adv}}$ and $\mathcal{J}_{\text{dis}}$, along with their disparity $\mathcal{J}_{\text{gap}}$. For fine-grained distributional fairness, we report the $\mathcal{J}_{\text{DUD}}$ to quantify the utility divergence across the entire activity spectrum. To ensure statistical reliability, all experiments are conducted 5 times with different seeds, and we report the average performance across these runs.

4.1.4 Implementation Details. All backbone models utilize an embedding size of 32 and are optimized via Adam [17]. For a fair comparison under similar computational constraints, we align the total training budget: while baselines are trained for 200 epochs, FairDiff allocates 100 epochs for backbone pre-training and 100 epochs for the diffusion process. Hyperparameters for all methods are tuned based on validation performance.

4.2 RQ1: Overall Comparison

As shown in Table 2, our proposed method consistently outperforms baseline models in both overall utility and fairness metrics.

In terms of overall utility, we observe that: (i) Reranking-based UFR and reweighting-based S-DRO fail to improve $\mathcal{J}_{\text{overall}}$ significantly, as they cannot recover the information lost due to data sparsity. (ii) While knowledge transfer-based In-UCDS offers marginal gains, it suffers from representation collapse by supervising a vast number of disadvantaged users with limited data from advantaged ones. (iii) In contrast, FairDiff achieves the most substantial utility improvement. This validates our analysis in Sec. 3.3.2: by leveraging the full spectrum of user data via Ambient Diffusion, our method prevents mode collapse and effectively reconstructs high-fidelity representations from degraded observations.

In terms of fairness evaluation, our proposed $\mathcal{J}_{\text{DUD}}$ demonstrates superior robustness and granularity over $\mathcal{J}_{\text{gap}}$. First, $\mathcal{J}_{\text{gap}}$ lacks robustness because it can minimize the gap by simply degrading the performance of advantaged users rather than uplifting disadvantaged ones, a compromise frequently observed in UFR. Second, it hides intra-group unfairness by averaging disadvantaged users into a single scalar; for instance, while FairDGCL achieves a deceptively low $\mathcal{J}_{\text{gap}}$ on Epinions, its persistently high $\mathcal{J}_{\text{DUD}}$ (0.0150 vs. our 0.0075) exposes severe unresolved intra-group disparity. In contrast, $\mathcal{J}_{\text{DUD}}$ utilizes Optimal Transport to evaluate full distributional alignment. Evaluated against these standards, our method achieves significant gains in both metrics, empirically validating the theoretical analysis in Sec. 3.4.3. Specifically, the improvement in coarse-grained $\mathcal{J}_{\text{gap}}$ implies that we have effectively improved the overall utility of disadvantaged users, which stems from modeling the full user spectrum to narrow group-level disparities. Simultaneously, the gains in fine-grained $\mathcal{J}_{\text{DUD}}$ indicate that users with lower activity levels leverage a broader spectrum of reference signals, thereby deriving greater performance improvements.

4.3 RQ2: Generalizability to Multimodal Recommendation

To demonstrate the broad applicability and robustness of FairDiff beyond traditional ID-based recommendation, we extend our evaluation to the multimodal recommendation domain. In this setting, items are associated with visual and textual features. We evaluate our method against the state-of-the-art In-UCDS baseline on the multimodal-extended MovieLens dataset. As shown in Table 3, our model consistently outperforms In-UCDS in both overall utility and fairness metrics across various backbones. This strongly highlights the exceptional generalizability of our proposed framework.

4.4 RQ3: Ablation Study

To evaluate the contributions of Ambient Diffusion alongside the three core components of our framework, we conduct extensive ablation studies. We report the experimental results on the MovieLens dataset. Similar trends are observed across other datasets and are thus omitted for brevity.

4.4.1 Ablation for Ambient Diffusion. We compare our method against a standard diffusion baseline trained on the top n% advantaged users. Different values of n represent two extreme cases: a small n relies only on a limited subset of clean data, while $n = 100$ simply uses all samples regardless of degradation. Since this baseline treats inputs as ground truth and lacks degradation calibration, it cannot utilize Calibrated Reverse Initialization. For a fair evaluation of distribution modeling, all methods sample from a standard Gaussian distribution using Interaction-Aware Semantic Guidance.

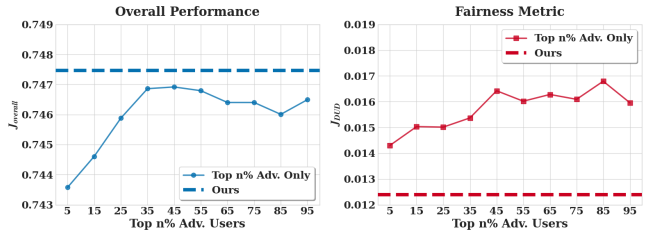

Figure 3: Ablation for Ambient Diffusion Training using PMF

As shown in Figure 3, the baseline exhibits an inverted-U trend in overall performance. Its utility initially improves with data scale but deteriorates after $n > 45\%$ as including disadvantaged users introduces degraded embeddings that interfere with learning high-fidelity representations. Regarding fairness, the baseline's $\mathcal{J}_{\text{DUD}}$ worsens consistently because the standard diffusion process treats

Table 2: Overall performance comparison. The top performance is highlighted in bold and the runner-up with underlining.

		MovieLens					Epinions					Gowalla				
		$\mathcal{J}_{\text{overall}}$ ↑	$\mathcal{J}_{\text{adv}}$	$\mathcal{J}_{\text{dis}}$ ↑	$\mathcal{J}_{\text{gap}}$ ↓	$\mathcal{J}_{\text{DUD}}$ ↓	$\mathcal{J}_{\text{overall}}$ ↑	$\mathcal{J}_{\text{adv}}$	$\mathcal{J}_{\text{dis}}$ ↑	$\mathcal{J}_{\text{gap}}$ ↓	$\mathcal{J}_{\text{DUD}}$ ↓	$\mathcal{J}_{\text{overall}}$ ↑	$\mathcal{J}_{\text{adv}}$	$\mathcal{J}_{\text{dis}}$ ↑	$\mathcal{J}_{\text{gap}}$ ↓	$\mathcal{J}_{\text{DUD}}$ ↓
PMF	Vanilla	0.7465	0.7760	0.7449	0.0311	0.0165	0.6947	0.7285	0.6929	0.0356	0.0170	0.6172	0.6514	0.6154	0.0360	0.0267
	S-DRO	0.7471	0.7751	0.7456	0.0295	0.0159	0.6961	0.7301	0.6943	0.0358	0.0169	0.6168	0.6542	0.6148	0.0394	0.0275
	UFR	0.7470	0.7759	0.7454	0.0304	0.0159	0.6945	0.7218	0.6930	0.0288	0.0164	0.6172	0.6465	0.6156	0.0309	0.0263
	In-UCDS	0.7475	0.7762	0.7460	0.0302	0.0149	0.7040	0.7289	0.7027	0.0262	0.0126	0.6212	0.6535	0.6195	0.0340	0.0237
	Ours	**0.7488**	0.7754	**0.7474**	**0.0280**	**0.0121**	**0.7171**	0.7248	**0.7167**	**0.0082**	**0.0071**	**0.6247**	0.6514	**0.6233**	**0.0281**	**0.0222**
NGCF	Vanilla	0.7445	0.7779	0.7427	0.0352	0.0174	0.6876	0.7156	0.6861	0.0295	0.0226	0.6218	0.6684	0.6193	0.0491	0.0211
	S-DRO	0.7449	0.7744	0.7433	**0.0311**	0.0161	0.6862	0.7136	0.6848	0.0288	0.0224	0.6221	0.6712	0.6195	0.0517	0.0211
	UFR	0.7446	0.7762	0.7430	0.0332	0.0170	0.6875	0.7103	0.6863	0.0240	0.0222	0.6217	0.6604	0.6197	0.0407	0.0207
	In-UCDS	0.7458	0.7763	0.7442	0.0321	0.0163	0.7019	0.7158	0.7011	0.0147	0.0143	0.6256	0.6622	0.6236	0.0386	0.0201
	Ours	**0.7475**	0.7778	**0.7459**	0.0319	**0.0132**	**0.7080**	0.7168	**0.7075**	**0.0093**	**0.0131**	**0.6365**	0.6635	**0.6351**	**0.0284**	**0.0116**
LightGCN	Vanilla	0.7466	0.7765	0.7450	0.0315	0.0166	0.7059	0.7469	0.7038	0.0431	0.0185	0.6279	0.6741	0.6255	0.0486	0.0169
	S-DRO	0.7467	0.7758	0.7452	0.0306	0.0164	0.7060	0.7475	0.7038	0.0437	0.0181	0.6278	0.6706	0.6256	0.0450	0.0164
	UFR	0.7465	0.7739	0.7451	**0.0288**	0.0164	0.7049	0.7397	0.7030	0.0367	0.0182	0.6282	0.6536	0.6268	0.0268	0.0152
	FairDGCL	0.7474	0.7768	0.7459	0.0309	0.0157	0.7084	0.7214	0.7077	**0.0137**	0.0150	0.6338	0.6560	0.6326	**0.0234**	0.0132
	In-UCDS	0.7473	0.7760	0.7458	0.0302	0.0158	0.7156	0.7447	0.7141	0.0306	0.0110	0.6309	0.6671	0.6290	0.0381	0.0124
	Ours	**0.7485**	0.7771	**0.7470**	0.0301	**0.0121**	**0.7241**	0.7444	**0.7230**	0.0214	**0.0075**	**0.6387**	0.6653	**0.6373**	0.0280	**0.0107**

Table 3: Results on the multimodal extended MovieLens dataset. The best results are highlighted in bold.

Backbone	Method	$\mathcal{J}_{\text{overall}}$ ↑	$\mathcal{J}_{\text{adv}}$	$\mathcal{J}_{\text{dis}}$ ↑	$\mathcal{J}_{\text{gap}}$ ↓	$\mathcal{J}_{\text{DUD}}$ ↓
VBPR	Vanilla	0.7523	0.7929	0.7502	0.0427	0.0151
	In-UCDS	0.7530	0.7918	0.7511	0.0407	0.0140
	Ours	**0.7537**	0.7917	**0.7517**	**0.0400**	**0.0128**
MMGCN	Vanilla	0.7528	0.7920	0.7507	0.0413	0.0185
	In-UCDS	0.7535	0.7921	0.7514	0.0407	0.0161
	Ours	**0.7541**	0.7913	**0.7521**	**0.0392**	**0.0140**
FREEDOM	Vanilla	0.7598	0.8082	0.7573	0.0509	0.0165
	In-UCDS	0.7603	0.8072	0.7579	0.0493	0.0154
	Ours	**0.7610**	0.8063	**0.7586**	**0.0477**	**0.0143**
LATTICE	Vanilla	0.7595	0.8005	0.7573	0.0432	0.0146
	In-UCDS	0.7602	0.8010	0.7580	0.0430	0.0139
	Ours	**0.7605**	0.7984	**0.7585**	**0.0399**	**0.0135**
PGL	Vanilla	0.7608	0.8117	0.7582	0.0535	0.0196
	In-UCDS	0.7615	0.8092	0.7590	**0.0503**	0.0173
	Ours	**0.7622**	0.8101	**0.7597**	0.0504	**0.0156**

all inputs as high-fidelity ground truth, forcing the model to learn and regenerate degraded distributions. In contrast, our method achieves optimal utility and fairness. By properly calibrating degraded embeddings to the appropriate timestep via Ambient Diffusion, we effectively utilize the full spectrum of data to aid restoration without allowing noisy signals to disrupt the high-fidelity manifold.

4.4.2 Ablation for Multi-trial Augmentation. To evaluate the Multi-trial Calibration mechanism, we conduct ablation studies on the number of trials N. As shown in Figure 4, increasing N from 1 to 30 significantly improves both $\mathcal{J}_{\text{overall}}$ and $\mathcal{J}_{\text{DUD}}$, confirming that multi-trial augmentation mitigates stochastic bias through diverse noise realizations. Since performance gains become marginal after $N =$ 30, we set 30 as the default to balance effectiveness and efficiency.

4.4.3 Ablation for Timestep-Guided Curriculum Learning. We evaluate the contribution of the Timestep-Guided Curriculum Learning

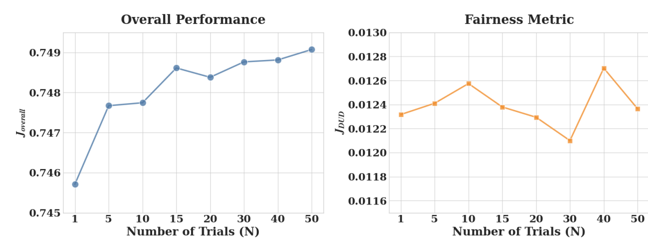

Figure 4: Ablation for Multi-trial Augmentation using PMF

by comparing our method (*w/ cur*) against a variant that uniformly samples $\mathbf{x}_{t_n}$ from the beginning of training (*w/o cur*).

Table 4: Ablation for Timestep-Guided Curriculum Learning.

Method	PMF		NGCF		LightGCN	
	$\mathcal{J}_{\text{overall}}$ ↑	$\mathcal{J}_{\text{DUD}}$ ↓	$\mathcal{J}_{\text{overall}}$ ↑	$\mathcal{J}_{\text{DUD}}$ ↓	$\mathcal{J}_{\text{overall}}$ ↑	$\mathcal{J}_{\text{DUD}}$ ↓
w/o cur	0.7486	0.0125	0.7473	0.0134	0.7481	0.0127
w/ cur	**0.7488**	**0.0121**	**0.7475**	**0.0132**	**0.7485**	**0.0121**

As shown in Table 4, the curriculum schedule yields consistent improvements in both utility and fairness across backbones, validating the *easy-to-hard* paradigm. By anchoring the diffusion manifold on high-fidelity embeddings, the curriculum establishes a stable prior. This prevents early-stage divergence caused by high-uncertainty samples and guides the restoration of severely degraded representations as training progresses to harder stages.

4.4.4 Ablation for Adaptive Conditional Guidance. We further investigate the ACG mechanism by isolating CRI and ISG modules.

As shown in Table 5, the full ACG strategy achieves the highest utility. While the baseline (*None*) yields the lowest $\mathcal{J}_{\text{DUD}}$, its inferior utility indicates that standard diffusion from pure noise generates generic distributions lacking personalized precision. Our gains stem

Table 5: Ablation for ACG. *None* denotes the standard diffusion sampling from Gaussian noise without condition.

Method	PMF		NGCF		LightGCN	
	$\mathcal{J}_{\text{overall}} \uparrow$	$\mathcal{J}_{\text{DUD}} \downarrow$	$\mathcal{J}_{\text{overall}} \uparrow$	$\mathcal{J}_{\text{DUD}} \downarrow$	$\mathcal{J}_{\text{overall}} \uparrow$	$\mathcal{J}_{\text{DUD}} \downarrow$
None	0.7117	**0.0099**	0.7104	**0.0118**	0.7124	**0.0114**
CRI	0.7443	0.0133	0.7430	0.0148	0.7438	0.0135
ISG	0.7475	0.0126	0.7467	0.0132	0.7478	0.0128
ACG (Both)	**0.7488**	0.0121	**0.7475**	0.0132	**0.7485**	0.0121

from coupling ISG with CRI. By launching restoration from the calibrated state $\mathbf{x}_{t_n}$, CRI preserves backbone information and reduces the Number of Function Evaluations (NFE) by bypassing unnecessary denoising steps. Meanwhile, ISG acts as a semantic anchor to align the generative trajectory with user preferences. Ultimately, this integration ensures the robust recovery of personalized user embeddings to deliver high-fidelity recommendations efficiently.

4.5 RQ4: Online A/B Experiment

We evaluate FairDiff via an online A/B test on a large-scale video recommendation platform with hundreds of millions of daily active users (DAUs). FairDiff is applied to pretrained user embeddings, and then deployed in the ranking stage for the finish-view prediction task. The experiment ran for five consecutive days and allocated 10% of user traffic to the treatment. Table 6 reports the results. We define the Finish View-Through Rate (FVTR) as a binary indicator equal to 1 when a user watches a video to completion and 0 otherwise. The proposed method improves both the primary task metric (FVTR) and overall user engagement, measured by app usage time, with particularly pronounced gains among low- and mid-activity users.

Online Metric	App Usage Time	FVTR
All users	+0.052%	+0.516%*
Low-activity users	+0.528%*	+1.127%*
Mid-activity users	+0.333%*	+0.793%*

Table 6: Results of the online A/B testing experiment. We define low-activity users as those with LT30 < 10, and mid-activity users as those with $10 \leq$ LT30 < 20. * indicates that the performance gains are statistically significant at $p < 0.05$.

5 Related Work

5.1 User-Oriented Fairness in Recommendation

Current research addressing UOF mainly falls into three categories: re-ranking strategies, distributionally robust optimization (DRO), and knowledge transfer methods. Re-ranking approaches [19] function as model-agnostic post-processing steps that adjust recommendation lists to bound the utility gap between advantaged and disadvantaged users. Conversely, DRO-based methods, exemplified by Streaming-DRO [41], intervene during training. These methods optimize for worst-case distributions by assigning higher gradients to disadvantaged users, thereby counteracting the dominance of advantaged users in empirical risk minimization. Furthermore, knowledge transfer frameworks [10, 11] address the scarcity of collaborative signals by transferring behavioral patterns from advantaged users to disadvantaged groups, typically employing clustering and embedding alignment to yield more informative representations.

5.2 Diffusion Models

Diffusion Probabilistic Models (DPMs) have emerged as a dominant generative paradigm, offering superior training stability and mode coverage by formulating generation as a sequential denoising process [15, 33]. Due to these advantages, DPMs have achieved success in computer vision tasks, ranging from high-fidelity image synthesis [31, 49] to complex video generation [27–29, 46]. In the domain of recommender systems, their application generally falls into three directions: serving as backbone recommenders for interaction generation [20, 36], facilitating data augmentation to synthesize pseudo-interactions [42], and enhancing representation robustness via iterative denoising [22, 23, 47]. To address scenarios where uncorrupted samples are prohibitively expensive to acquire (e.g., in medical imaging such as MRI [1]), recent paradigms like Ambient Diffusion have extended DPMs to train directly on corrupted data. By utilizing further corruption mechanisms and Tweedie's formula for score estimation, these models enable sampling from the underlying high-fidelity distribution [6, 7].

5.3 Curriculum Learning

Curriculum Learning (CL) [2, 8, 18, 37] mimics the human cognitive process of learning from easy to hard. By organizing training data via a difficulty measurer and a training scheduler, CL smooths the optimization landscape of non-convex functions, guiding models toward better local minima while accelerating convergence. In recommendation systems, this paradigm is widely adopted to mitigate noise and sparsity [5]. Models are typically pre-trained on high-confidence interactions, before progressively introducing harder or noisier samples to enhance robustness [3, 42]. Extending this to diffusion models [9, 21], CL addresses heterogeneity across timesteps and data distributions. Existing works often employ timestep scheduling to prioritize easier global structure generation at high noise levels prior to harder fine-grained detail restoration [16, 43]. Alternatively, data-level strategies like synthetic-to-real curricula, including DisCL [21] and EvoGen [9], are adopted to bridge distribution gaps in long-tail or complex generation scenarios.

6 Conclusions

In this work, we move beyond the coarse-grained binary dichotomy in User-Oriented Fairness by recognizing the continuous spectrum of user activity. To accurately evaluate distributional fairness across the entire population, we first introduce a fine-grained metric based on Optimal Transport. Furthermore, to overcome the interest distribution collapse and overlooked intra-group unfairness prevalent in existing approaches, we propose FairDiff. This novel Ambient Diffusion framework adaptively restores degraded user embeddings by capturing high-fidelity interest distributions. Extensive experiments across ID-based and multimodal scenarios, alongside online A/B testing on a short-video platform, demonstrate that FairDiff significantly enhances both overall utility and fairness.

Acknowledgments

This work is supported by the National Key Research and Development Program of China under Grant No. 2022ZD0115903. It is also partially supported by Kuaishou Technology.

References

[1] Asad Aali, Marius Arvinte, Sidharth Kumar, and Jonathan I. Tamir. 2023. Solving Inverse Problems with Score-Based Generative Priors learned from Noisy Data. In *2023 57th Asilomar Conference on Signals, Systems, and Computers*. IEEE, 837–843. doi:10.1109/ieeeconf59524.2023.10477042

[2] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. 2009. Curriculum learning. In *Proceedings of the 26th Annual International Conference on Machine Learning* (Montreal, Quebec, Canada) *(ICML '09)*. Association for Computing Machinery, New York, NY, USA, 41–48. doi:10.1145/1553374.1553380

[3] Hong Chen, Yudong Chen, Xin Wang, Ruobing Xie, Rui Wang, Feng Xia, and Wenwu Zhu. 2021. Curriculum disentangled recommendation with noisy multi-feedback. In *Proceedings of the 35th International Conference on Neural Information Processing Systems (NIPS '21)*. Curran Associates Inc., Red Hook, NY, USA, Article 2062, 13 pages.

[4] Wei Chen, Meng Yuan, Zhao Zhang, Ruobing Xie, Fuzhen Zhuang, Deqing Wang, and Rui Liu. 2025. FairDgcl: Fairness-Aware Recommendation With Dynamic Graph Contrastive Learning. *IEEE Transactions on Knowledge and Data Engineering* 37, 9 (2025), 5230–5242. doi:10.1109/TKDE.2025.3580087

[5] Yudong Chen, Xin Wang, Miao Fan, Jizhou Huang, Shengwen Yang, and Wenwu Zhu. 2021. Curriculum Meta-Learning for Next POI Recommendation. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining* (Virtual Event, Singapore) *(KDD '21)*. Association for Computing Machinery, New York, NY, USA, 2692–2702. doi:10.1145/3447548.3467132

[6] Giannis Daras, Alexandros G. Dimakis, and Constantinos Daskalakis. 2024. Consistent Diffusion Meets Tweedie: Training Exact Ambient Diffusion Models with Noisy Data. arXiv:2404.10177 [cs.CV] https://arxiv.org/abs/2404.10177

[7] Giannis Daras, Adrian Rodriguez-Munoz, Adam Klivans, Antonio Torralba, and Constantinos Daskalakis. 2025. Ambient Diffusion Omni: Training Good Models with Bad Data. arXiv:2506.10038 [cs.GR] https://arxiv.org/abs/2506.10038

[8] Chendi Ge, Xin Wang, Zeyang Zhang, Hong Chen, Jiapei Fan, Longtao Huang, Hui Xue, and Wenwu Zhu. 2025. Dynamic Mixture of Curriculum LoRA Experts for Continual Multimodal Instruction Tuning. arXiv:2506.11672 [cs.CV] https://arxiv.org/abs/2506.11672

[9] Evans Xu Han, Linghao Jin, Xiaofeng Liu, and Paul Pu Liang. 2025. Progressive Compositionality in Text-to-Image Generative Models. arXiv:2410.16719 [cs.CV] https://arxiv.org/abs/2410.16719

[10] Zhongxuan Han, Chaochao Chen, Xiaolin Zheng, Weiming Liu, Jun Wang, Wenjie Cheng, and Yuyuan Li. 2023. In-processing User Constrained Dominant Sets for User-Oriented Fairness in Recommender Systems. In *Proceedings of the 31st ACM International Conference on Multimedia (MM '23)*. ACM, 6190–6201. doi:10.1145/3581783.3613831

[11] Zhongxuan Han, Chaochao Chen, Xiaolin Zheng, Li Zhang, and Yuyuan Li. 2024. Hypergraph Convolutional Network for User-Oriented Fairness in Recommender Systems. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Washington DC, USA) *(SIGIR '24)*. Association for Computing Machinery, New York, NY, USA, 903–913. doi:10.1145/3626772.3657737

[12] F. Maxwell Harper and Joseph A. Konstan. 2015. The MovieLens Datasets: History and Context. *ACM Trans. Interact. Intell. Syst.* 5, 4, Article 19 (Dec. 2015), 19 pages. doi:10.1145/2827872

[13] Ruining He and Julian McAuley. 2015. VBPR: Visual Bayesian Personalized Ranking from Implicit Feedback. arXiv:1510.01784 [cs.IR] https://arxiv.org/abs/1510.01784

[14] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yongdong Zhang, and Meng Wang. 2020. LightGCN: Simplifying and Powering Graph Convolution Network for Recommendation. arXiv:2002.02126 [cs.IR] https://arxiv.org/abs/2002.02126

[15] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising Diffusion Probabilistic Models. arXiv:2006.11239 [cs.LG] https://arxiv.org/abs/2006.11239

[16] Jin-Young Kim, Hyojun Go, Soonwoo Kwon, and Hyun-Gyoon Kim. 2025. Denoising Task Difficulty-based Curriculum for Training Diffusion Models. arXiv:2403.10348 [cs.CV] https://arxiv.org/abs/2403.10348

[17] Diederik P. Kingma and Jimmy Ba. 2017. Adam: A Method for Stochastic Optimization. arXiv:1412.6980 [cs.LG] https://arxiv.org/abs/1412.6980

[18] Haoyang Li, Xin Wang, Zeyang Zhang, Zongyuan Wu, Linxin Xiao, and Wenwu Zhu. 2025. Self-supervised Masked Graph Autoencoder via Structure-aware Curriculum. In *Proceedings of the 42nd International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 267)*, Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (Eds.). PMLR, 36215–36235. https://proceedings.mlr.press/v267/li25ct.html

[19] Yunqi Li, Hanxiong Chen, Zuohui Fu, Yingqiang Ge, and Yongfeng Zhang. 2021. User-oriented Fairness in Recommendation. In *Proceedings of the Web Conference 2021 (WWW '21)*. ACM, 624–632. doi:10.1145/3442381.3449866

[20] Zihao Li, Aixin Sun, and Chenliang Li. 2023. DiffuRec: A Diffusion Model for Sequential Recommendation. arXiv:2304.00686 [cs.IR] https://arxiv.org/abs/2304.00686

[21] Yijun Liang, Shweta Bhardwaj, and Tianyi Zhou. 2025. Diffusion Curriculum: Synthetic-to-Real Data Curriculum via Image-Guided Diffusion. arXiv:2410.13674 [cs.CV] https://arxiv.org/abs/2410.13674

[22] Jianghao Lin, Jiaqi Liu, Jiachen Zhu, Yunjia Xi, Chengkai Liu, Yangtian Zhang, Yong Yu, and Weinan Zhang. 2024. A Survey on Diffusion Models for Recommender Systems. arXiv:2409.05033 [cs.IR] https://arxiv.org/abs/2409.05033

[23] Xiao Lin, Xiaokai Chen, Chenyang Wang, Hantao Shu, Linfeng Song, Biao Li, and Peng jiang. 2023. Discrete Conditional Diffusion for Reranking in Recommendation. arXiv:2308.06982 [cs.IR] https://arxiv.org/abs/2308.06982

[24] Yiding Liu, Tuan-Anh Nguyen Pham, Gao Cong, and Quan Yuan. 2017. An experimental evaluation of point-of-interest recommendation in location-based social networks. *Proc. VLDB Endow.* 10, 10 (June 2017), 1010–1021. doi:10.14778/3115404.3115407

[25] Paolo Massa and Paolo Avesani. 2007. Trust-aware recommender systems *(RecSys '07)*. Association for Computing Machinery, New York, NY, USA, 17–24. doi:10.1145/1297231.1297235

[26] Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. 2022. SDEdit: Guided Image Synthesis and Editing with Stochastic Differential Equations. arXiv:2108.01073 [cs.CV] https://arxiv.org/abs/2108.01073

[27] Zirui Pan, Xin Wang, Yipeng Zhang, Hong Chen, Kwan Man Cheng, Yaofei Wu, and Wenwu Zhu. 2025. Modular-cam: Modular dynamic camera-view video generation with llm. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39. 6363–6371.

[28] Zirui Pan, Xin Wang, Yipeng Zhang, Hong Chen, Kecheng Zheng, and Wenwu Zhu. 2026. Reasoning Diffusion for Unpaired Test Time Out-of-distribution Text-Image to Video Generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 636–646.

[29] Zirui Pan, Xin Wang, Yipeng Zhang, Yuwei Zhou, and Wenwu Zhu. 2026. Temporal-aware Flow Matching for Video Generation with Temporally Coherent Motion. In *Forty-third International Conference on Machine Learning*.

[30] Hossein A. Rahmani, Mohammadmehdi Naghiaei, Mahdi Dehghan, and Mohammad Aliannejadi. 2022. Experiments on Generalizability of User-Oriented Fairness in Recommender Systems *(SIGIR '22)*. Association for Computing Machinery, New York, NY, USA, 2755–2764. doi:10.1145/3477495.3531718

[31] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-Resolution Image Synthesis with Latent Diffusion Models. arXiv:2112.10752 [cs.CV] https://arxiv.org/abs/2112.10752

[32] Ruslan Salakhutdinov and Andriy Mnih. 2007. Probabilistic Matrix Factorization *(NIPS'07)*. Curran Associates Inc., Red Hook, NY, USA, 1257–1264.

[33] Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. 2021. Score-Based Generative Modeling through Stochastic Differential Equations. arXiv:2011.13456 [cs.LG] https://arxiv.org/abs/2011.13456

[34] Giuseppe Spillo, Elio Musacchio, Cataldo Musto, Marco de Gemmis, Pasquale Lops, and Giovanni Semeraro. 2025. See the Movie, Hear the Song, Read the Book: Extending MovieLens-1M, Last.fm-2K, and DBbook with Multimodal Data. In *Proceedings of the Nineteenth ACM Conference on Recommender Systems (RecSys '25)*. Association for Computing Machinery, New York, NY, USA, 847–856. doi:10.1145/3705328.3748162

[35] Cédric Villani. 2009. *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften, Vol. 338. Springer Berlin Heidelberg, Berlin, Heidelberg. doi:10.1007/978-3-540-71050-9

[36] Wenjie Wang, Yiyan Xu, Fuli Feng, Xinyu Lin, Xiangnan He, and Tat-Seng Chua. 2025. Diffusion Recommender Model. arXiv:2304.04971 [cs.IR] https://arxiv.org/abs/2304.04971

[37] Xin Wang, Yudong Chen, and Wenwu Zhu. 2021. A Survey on Curriculum Learning. arXiv:2010.13166 [cs.LG] https://arxiv.org/abs/2010.13166

[38] Xiang Wang, Xiangnan He, Meng Wang, Fuli Feng, and Tat-Seng Chua. 2019. Neural Graph Collaborative Filtering. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '19)*. ACM, 165–174. doi:10.1145/3331184.3331267

[39] Yifan Wang, Weizhi Ma, Min Zhang, Xiaoxiao Xu, Zhiqiang Liu, and Shaoping Ma. 2025. Improving Long-tail User CTR Prediction via Hierarchical Distribution Alignment. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2* (Toronto ON, Canada) *(KDD '25)*. Association for Computing Machinery, New York, NY, USA, 3079–3090. doi:10.1145/3711896.3737003

[40] Yinwei Wei, Xiang Wang, Liqiang Nie, Xiangnan He, Richang Hong, and {Tat Seng} Chua. 2019. MMGCN: Multi-modal graph convolution network for personalized recommendation of micro-video. In *Proceedings of the 27th ACM International Conference on Multimedia*, Guillaume Gravier, Hayley Hung, Chong-Wah Ngo, and Wei {Tsang Ooi} (Eds.). Association for Computing Machinery (ACM), United States of America, 1437–1445. doi:10.1145/3343031.3351034

[41] Hongyi Wen, Xinyang Yi, Tiansheng Yao, Jiaxi Tang, Lichan Hong, and Ed H. Chi. 2022. Distributionally-robust Recommendations for Improving Worst-case User Experience. In *Proceedings of the ACM Web Conference 2022* (Virtual Event, Lyon, France) *(WWW '22)*. Association for Computing Machinery, New York, NY, USA, 3606–3610. doi:10.1145/3485447.3512255

[42] Zihao Wu, Xin Wang, Hong Chen, Kaidong Li, Yi Han, Lifeng Sun, and Wenwu Zhu. 2023. Diff4Rec: Sequential Recommendation with Curriculum-scheduled Diffusion Augmentation *(MM '23)*. Association for Computing Machinery, New York, NY, USA, 7 pages. doi:10.1145/3581783.3612709

[43] Xuanyu Yi, Zike Wu, Qingshan Xu, Pan Zhou, Joo-Hwee Lim, and Hanwang Zhang. 2024. Diffusion Time-step Curriculum for One Image to 3D Generation. arXiv:2404.04562 [cs.CV] https://arxiv.org/abs/2404.04562

[44] Penghang Yu, Zhiyi Tan, Guanming Lu, and Bing-Kun Bao. 2025. Mind Individual Information! Principal Graph Learning for Multimedia Recommendation. *Proceedings of the AAAI Conference on Artificial Intelligence* 39, 12 (Apr. 2025), 13096–13105. doi:10.1609/aaai.v39i12.33429

[45] Jinghao Zhang, Yanqiao Zhu, Qiang Liu, Shu Wu, Shuhui Wang, and Liang Wang. 2021. Mining Latent Structures for Multimedia Recommendation. In *Proceedings of the 29th ACM International Conference on Multimedia (MM '21)*. ACM, 3872–3880. doi:10.1145/3474085.3475259

[46] Yipeng Zhang, Xin Wang, Hong Chen, Chenyang Qin, Yibo Hao, Wenwu Zhu, and Hong Mei. 2025. ScenarioDiff: Text-to-video Generation with Dynamic Transformations of Scene Conditions. *International Journal of Computer Vision* 133 (2025), 4909–4922. doi:10.1007/s11263-025-02413-7

[47] Jujia Zhao, Wenjie Wang, Yiyan Xu, Teng Sun, Fuli Feng, and Tat-Seng Chua. 2024. Denoising Diffusion Recommender Model. arXiv:2401.06982 [cs.IR] https://arxiv.org/abs/2401.06982

[48] Xin Zhou and Zhiqi Shen. 2023. A Tale of Two Graphs: Freezing and Denoising Graph Structures for Multimodal Recommendation. In *Proceedings of the 31st ACM International Conference on Multimedia (MM '23)*. ACM, 935–943. doi:10.1145/3581783.3611943

[49] Yuwei Zhou, Xin Wang, Hong Chen, Yipeng Zhang, Zeyang Zhang, and Wenwu Zhu. 2025. ModuleTeam: Open-Set Multi-Conditional Image Generation with Training-Free Latent Mixture of Any Control Module. In *Proceedings of the 33rd ACM International Conference on Multimedia* (Dublin, Ireland) *(MM '25)*. Association for Computing Machinery, New York, NY, USA, 10467–10475. doi:10.1145/3746027.3755686