MODALITY DISAGREEMENT GUIDED FINE-TUNING FOR MLLM-BASED SEQUENTIAL RECOMMENDATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Multimodal sequential recommendation combines interaction histories with item images and text to model users' preferences. Recent approaches fine-tune multimodal large language models (MLLMs) to rank candidate items from these inputs. However, candidates that are similar in one modality may differ substantially in the other. We ask whether such cross-modal disagreement can serve as a content-based criterion for deciding which comparisons to emphasize during recommendation training. We propose MoDiFy-Rec, a modality-disagreement-guided fine-tuning method that uses these differences to weight candidate comparisons in MLLM-based sequential recommendation. Specifically, MoDiFy-Rec measures disagreement using frozen visual and textual representations. It first maps pairwise similarities to percentiles in their respective training-set distributions, making their relative positions comparable across modalities. An exponential kernel and symmetric normalization then convert the absolute differences between these normalized similarities into fixed positive comparison weights. Following cross-entropy initialization, MoDiFy-Rec fine-tunes the MLLM with full-candidate cross-entropy and a weighted pairwise logistic loss comparing the observed next item against every other candidate. For fixed contexts with positive conditional label probabilities, we show that the combined objective preserves cross-entropy's population minimizers over unrestricted scores for any nonnegative auxiliary coefficient. On three recommendation datasets, MoDiFy-Rec improves ranking quality over cross-entropy fine-tuning and pairwise baselines.

1 INTRODUCTION

Sequential recommendation uses ordered interaction histories to model user preferences and recommend the next item (Kang & McAuley, 2018; Sun et al., 2019; Xie et al., 2022). Large language models (LLMs) support this task through prompts containing user histories and candidate items (Liao et al., 2024b; Yue et al., 2023). Multimodal approaches incorporate item images into user modeling and recommendation (Liu et al., 2024; Zhou et al., 2025; Ye et al., 2025), supported by multimodal large language models (MLLMs) that jointly process visual and textual inputs (Bai et al., 2025; Li et al., 2024). Ranking these candidates requires relating their content to the preferences reflected in the user's history.

Prior work improves recommendation through representation learning and preference optimization. Representation methods align language features with collaborative signals or adapt language models to recommendation (Ren et al., 2024; Ren & Huang, 2025). Building on direct preference optimization (DPO) (Rafailov et al., 2023), recommendation methods introduce personalized objectives and self-play feedback (Liao et al., 2024a; Gao et al., 2025). MLLM alignment methods use factually augmented reward models, style-consistent preference pairs, and input perturbations (Sun et al., 2024; Zhao et al., 2025; Zhou et al., 2024), with adaptive objectives incorporating sample hardness and model responses (Lu et al., 2025). For sequential recommendation, HaNoRec combines content-based hardness estimates and model responsiveness with noise-regularized preference optimization (Wang et al., 2026b). We study whether disagreement between visual and textual candidate similarities provides additional information for weighting ranking comparisons.

Consider two shoes described as white casual sneakers, one plain and the other patterned. Their visual difference may matter to a user whose recent purchases favor plain designs, despite their

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

similar descriptions. Conversely, visually similar items can differ in material or intended use. These examples motivate using cross-modal disagreement to weight candidate comparisons, with observed interactions providing the ranking supervision. Raw visual and textual similarities can follow different distributions, so their direct difference may reflect numerical scale as well as content. We address this measurement issue and design a content-dependent auxiliary loss to share the population target of full-candidate cross-entropy.

We propose MoDiFy-Rec, a modality-disagreement-guided fine-tuning method for MLLM-based sequential recommendation. For candidate sets retrieved without using the next-item label, MoDiFy-Rec computes pairwise cosine similarities from frozen visual and textual representations. It maps each similarity to its percentile in the corresponding training-set distribution, then applies an exponential kernel and symmetric normalization to the absolute percentile differences, yielding fixed, symmetric positive comparison weights. Starting from a full-candidate cross-entropy checkpoint, we update the MLLM's LoRA parameters (Hu et al., 2022) using cross-entropy and a weighted pairwise logistic loss comparing the observed next item with every other candidate. All candidate scores are computed in a shared context containing the user history and item images and text. At inference, the MLLM directly ranks the retrieved candidates by its output scores. For each fixed context with positive conditional label probabilities, we show that the combined objective preserves cross-entropy's population minimizers over unrestricted scores for any nonnegative auxiliary coefficient. This result provides a design constraint shared by all symmetric positive weighting kernels. We also characterize how the weights affect local curvature at the shared optimum and how they enter the finite-sample gradient.

- We propose MoDiFy-Rec, which uses cross-modal disagreement to construct symmetric comparison weights for fine-tuning MLLMs with full-candidate cross-entropy and a pairwise ranking objective.
- We show that, under the stated assumptions, the combined objective preserves cross-entropy's population minimizers for any symmetric positive weights. We also characterize their effects on local curvature at the shared optimum and on the finite-sample gradient.
- On three recommendation datasets, MoDiFy-Rec improves ranking quality over cross-entropy fine-tuning and pairwise baselines.

2 Related Works

2.1 MLLMs for Recommendation

Visual-language models enter recommendation at different points in the pipeline, from producing item representations to predicting the next item (Valencia et al., 2026; Zhang et al., 2025; Wu et al., 2025). VLM4Rec converts images into cached descriptions, encodes them, and retrieves items by cosine similarity to a mean-pooled history profile (Valencia et al., 2026). NoteLLM-2 addresses item-to-item recommendation, using multimodal in-context prompting and gated late fusion with contrastive objectives on visual and multimodal note embeddings (Zhang et al., 2025). For sequential recommendation, Molar builds dynamic user representations from multimodal item embeddings and aligns them contrastively with ID-based user embeddings (Luo et al., 2024).

Direct next-item prediction also admits different input and scoring designs (Wu et al., 2025; Wang et al., 2025a; 2026a). Multimodal information may reach the recommender through visual inputs or through prior textualization (Wu et al., 2025; Wang et al., 2026a). AlmostRec supplies images, text, and projected sequential-recommender ID embeddings to an LMM, training a behavior adapter, LoRA parameters, and an item-output projection with cross-entropy (Wu et al., 2025). MMSRARec summarizes item content into keywords and adds subsequent-item keywords from similar training-set users, then scores each candidate through an MLLM's first-token yes/no probabilities (Wang et al., 2025a). MLLMRec-R1 textualizes visual content offline; its recommender uses supervised fine-tuning followed by GRPO with format and next-item-hit rewards, after filtering reasoning traces for modality and prediction consistency (Wang et al., 2026a).

2.2 Preference Alignment in MLLMs

DPO derives a preference classification objective from RLHF (Rafailov et al., 2023). RLHF-V learns through dense preference optimization from human corrections to hallucinated segments (Yu et al., 2024). OPA-DPO aligns expert revisions with the reference policy (Yang et al., 2025). RLAIF-V assesses atomic claims with open-source AI feedback and refreshes preferences between DPO iterations (Yu et al., 2025). mDPO adds image preferences and a chosen-response reward anchor (Wang et al., 2024). ASPO weights chosen-response sentence rewards by similarity between images and text and textual perplexity (Wang et al., 2025b). DAMA adapts DPO's coefficient using hardness derived from image-response similarity and current implicit reward gaps (Lu et al., 2025), while HaNoRec combines content-derived hardness, model responsiveness, and noise regularization for multimodal sequential recommendation (Wang et al., 2026b). SymMPO's input-response symmetry reverses preferred responses across contrastive image contexts and regularizes margin consistency (Liu et al., 2025). JoMPO combines response preferences with listwise visual-context rankings, selecting image pairs with high CLIP and low DINOv2 similarity (Kim & Yoon, 2026).

3 Method

MoDiFy-Rec uses disagreement between visual and textual candidate relations to weight additional ranking comparisons. The weights are computed from frozen item representations, while the multimodal large language model (MLLM) learns user-dependent scores from the interaction history and item content. Figure 1 summarizes the training pipeline.

3.1 Task Formulation and Candidate Scoring

Given an ordered interaction history H, we rank a retrieved candidate set $\mathcal{C}(H)$ of $M \geq 2$ items to predict the next interaction. The candidate set and its ordering are fixed by a retriever that does not use the next-item label (Section 4.1). Training uses events whose observed next item belongs to this set; $i \in \{1, \ldots, M\}$ denotes its candidate index. For each event, x is the shared context containing the ordered history, the images and text of the history and candidate items, the candidate identifiers, and the task instruction. Each candidate has a distinct single-token identifier. The MLLM produces its score s_i as the identifier's logit at a common answer position, so all candidate comparisons use the same context and score vector. The candidate probabilities are normalized over the current candidate set only,

$$\pi_\theta(i \mid x) = \frac{\exp(s_i)}{\sum_{k=1}^{M} \exp(s_k)},$$

where θ denotes the trainable recommender parameters. This distribution defines the full-candidate cross-entropy used in both training stages; ranking uses s_i alone.

3.2 Cross-Modal Disagreement

Let v_i and t_i denote the frozen visual and textual representations of candidate i. For each distinct candidate pair, we compute the cosine similarities $c_{ij}^V = \cos(v_i, v_j)$ and $c_{ij}^T = \cos(t_i, t_j)$. To compare their relative positions across modalities, we map each similarity to its percentile rank in the corresponding modality's training-set distribution, obtaining $\hat{c}_{ij}^V, \hat{c}_{ij}^T \in [0, 1]$: $\hat{c}_{ij}^V = F_V(c_{ij}^V)$, where F_V is the empirical cumulative distribution function of visual similarities over all candidate pairs of the training events (its numerical implementation is given in Appendix A), and $\hat{c}_{ij}^T = F_T(c_{ij}^T)$ likewise. Both mappings are estimated once from training data and remain fixed throughout fine-tuning. We define disagreement as

$$d_{ij} = \left|\hat{c}_{ij}^V - \hat{c}_{ij}^T\right|. \tag{1}$$

A large d_{ij} indicates that the pair occupies different relative similarity positions in the two modalities. This definition covers both visually similar items with different textual attributes and textually similar items with different appearances. This content relation supplies only the comparison weights, and the observed interaction determines the direction of ranking supervision. The relation does not express user preference or the confidence of the model being trained.

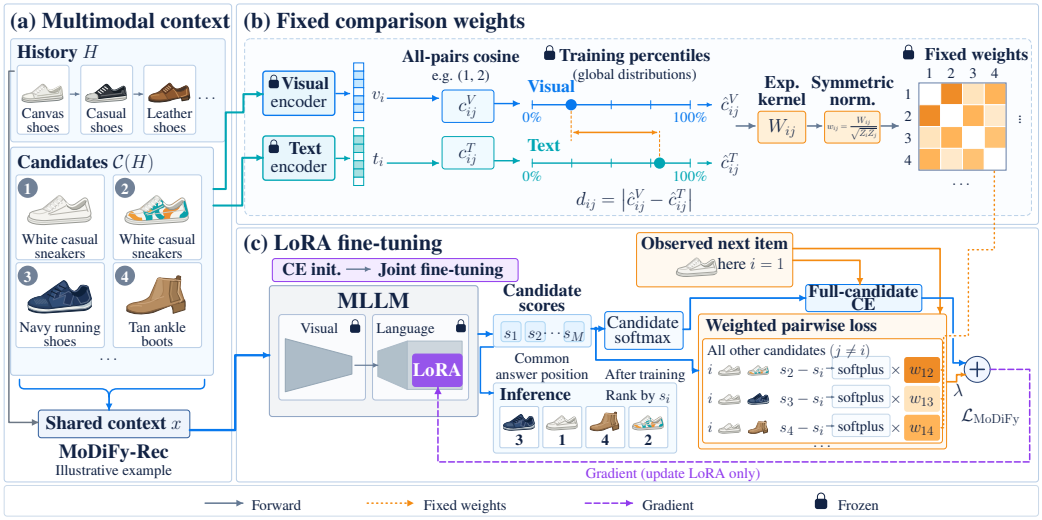


Figure 1: Overview of MoDiFy-Rec. Pairwise visual and textual similarities from frozen representations are mapped to training-set percentiles, and their cross-modal disagreement is converted into fixed symmetric comparison weights. Following cross-entropy initialization, full-candidate cross-entropy and a weighted pairwise logistic loss jointly update the MLLM's LoRA parameters, while inference ranks candidates directly by model scores conditioned on the interaction history and candidate content.

3.3 Symmetric Comparison Weights

For a temperature $\tau > 0$, we transform disagreement into positive kernel values and normalize them symmetrically:

$$W_{ij} = \exp(d_{ij}/\tau), \qquad Z_i = \sum_{k \neq i} W_{ik}, \qquad w_{ij} = \frac{W_{ij}}{\sqrt{Z_i Z_j}}, \qquad w_{ij} = w_{ji} > 0. \tag{2}$$

All sums and weights concern distinct candidates; W_{ii} is never used. The exponential kernel assigns greater raw weight to larger disagreement and retains every comparison with a positive weight. Appendix B reports the resulting weight dynamic range. The denominator accounts for the total kernel mass associated with both candidates, while preserving the coefficient under index exchange; when $Z_i \approx Z_j$ the total weight received by candidate i, $\sum_{j \neq i} w_{ij}$, is close to one, although the rows are not exactly normalized. Consequently, the final weights depend on both pairwise disagreement and the composition of the candidate set. Symmetry makes the offset of Proposition 2 vanish; it is a sufficient condition for the population analysis in Section 3.5, not a necessary one.

For a fixed candidate set, the representations, percentile mappings, and temperature determine all weights before the target index is used in the loss. These quantities remain frozen during fine-tuning. Gradients propagate only through the MLLM scores; the weights receive no gradient and are not updated from the current scores.

3.4 Disagreement-Guided Fine-Tuning

Both training stages update only the language component's LoRA parameters (Hu et al., 2022); the pretrained backbone, including the visual tower, remains frozen. We first obtain a checkpoint by minimizing full-candidate cross-entropy, then continue fine-tuning from this checkpoint with an additional pairwise objective. For a training event (x, i), the auxiliary loss compares the observed next item with every other candidate:

$$\mathcal{L}_{\text{emph}}(x, i) = \sum_{j \neq i} w_{ij} \ \text{softplus}(s_j - s_i). \tag{3}$$

Here $\text{softplus}(z) = \log(1 + \exp(z))$. Each comparison penalizes a competing candidate's score relative to the observed target, with its strength scaled by w_{ij}. The joint objective is

$$\mathcal{L}_{\text{MoDiFy}} = \mathbb{E}_{x,i}\Big[-\log \pi_\theta(i \mid x) + \lambda \, \mathcal{L}_{\text{emph}}(x, i) \Big], \tag{4}$$

where the expectation is over the retained next-item events and $\lambda \geq 0$ controls the auxiliary contribution. Setting $\lambda = 0$ recovers full-candidate cross-entropy. Both terms use the same candidate scores and update the same LoRA parameters.

At inference, the MLLM receives the history and retrieved candidates with their images and text, and ranks the candidates by descending output logits s_i. The disagreement weights enter only the training objective, and no auxiliary quantity is computed at inference.

3.5 Population Analysis

Fix a shared context x, and let $p_i > 0$ be the conditional probability that candidate i is the observed next item, given x and given that the target lies in the retrieved set, with $\sum_i p_i = 1$; p is the logged conditional distribution under the candidate protocol, not an exposure-free preference. In this subsection, population risks are conditioned on x, weights are fixed, and scores are unrestricted real values.

Proposition 1 (Population minimizers)**.** *For any fixed symmetric positive weights $w_{ij} = w_{ji} > 0$, the conditional population risks of $\mathcal{L}_{\mathrm{emph}}$, full-candidate cross-entropy, and $\mathcal{L}_{\mathrm{MoDiFy}}$ for any $\lambda \geq 0$ share the minimizers $s_i = \log p_i + c(x)$. The candidate probabilities and score differences are unique, while $c(x)$ is an arbitrary common constant.*

Proof sketch. Because $w_{ij} = w_{ji}$, the population auxiliary risk decomposes over unordered pairs into strictly convex terms in $G = s_i - s_j$, each minimized uniquely at $G = \log p_i - \log p_j$; these minima are attained simultaneously by $s_i = \log p_i + c(x)$, which also minimizes cross-entropy, and positivity of every w_{ij} forces any minimizer to attain them all (here $\sigma(G) = (1 + \exp(-G))^{-1}$). The full proof is in Appendix E.2; the statement holds for every symmetric positive kernel, including the uniform kernel of Section 4.1, so it constrains the design rather than separating kernels.

Local curvature. At the shared optimum, the second derivative of the auxiliary risk for pair (i, j) with respect to G equals $w_{ij} p_i p_j / (p_i + p_j)$. For any score perturbation $u \in \mathbb{R}^M$, the Hessian of the conditional population risk, evaluated at $s_i = \log p_i + c(x)$ and taken with respect to the candidate scores, satisfies

$$u^\top \nabla^2 \mathcal{L}_{\mathrm{MoDiFy}}\, u = \sum_{i<j} \Big[p_i p_j + \lambda\, w_{ij} \frac{p_i p_j}{p_i + p_j} \Big] (u_i - u_j)^2. \tag{5}$$

Eq. (5) describes the risk only locally at the shared unrestricted population optimum: it gives the exact Hessian with respect to candidate scores, not LoRA parameters (derivation in Appendix E.3). The first term is the cross-entropy contribution; the second makes the local geometry of score differences depend on the comparison weights while preserving that optimum.

Under the restricted LoRA parameterization, the auxiliary term can affect both the optimization trajectory and the attainable optimum. Its effect on recommendation quality is an empirical question.

Training observes one target i^* per event. The empirical objective over unrestricted per-event scores is therefore minimized only in the limit and is not described by Proposition 1. The weights change the gradient for every competitor $j \neq i^*$:

$$\partial_{s_j} \big[-\log \pi_\theta(i^* \mid x) + \lambda\, \mathcal{L}_{\mathrm{emph}}(x, i^*) \big] = \pi_\theta(j \mid x) + \lambda\, w_{i^* j}\, \sigma(s_j - s_{i^*}). \tag{6}$$

The weighted push on each competitor, the second term, is proportional to $w_{i^* j}$ and decays with the score margin. In finite training, the kernel acts through this gradient while leaving the population target unchanged.

4 Experiments

The experiments answer three pre-specified empirical questions. **Q1**: on clean inputs, does MoDiFy-Rec beat full-candidate cross-entropy (Full-CE) on the high-disagreement slice while remaining non-inferior on the low slice, with a positive high-minus-low difference? **Q2**: does the disagreement kernel beat a uniform kernel, and how does symmetric normalization compare with row normalization with the exact correction? **Q3** (diagnostic): is MoDiFy-Rec more sensitive to test-time modality interventions, in the direction of its clean-input gain?

4.1 SETUP

Data. MicroLens (Ni et al., 2025) is the primary dataset. We use each item's cover image and title, the content fields used by HaNoRec. The secondary datasets are MovieLens-1M (Harper & Konstan, 2016), with poster images and titles plus genres, and Netflix, with titles, years and posters from the collection released by LLMRec (Wei et al., 2024). These are the three datasets used by HaNoRec. The MovieLens-1M release contains no images; Appendix A gives the poster provenance. We apply per-user temporal hold-out (last valid event for test, second-to-last for validation), require every item in scope to have a valid image and text, and expose the most recent 20 history items; no interaction that post-dates the target enters the input. Table 4 (Appendix A) reports the filtered statistics and retrieval-hit events.

Label-agnostic candidates. Two retrieval channels (a frozen visual encoder (Radford et al., 2021) and a frozen sentence-text encoder (Reimers & Gurevych, 2019) over the history prefix) each return the top $M/2$ items; the union is deduplicated, filled alternately, and frozen before the target is read. Training keeps only events whose target lies in the frozen set, and validation and test are scored on the same kind of event (*covered* events; their counts are in Table 4). The same $M = 16$ candidate sets, identifier assignments and permutations are shared by all objectives; a random-candidate task (target plus $M - 1$ random unobserved items) is in Appendix C. Because both channels use the encoders that define d_{ij}, cross-channel pairs may be over-represented among high-disagreement pairs; Appendix B quantifies this.

Backbone and optimization. The reranker is Qwen2.5-VL-3B-Instruct (Bai et al., 2025), with a frozen vision tower and rank-8 LoRA ($\alpha = 32$) on the q/k/v/o projections of the language tower. The identifiers are the single-token letters A–P, read at the shared answer position. History and candidate images use budgets of 64 and 256 visual tokens. We use AdamW (Loshchilov & Hutter, 2019) with a learning rate of 10^{-4} and an effective batch size of 16, training for one shared Full-CE epoch and then two epochs per objective. The loss of Eqs. (3)–(4) is evaluated in double precision on bf16 logits. Checkpoints are chosen by validation covered NDCG@5, with ties resolved in favor of the earlier checkpoint. We use 3 training seeds per objective and re-draw the shared initialization for each seed. Per-user metrics are averaged over seeds before any statistic is computed. We select τ on validation from $\{0.05, 0.1, 0.2, 0.5\}$, choosing $\tau = 0.1$, and set $\lambda = 1$.

Baselines and variants. We compare MoDiFy-Rec with five baselines trained under the identical protocol, and with ablation variants of its own kernel. *Full-CE* ($\lambda = 0$). *Uniform-Pair*: Eq. (3) with $w_{ij} = 1/(M-1)$, approximately matching the mean row sum of MoDiFy-Rec. *Global-Sym*: $w_{ij} = W_{ij}/\sum_{k<l} W_{kl}$ as pre-specified; since its row sums are about $2/M$ of MoDiFy-Rec's, a single λ confounds weight structure with auxiliary strength, so we also report a *mass-matched* Global-Sym whose λ equalizes the auxiliary mass per event (defined in Appendix A). *Row-Corrected*: $q_{ij} = W_{ij}/Z_i$ with the exact $\log(q_{ij}/q_{ji})$ correction (Proposition 2 in Appendix E.4). *Visual-only* and *text-only* kernels replace d_{ij} with $1 - \hat{c}^V_{ij}$ and $1 - \hat{c}^T_{ij}$. We train *Standard-DPO* (Rafailov et al., 2023) and *HaNoRec* (Wang et al., 2026b) on target-versus-candidate pairs with the shared initialization as reference.

Metrics and statistics. The primary metric is NDCG@5 on covered events; Recall@5, NDCG@3 and Recall@3 are supplementary (Recall@K equals Hit Rate@K here). Disagreement slices are terciles of the median pairwise d_{ij} of each candidate set, fixed before any test result is seen. All intervals are user-level paired bootstrap intervals with 10,000 resamples. The pre-specified thresholds are a relative high-slice gain $\Delta_{\text{high}} \geq 5\%$, a non-inferiority margin $\Delta_{\text{NI}} = 1\%$ on the low slice ($\text{LCB}_{95}(\Delta_{\text{low}}) > -1\%$), and a positive lower confidence bound for $\Delta_{\text{high}} - \Delta_{\text{low}}$. The primary family is MoDiFy-Rec versus the three pre-specified baselines (Full-CE, Uniform-Pair, Global-Sym at $\lambda = 1$) on high-slice NDCG@5 with Holm step-down correction; a secondary family compares MoDiFy-Rec with the three closest methods (mass-matched Global-Sym, Row-Corrected and HaNoRec) under its own Holm correction. Intervals reflect test-user sampling of the seed-averaged per-user metrics; seed-to-seed variability is reported separately in Table 1.

Table 1: Main comparison under the frozen label-agnostic candidate protocol. Metrics are computed on covered test events, whose target lies in the 16 retrieved candidates. All objectives share initialization, events, candidates and identifier permutations; Global-Sym is shown in its mass-matched form (the $\lambda = 1$ form is in Table 3). Best per column in bold. Values are means over 3 seeds with the across-seed standard deviation in small type; Table 6 lists the per-seed primary metric.

Objective	NDCG@5	Recall@5	NDCG@3	Recall@3
MicroLens				
Full-CE	0.4337±.0050	0.6501±.0043	0.3608±.0069	0.4727±.0089
Uniform-Pair	0.4381±.0051	0.6560±.0058	0.3657±.0051	0.4799±.0061
Global-Sym (mass-matched)	0.4408±.0069	0.6566±.0093	0.3692±.0071	0.4822±.0098
Standard-DPO	0.4307±.0061	0.6463±.0054	0.3575±.0067	0.4682±.0066
HaNoRec	0.4405±.0066	0.6558±.0070	0.3686±.0060	0.4809±.0054
MoDiFy-Rec (ours)	**0.4475**±.0064	**0.6649**±.0073	**0.3745**±.0067	**0.4868**±.0076
MovieLens-1M				
Full-CE	0.4147±.0095	0.6734±.0143	0.3241±.0065	0.4526±.0063
Uniform-Pair	0.4207±.0041	0.6821±.0052	0.3306±.0046	**0.4629**±.0062
Global-Sym (mass-matched)	0.4244±.0051	0.6856±.0051	**0.3318**±.0047	0.4606±.0048
Standard-DPO	0.4116±.0095	0.6700±.0123	0.3194±.0088	0.4454±.0105
HaNoRec	0.4246±.0066	0.6899±.0075	0.3295±.0062	0.4585±.0064
MoDiFy-Rec (ours)	**0.4259**±.0121	**0.6932**±.0163	0.3314±.0115	**0.4629**±.0149
Netflix				
Full-CE	0.4657±.0045	0.6605±.0068	0.4034±.0031	0.5086±.0035
Uniform-Pair	0.4702±.0027	0.6667±.0033	0.4057±.0052	0.5097±.0070
Global-Sym (mass-matched)	0.4759±.0041	0.6697±.0042	0.4142±.0024	0.5194±.0020
Standard-DPO	0.4657±.0028	0.6565±.0040	0.4043±.0033	0.5068±.0051
HaNoRec	0.4770±.0022	0.6705±.0026	0.4149±.0017	0.5193±.0009
MoDiFy-Rec (ours)	**0.4788**±.0039	**0.6745**±.0034	**0.4164**±.0042	**0.5223**±.0046

4.2 Does disagreement weighting improve ranking? (Q1)

On MicroLens, every objective except Standard-DPO improves on Full-CE (Table 1). The covered NDCG@5 gains are 3.2% for MoDiFy-Rec, 1.6% for HaNoRec, 1.6% for mass-matched Global-Sym and 1.0% for Uniform-Pair. Standard-DPO stays slightly below the baseline. The supplementary metrics move in the same direction for every objective. MoDiFy-Rec is also first on the primary metric on MovieLens-1M and Netflix, where it gains 2.7% and 2.8%, while the baselines change places within their seed variation. On MovieLens-1M, the supplementary metrics do not separate MoDiFy-Rec from the best baseline.

Seed replication supports the ranking: the MoDiFy-Rec−Full-CE difference is positive in every one of the 9 dataset–seed runs (0.7–1.7 NDCG points; Table 6), and MoDiFy-Rec ranks first in 7 of them (mass-matched Global-Sym leads one MovieLens-1M seed and one Netflix seed; Appendix A lists the per-seed differences from the two closest baselines). Because every objective within a seed starts from that seed's initialization, the across-seed standard deviation of the paired MoDiFy-Rec−Full-CE difference is smaller than the single-objective standard deviations of Full-CE and MoDiFy-Rec in Table 1. The paired standard deviations are 0.30, 0.54 and 0.35 NDCG points on the three datasets. The single-objective standard deviations are comparable to the differences among the baselines, so the relative ordering of the baselines is indicative only.

Relative to Full-CE, MoDiFy-Rec improves the high-disagreement tercile by 6.7%, the mid tercile by 2.1% and the low tercile by only 0.8% (Table 2 and Figure 3). The gain is concentrated where the two modalities disagree, and the high-minus-low interval excludes zero. All three pre-specified conditions therefore hold on MicroLens. On MovieLens-1M the high-slice gain (6.9%) and the high-minus-low contrast replicate, but the covered test set is a third of the MicroLens size and the low-slice interval is too wide to establish non-inferiority; on Netflix all three conditions hold again (high-slice gain 5.3%). Q1 is thus met in full on two datasets and in two of three conditions on the third (Table 7). The high-slice gains on MovieLens-1M and Netflix exceed the pre-specified threshold, but their lower bounds fall below it, making this part of the conclusion sensitive to the threshold choice. No baseline satisfies all three conditions: HaNoRec shows the same concentration at a smaller magnitude (3.4% on the high slice), and Standard-DPO improves on no slice.

Table 2: Slice analysis on MicroLens (relative NDCG@5 change versus Full-CE in %, with 95% user-level paired bootstrap intervals). Slices are terciles of the median pairwise disagreement of each candidate set, fixed before evaluation (high/mid/low sizes: 1,918/1,912/1,912). The last column checks the three pre-specified Q1 conditions in order: high-slice gain $\geq 5\%$ with a positive lower bound; low-slice lower bound above -1%; positive lower bound on the high-minus-low difference. Intervals are rounded to one decimal; the condition column uses unrounded bounds (MoDiFy-Rec's low-slice lower bound is -0.27%).

Objective	Overall	High	Mid	Low	High − Low	Q1
Uniform-Pair	+1.0 [+0.4, +1.6]	+0.9 [-0.2, +2.1]	+1.0 [-0.1, +2.0]	+1.1 [+0.0, +2.2]	-0.1 [-1.7, +1.5]	✗✓✗
Global-Sym (λ=1)	+1.1 [+0.4, +1.7]	+2.5 [+1.4, +3.7]	+0.9 [-0.2, +1.9]	-0.1 [-1.2, +1.0]	+2.6 [+1.0, +4.3]	✗✗✓
Global-Sym (mass-matched)	+1.6 [+1.0, +2.3]	+3.7 [+2.5, +4.9]	+0.9 [-0.1, +2.0]	+0.3 [-0.8, +1.4]	+3.4 [+1.8, +5.0]	✗✓✓
Standard-DPO	-0.7 [-1.3, -0.1]	-1.1 [-2.2, +0.1]	-0.9 [-2.0, +0.1]	-0.1 [-1.2, +1.0]	-1.0 [-2.5, +0.6]	✗✗✗
HaNoRec	+1.6 [+0.9, +2.2]	+3.4 [+2.2, +4.5]	+0.3 [-0.8, +1.3]	+1.1 [+0.0, +2.2]	+2.3 [+0.6, +3.9]	✗✓✓
MoDiFy-Rec (ours)	+3.2 [+2.5, +3.8]	+6.7 [+5.4, +8.1]	+2.1 [+1.1, +3.2]	+0.8 [-0.3, +1.9]	+5.9 [+4.2, +7.6]	✓✓✓

Table 3: Kernel and normalization ablations on MicroLens. Absolute covered NDCG@5 (overall, high slice, low slice) and relative change versus Full-CE (%, 95% intervals). Global-Sym with $\lambda = 1$ carries about $2/M$ of MoDiFy-Rec's auxiliary mass; mass-matched rescales λ to equalize it. Per-set mass-matched Uniform-Pair equalizes the auxiliary mass of every event with that of MoDiFy-Rec; the raw-difference kernel applies Eq. (2) to $|c_{ij}^V - c_{ij}^T|$ without the percentile transform; the permuted-weight variant applies a candidate permutation P, drawn before the target is read, to the normalized weight matrix, $w' = PwP^\top$, which keeps the weight values and row sums but detaches them from the candidates.

	NDCG@5 (absolute)			Δ vs. Full-CE (%)		
Objective	Overall	High	Low	Overall	High	Low
Full-CE	0.4337	0.4245	0.4413	—	—	—
Uniform-Pair	0.4381	0.4285	0.4461	+1.0 [+0.4, +1.6]	+0.9 [-0.2, +2.1]	+1.1 [+0.0, +2.2]
Uniform-Pair (per-set mass-matched)	0.4369	0.4275	0.4450	+0.7 [+0.1, +1.4]	+0.7 [-0.4, +1.9]	+0.8 [-0.3, +1.9]
Global-Sym (λ=1)	0.4384	0.4352	0.4407	+1.1 [+0.4, +1.7]	+2.5 [+1.4, +3.7]	-0.1 [-1.2, +1.0]
Global-Sym (mass-matched)	0.4408	0.4402	0.4427	+1.6 [+1.0, +2.3]	+3.7 [+2.5, +4.9]	+0.3 [-0.8, +1.4]
Row-Corrected	0.4439	0.4493	0.4415	+2.3 [+1.7, +3.0]	+5.8 [+4.6, +7.1]	+0.0 [-1.0, +1.1]
Visual-only kernel	0.4384	0.4328	0.4435	+1.1 [+0.5, +1.7]	+2.0 [+0.8, +3.1]	+0.5 [-0.6, +1.6]
Text-only kernel	0.4376	0.4347	0.4411	+0.9 [+0.2, +1.5]	+2.4 [+1.3, +3.5]	-0.1 [-1.1, +1.0]
Raw-difference kernel (no percentile transform)	0.4415	0.4417	0.4438	+1.8 [+1.2, +2.4]	+4.1 [+2.9, +5.3]	+0.6 [-0.5, +1.6]
Permuted weights ($w' = PwP^\top$)	0.4332	0.4251	0.4406	-0.1 [-0.8, +0.5]	+0.1 [-1.0, +1.3]	-0.2 [-1.2, +0.9]
MoDiFy-Rec, gated d_{ij}^{gate} (Eq. 7)	0.4460	0.4501	0.4445	+2.8 [+2.2, +3.5]	+6.0 [+4.8, +7.3]	+0.7 [-0.4, +1.8]
MoDiFy-Rec (ours)	0.4475	0.4530	0.4449	+3.2 [+2.5, +3.8]	+6.7 [+5.4, +8.1]	+0.8 [-0.3, +1.9]

Full-CE itself is only slightly weaker on the high slice than on the low slice. Table 14 conditions on a pre-treatment difficulty proxy, the target's margin at the shared initialization, and shows the largest gain over Full-CE in the high-disagreement column of every difficulty tercile (+2.2, +2.9 and +3.4 NDCG points on hard, medium and easy sets), whereas the low-disagreement column reaches at most 1.0 points. The high-disagreement advantage persists within each tercile of the initialization margin. Disagreement retains explanatory power given difficulty, although it is also partly a difficulty proxy: Full-CE's own score falls with set-level disagreement, and the content-weighted baselines also gain more on the high slice than on the low slice, whereas the uniform kernel does not (Table 2).

4.3 Is the disagreement kernel and its normalization necessary? (Q2)

A uniform pairwise term with approximately the same mean row sum gains only 0.9% on the high slice against 6.7% for MoDiFy-Rec, a significant difference (Table 3; Holm-corrected family tests in Table 9). Generic pairwise regularization does not account for the effect. Single-modality kernels gain 2.0% and 2.4%, raw cosine differences without the percentile transform 4.1%, and permuting the normalized weight matrix by a candidate permutation (same weight values and row sums, detached from the candidates) leaves 0.1%: the disagreeing pairs themselves carry the effect, not the weight shape or a single-modality distance. Eq. (1) was fixed as the definition of disagreement before any experiment; the gated variant (Eq. 7), reported as an ablation only, gains 6.0%.

Global-Sym at $\lambda = 1$ gains 2.5% on the high slice and nothing on the low slice (-0.1%). Matching its auxiliary mass to MoDiFy-Rec's raises the high-slice gain to 3.7% at no low-slice cost (+0.3%), yet it stays significantly below MoDiFy-Rec. The two symmetric kernels thus differ for two separable reasons: auxiliary strength explains part of the difference and per-endpoint normalization the rest. The row sums of MoDiFy-Rec are flat across slices (Table 13), and a uniform kernel with MoDiFy-Rec's auxiliary mass matched event by event gains only 0.7% on the high slice. Auxiliary mass therefore does not drive the slice pattern.

Row-Corrected gains 5.8% on the high slice; its difference from MoDiFy-Rec is not significant, which does not establish equivalence (the interval in Table 9 also admits a small advantage for MoDiFy-Rec). Both share the population target (Proposition 2); we keep the symmetric form because it needs no correction term.

Because both retrieval channels use the encoders that define d_{ij}, Table 11 checks whether high disagreement mainly marks cross-channel pairs. Cross-channel pairs are over-represented in the top disagreement decile and receive more than their share of the auxiliary mass, while accounting for only part of each. Because the cross-channel fraction of a candidate set is nearly constant by construction (each channel contributes $M/2$ candidates), the slice pattern cannot be an effect of the cross-channel pair count. MoDiFy-Rec improves on Full-CE by 2.2% to 3.5% on events retrieved by the visual channel only, the textual channel only, or both (Table 12), so the gain does not depend on which channel retrieved the target. A trained-retriever protocol remains untested.

4.4 Are the gains content-dependent? (Q3)

Both models degrade under every modality intervention (Figure 2 and Table 8, both in the appendix). Text interventions hurt more than image interventions for either model, consistent with text carrying more of the recommendation signal (Ye et al., 2026). The MoDiFy-Rec−Full-CE gap is 1.38 NDCG points on clean inputs. Replacing candidate images shrinks it to between 0.45 (cross-candidate permutation) and 1.00 (constant image) points, whereas replacing candidate texts leaves it at 1.09–1.52 points. Equivalently, MoDiFy-Rec loses 9.2% of its clean score against 7.4% for Full-CE when images are permuted, but the two lose nearly the same share when texts are permuted.

As the construction anticipates, the advantage rests on the candidates' actual images being present and correctly paired, and it appears on the high-disagreement events where the clean gain concentrates. Because the interventions shift the input distribution, we treat the results as a consistency check and do not infer stronger visual reasoning. The larger relative drop under image interventions is also a deployment cost when images are missing or replaced, although MoDiFy-Rec stays above Full-CE under every intervention.

Sensitivity. Figure 5 and Table 10 in the appendix report sensitivity to λ and τ. Halving λ keeps every slice non-negative but lowers the high-slice gain to 4.1%. Doubling λ or sharpening the kernel to $\tau = 0.05$ retains most of the high-slice gain (5.3% and 4.9%) but pushes the low slice past the non-inferiority margin (lower bound -2.0% at $\lambda = 2$). A flat kernel ($\tau = 0.5$) gains only 3.0% on the high slice and moves toward Uniform-Pair. Low-slice non-inferiority, not the high-slice gain, is the binding constraint. The validation column confirms that the main setting is the validation maximum, so the test ordering does not rest on test-set selection.

5 Conclusion

MoDiFy-Rec adds a symmetric, disagreement-weighted pairwise term to full-candidate cross-entropy; under the stated assumptions, the symmetric auxiliary term preserves the population minimizers of cross-entropy and changes the score-space Hessian at those minimizers, so no probability-ratio correction is needed. Under a pre-specified protocol, its gain over cross-entropy, uniform and globally normalized kernels concentrates on high-disagreement sets, including after conditioning on pre-treatment difficulty. On low-disagreement sets, non-inferiority holds on MicroLens and Netflix and remains inconclusive on MovieLens-1M. The gain shrinks most under test-time image interventions. A trained-retriever protocol and an external baseline are the next step.

AI use statement

We used a large language model as a writing assistant for language editing of the manuscript and for drafting the plotting scripts, both of which were reviewed and edited by the authors. We did not use generative AI to generate research ideas, to design experiments, or to produce results. We take responsibility for the final content of this work, including text, claims and artifacts produced with the aid of generative AI.

Ethics statement

The datasets are public research benchmarks of user–item interactions with item images and text; we use them under their released terms and do not attempt to re-identify users. The method does not add user attributes, modality-need labels or inferred demographics; the weights are functions of item content only. As with any recommender objective, emphasizing content-relation differences could shift exposure across items, and we report slice-level results in full so that non-uniform effects are visible rather than averaged away. We see no additional risk beyond those of standard sequential recommendation.

Reproducibility statement

Sections 3 and 4.1 give the complete objective (Eqs. 1–4), backbone, LoRA configuration, optimizer, candidate protocol, identifier interface, metrics and statistical procedure. Appendix E contains the full proof of Proposition 1, the derivation of Eq. (5), the row-normalization result behind the Row-Corrected ablation; Appendix A lists data filtering, split, retrieval and percentile-transform details, the gated variant, and the run manifest fields we freeze (split, eligibility filter, candidate version, percentile file, feature checkpoints, identifier token IDs, prompt template, model revision, LoRA configuration and checkpoint selection log); Appendix B reports the retrieval-channel, weight, difficulty and retrieval-source diagnostics; Table 6 lists the per-seed primary metric of every main-table objective; Appendix D reports compute. The supplementary material contains the training and evaluation code, the frozen candidate lists, the percentile transform files, the per-user and per-seed evaluation outputs and the bootstrap scripts used to produce every table and figure.

References

Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, Humen Zhong, Yuanzhi Zhu, Ming-Hsuan Yang, Zhaohai Li, Jianqiang Wan, Pengfei Wang, Wei Ding, Zheren Fu, Yiheng Xu, Jiabo Ye, Xi Zhang, Tianbao Xie, Zesen Cheng, Hang Zhang, Zhibo Yang, Haiyang Xu, and Junyang Lin. Qwen2.5-vl technical report. *CoRR*, abs/2502.13923, 2025. doi: 10.48550/ARXIV.2502.13923. URL https://doi.org/10.48550/arXiv.2502.13923.

Chongming Gao, Ruijun Chen, Shuai Yuan, Kexin Huang, Yuanqing Yu, and Xiangnan He. Sprec: Self-play to debias llm-based recommendation. In Guodong Long, Michael Blumenstein, Yi Chang, Liane Lewin-Eytan, Zi Helen Huang, and Elad Yom-Tov (eds.), *Proceedings of the ACM on Web Conference 2025, WWW 2025, Sydney, NSW, Australia, 28 April 2025- 2 May 2025*, pp. 5075–5084. ACM, 2025. doi: 10.1145/3696410.3714524. URL https://doi.org/10.1145/3696410.3714524.

F. Maxwell Harper and Joseph A. Konstan. The movielens datasets: History and context. *ACM Trans. Interact. Intell. Syst.*, 5(4):19:1–19:19, 2016. doi: 10.1145/2827872. URL https://doi.org/10.1145/2827872.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022. URL https://openreview.net/forum?id=nZeVKeeFYf9.

Wang-Cheng Kang and Julian J. McAuley. Self-attentive sequential recommendation. In *IEEE International Conference on Data Mining, ICDM 2018, Singapore, November 17-20, 2018*, pp. 197–206. IEEE Computer Society, 2018. doi: 10.1109/ICDM.2018.00035. URL https://doi.org/10.1109/ICDM.2018.00035.

Jiwon Kim and Hyunsoo Yoon. Joint multimodal preference optimization for fine-grained visual-textual alignment. In Vera Demberg, Kentaro Inui, and Lluís Marquez (eds.), *Findings of the Association for Computational Linguistics: EACL 2026, Rabat, Morocco, March 24-29, 2026*, Findings of ACL, pp. 79–94. Association for Computational Linguistics, 2026. doi: 10.18653/V1/2026.FINDINGS-EACL.5. URL https://doi.org/10.18653/v1/2026.findings-eacl.5.

Feng Li, Renrui Zhang, Hao Zhang, Yuanhan Zhang, Bo Li, Wei Li, Zejun Ma, and Chunyuan Li. Llava-next-interleave: Tackling multi-image, video, and 3d in large multimodal models. *CoRR*, abs/2407.07895, 2024. doi: 10.48550/ARXIV.2407.07895. URL https://doi.org/10.48550/arXiv.2407.07895.

Jiayi Liao, Xiangnan He, Ruobing Xie, Jiancan Wu, Yancheng Yuan, Xingwu Sun, Zhanhui Kang, and Xiang Wang. Rosepo: Aligning llm-based recommenders with human values. *CoRR*, abs/2410.12519, 2024a. doi: 10.48550/ARXIV.2410.12519. URL https://doi.org/10.48550/arXiv.2410.12519.

Jiayi Liao, Sihang Li, Zhengyi Yang, Jiancan Wu, Yancheng Yuan, Xiang Wang, and Xiangnan He. Llara: Large language-recommendation assistant. In Grace Hui Yang, Hongning Wang, Sam Han, Claudia Hauff, Guido Zuccon, and Yi Zhang (eds.), *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2024, Washington DC, USA, July 14-18, 2024*, pp. 1785–1795. ACM, 2024b. doi: 10.1145/3626772.3657690. URL https://doi.org/10.1145/3626772.3657690.

Wenqi Liu, Xuemeng Song, Jiaxi Li, Yinwei Wei, Na Zheng, Jianhua Yin, and Liqiang Nie. Mitigating hallucination through theory-consistent symmetric multimodal preference optimization. In Danielle Belgrave, Cheng Zhang, Laura N. Montoya, Hsuan-Tien Lin, Razvan Pascanu, Piotr Koniusz, Marzyeh Ghassemi, Nancy Chen, Iván Vladimir Meza Ruíz, and Arturo Loaiza-Bonilla (eds.), *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2025, NeurIPS 2025, San Diego, CA, USA, December 2-7, 2025 / Mexico City, Mexico, November 30 - December 5, 2025*, 2025. URL http://papers.nips.cc/paper_files/paper/2025/hash/a1718f361df32ff3a1fc224f8673c556-Abstract-Conference.html.

Yuqing Liu, Yu Wang, Lichao Sun, and Philip S. Yu. Rec-gpt4v: Multimodal recommendation with large vision-language models. *CoRR*, abs/2402.08670, 2024. doi: 10.48550/ARXIV.2402.08670. URL https://doi.org/10.48550/arXiv.2402.08670.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019. URL https://openreview.net/forum?id=Bkg6RiCqY7.

Jinda Lu, Junkang Wu, Jinghan Li, Xiaojun Jia, Shuo Wang, Yifan Zhang, Junfeng Fang, Xiang Wang, and Xiangnan He. DAMA: data- and model-aware alignment of multi-modal llms. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Forty-second International Conference on Machine Learning, ICML 2025, Vancouver, BC, Canada, July 13-19, 2025*, volume 267 of *Proceedings of Machine Learning Research*, pp. 40726–40740. PMLR / OpenReview.net, 2025. URL https://proceedings.mlr.press/v267/lu25m.html.

Yucong Luo, Qitao Qin, Hao Zhang, Mingyue Cheng, Ruiran Yan, Kefan Wang, and Jie Ouyang. Molar: Multimodal llms with collaborative filtering alignment for enhanced sequential recommendation. *CoRR*, abs/2412.18176, 2024. doi: 10.48550/ARXIV.2412.18176. URL https://doi.org/10.48550/arXiv.2412.18176.

Yongxin Ni, Yu Cheng, Xiangyan Liu, Junchen Fu, Youhua Li, Xiangnan He, Yongfeng Zhang, and Fajie Yuan. A content-driven micro-video recommendation dataset at scale. In Meeyoung Cha, Chanyoung Park, Noseong Park, Carl Yang, Senjuti Basu Roy, Jessie Li, Jaap Kamps, Kijung Shin, Bryan Hooi, and Lifang He (eds.), *Proceedings of the 34th ACM International Conference on Information and Knowledge Management, CIKM 2025, Seoul, Republic of Korea, November 10-14, 2025*, pp. 6486–6491. ACM, 2025. doi: 10.1145/3746252.3761655. URL https://doi.org/10.1145/3746252.3761655.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, volume 139 of *Proceedings of Machine Learning*

Research, pp. 8748–8763. PMLR, 2021. URL http://proceedings.mlr.press/v139/radford21a.html.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D. Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine (eds.), *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023. URL http://papers.nips.cc/paper_files/paper/2023/hash/a85b405ed65c6477a4fe8302b5e06ce7-Abstract-Conference.html.

Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan (eds.), *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pp. 3980–3990. Association for Computational Linguistics, 2019. doi: 10.18653/V1/D19-1410. URL https://doi.org/10.18653/v1/D19-1410.

Xubin Ren and Chao Huang. Easyrec: Simple yet effective language models for recommendation. In Christos Christodoulopoulos, Tanmoy Chakraborty, Carolyn Rose, and Violet Peng (eds.), *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing, EMNLP 2025, Suzhou, China, November 4-9, 2025*, pp. 17717–17732. Association for Computational Linguistics, 2025. doi: 10.18653/V1/2025.EMNLP-MAIN.894. URL https://doi.org/10.18653/v1/2025.emnlp-main.894.

Xubin Ren, Wei Wei, Lianghao Xia, Lixin Su, Suqi Cheng, Junfeng Wang, Dawei Yin, and Chao Huang. Representation learning with large language models for recommendation. In Tat-Seng Chua, Chong-Wah Ngo, Ravi Kumar, Hady W. Lauw, and Roy Ka-Wei Lee (eds.), *Proceedings of the ACM on Web Conference 2024, WWW 2024, Singapore, May 13-17, 2024*, pp. 3464–3475. ACM, 2024. doi: 10.1145/3589334.3645458. URL https://doi.org/10.1145/3589334.3645458.

Fei Sun, Jun Liu, Jian Wu, Changhua Pei, Xiao Lin, Wenwu Ou, and Peng Jiang. Bert4rec: Sequential recommendation with bidirectional encoder representations from transformer. In Wenwu Zhu, Dacheng Tao, Xueqi Cheng, Peng Cui, Elke A. Rundensteiner, David Carmel, Qi He, and Jeffrey Xu Yu (eds.), *Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM 2019, Beijing, China, November 3-7, 2019*, pp. 1441–1450. ACM, 2019. doi: 10.1145/3357384.3357895. URL https://doi.org/10.1145/3357384.3357895.

Zhiqing Sun, Sheng Shen, Shengcao Cao, Haotian Liu, Chunyuan Li, Yikang Shen, Chuang Gan, Liangyan Gui, Yu-Xiong Wang, Yiming Yang, Kurt Keutzer, and Trevor Darrell. Aligning large multimodal models with factually augmented RLHF. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar (eds.), *Findings of the Association for Computational Linguistics, ACL 2024, Bangkok, Thailand and virtual meeting, August 11-16, 2024*, volume ACL 2024 of *Findings of ACL*, pp. 13088–13110. Association for Computational Linguistics, 2024. doi: 10.18653/V1/2024.FINDINGS-ACL.775. URL https://doi.org/10.18653/v1/2024.findings-acl.775.

Ty Valencia, Burak Barlas, Varun Singhal, Ruchir Bhatia, and Wei Yang. Vlm4rec: Multimodal semantic representation for recommendation with large vision-language models. *CoRR*, abs/2603.12625, 2026. doi: 10.48550/ARXIV.2603.12625. URL https://doi.org/10.48550/arXiv.2603.12625.

Fei Wang, Wenxuan Zhou, James Y. Huang, Nan Xu, Sheng Zhang, Hoifung Poon, and Muhao Chen. mdpo: Conditional preference optimization for multimodal large language models. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen (eds.), *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, EMNLP 2024, Miami, FL, USA, November 12-16, 2024*, pp. 8078–8088. Association for Computational Linguistics, 2024. doi: 10.18653/V1/2024.EMNLP-MAIN.460. URL https://doi.org/10.18653/v1/2024.emnlp-main.460.

Haoyu Wang, Yitong Wang, and Jining Wang. Mmsrarec: Summarization and retrieval augumented sequential recommendation based on multimodal large language model. *CoRR*, abs/2512.20916,

2025a. doi: 10.48550/ARXIV.2512.20916. URL https://doi.org/10.48550/arXiv.2512.20916.

Yeyuan Wang, Dehong Gao, Rujiao Long, Lei Yi, Linbo Jin, Libin Yang, and Xiaoyan Cai. ASPO: adaptive sentence-level preference optimization for fine-grained multimodal reasoning. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar (eds.), *Findings of the Association for Computational Linguistics, ACL 2025, Vienna, Austria, July 27 - August 1, 2025*, volume ACL 2025 of *Findings of ACL*, pp. 5149–5160. Association for Computational Linguistics, 2025b. doi: 10.18653/V1/2025.FINDINGS-ACL.267. URL https://doi.org/10.18653/v1/2025.findings-acl.267.

Yu Wang, Yonghui Yang, Le Wu, Jiancan Wu, Hefei Xu, and Hui Lin. Mllmrec-r1: Incentivizing reasoning capability in large language models for multimodal sequential recommendation. *CoRR*, abs/2603.06243, 2026a. doi: 10.48550/ARXIV.2603.06243. URL https://doi.org/10.48550/arXiv.2603.06243.

Yu Wang, Yonghui Yang, Le Wu, Yi Zhang, Fei Liu, and Richang Hong. Multimodal large language models with adaptive preference optimization for sequential recommendation. In Alistair Moffat, Falk Scholer, Hannah Bast, Marc Najork, and Min Zhang (eds.), *Proceedings of the 49th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2026, MelbourneVICAustralia, July 20-24, 2026*, pp. 1903–1913. ACM, 2026b. doi: 10.1145/3805712.3809590. URL https://doi.org/10.1145/3805712.3809590.

Wei Wei, Xubin Ren, Jiabin Tang, Qinyong Wang, Lixin Su, Suqi Cheng, Junfeng Wang, Dawei Yin, and Chao Huang. LLMRec: large language models with graph augmentation for recommendation. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining, WSDM 2024*, pp. 806–815. ACM, 2024. doi: 10.1145/3616855.3635853.

Zihao Wu, Xin Wang, Heng Chang, Hong Chen, Lifeng Sun, and Wenwu Zhu. Aligning large multimodal model with sequential recommendation via content-behavior guidance. In Zhongfei (Mark) Zhang, Elisa Ricci, Yan Yan, Liqiang Nie, Vincent Oria, and Lamberto Ballan (eds.), *Proceedings of the 2025 International Conference on Multimedia Retrieval, ICMR 2025, Chicago, IL, USA, 30 June 2025 - 3 July 2025*, pp. 1507–1516. ACM, 2025. doi: 10.1145/3731715.3733273. URL https://doi.org/10.1145/3731715.3733273.

Xu Xie, Fei Sun, Zhaoyang Liu, Shiwen Wu, Jinyang Gao, Jiandong Zhang, Bolin Ding, and Bin Cui. Contrastive learning for sequential recommendation. In *38th IEEE International Conference on Data Engineering, ICDE 2022, Kuala Lumpur, Malaysia, May 9-12, 2022*, pp. 1259–1273. IEEE, 2022. doi: 10.1109/ICDE53745.2022.00099. URL https://doi.org/10.1109/ICDE53745.2022.00099.

Zhihe Yang, Xufang Luo, Dongqi Han, Yunjian Xu, and Dongsheng Li. Mitigating hallucinations in large vision-language models via DPO: on-policy data hold the key. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2025, Nashville, TN, USA, June 11-15, 2025*, pp. 10610–10620. Computer Vision Foundation / IEEE, 2025. doi: 10.1109/CVPR52734.2025.00992. URL https://openaccess.thecvf.com/content/CVPR2025/html/Yang_Mitigating_Hallucinations_in_Large_Vision-Language_Models_via_DPO_On-Policy_Data_CVPR_2025_paper.html.

Yu Ye, Junchen Fu, Yu Song, Kaiwen Zheng, and Joemon M. Jose. A reproducibility study of multimodal embeddings for recommender systems. *Int. J. Multim. Inf. Retr.*, 15 (2):14, 2026. doi: 10.1007/S13735-026-00408-9. URL https://doi.org/10.1007/s13735-026-00408-9.

Yuyang Ye, Zhi Zheng, Yishan Shen, Tianshu Wang, Hengruo Zhang, Peijun Zhu, Runlong Yu, Kai Zhang, and Hui Xiong. Harnessing multimodal large language models for multimodal sequential recommendation. In Toby Walsh, Julie Shah, and Zico Kolter (eds.), *Thirty-Ninth AAAI Conference on Artificial Intelligence, Thirty-Seventh Conference on Innovative Applications of Artificial Intelligence, Fifteenth Symposium on Educational Advances in Artificial Intelligence, AAAI 2025, Philadelphia, PA, USA, February 25 - March 4, 2025*, pp. 13069–13077. AAAI Press, 2025. doi: 10.1609/AAAI.V39I12.33426. URL https://doi.org/10.1609/aaai.v39i12.33426.

Tianyu Yu, Yuan Yao, Haoye Zhang, Taiwen He, Yifeng Han, Ganqu Cui, Jinyi Hu, Zhiyuan Liu, Hai-Tao Zheng, and Maosong Sun. RLHF-V: towards trustworthy mllms via behavior alignment from fine-grained correctional human feedback. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2024, Seattle, WA, USA, June 16-22, 2024*, pp. 13807–13816. IEEE, 2024. doi: 10.1109/CVPR52733.2024.01310. URL https://doi.org/10.1109/CVPR52733.2024.01310.

Tianyu Yu, Haoye Zhang, Qiming Li, Qixin Xu, Yuan Yao, Da Chen, Xiaoman Lu, Ganqu Cui, Yunkai Dang, Taiwen He, Xiaocheng Feng, Jun Song, Bo Zheng, Zhiyuan Liu, Tat-Seng Chua, and Maosong Sun. RLAIF-V: open-source AI feedback leads to super GPT-4V trustworthiness. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2025*, pp. 19985–19995, 2025. URL https://openaccess.thecvf.com/content/CVPR2025/html/Yu_RLAIF-V_Open-Source_AI_Feedback_Leads_to_Super_GPT-4V_Trustworthiness_CVPR_2025_paper.html.

Zhenrui Yue, Sara Rabhi, Gabriel de Souza Pereira Moreira, Dong Wang, and Even Oldridge. Llamarec: Two-stage recommendation using large language models for ranking. *CoRR*, abs/2311.02089, 2023. doi: 10.48550/ARXIV.2311.02089. URL https://doi.org/10.48550/arXiv.2311.02089.

Chao Zhang, Haoxin Zhang, Shiwei Wu, Di Wu, Tong Xu, Xiangyu Zhao, Yan Gao, Yao Hu, and Enhong Chen. Notellm-2: Multimodal large representation models for recommendation. In Yizhou Sun, Flavio Chierichetti, Hady W. Lauw, Claudia Perlich, Wee Hyong Tok, and Andrew Tomkins (eds.), *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining, V.1, KDD 2025, Toronto, ON, Canada, August 3-7, 2025*, pp. 2815–2826. ACM, 2025. doi: 10.1145/3690624.3709440. URL https://doi.org/10.1145/3690624.3709440.

Zhiyuan Zhao, Bin Wang, Linke Ouyang, Xiaoyi Dong, Jiaqi Wang, and Conghui He. Beyond multimodal hallucinations: Enhancing lvlms through hallucination-aware direct preference optimization. In *IEEE International Conference on Multimedia and Expo, ICME 2025, Nantes, France, June 30 - July 4, 2025*, pp. 1–6. IEEE, 2025. doi: 10.1109/ICME59968.2025.11209377. URL https://doi.org/10.1109/ICME59968.2025.11209377.

Peilin Zhou, Chao Liu, Jing Ren, Xinfeng Zhou, Yueqi Xie, Meng Cao, Zhongtao Rao, You-Liang Huang, Dading Chong, Junling Liu, Jae Boum Kim, Shoujin Wang, Raymond Chi-Wing Wong, and Sunghun Kim. When large vision language models meet multimodal sequential recommendation: An empirical study. In Guodong Long, Michael Blumenstein, Yi Chang, Liane Lewin-Eytan, Zi Helen Huang, and Elad Yom-Tov (eds.), *Proceedings of the ACM on Web Conference 2025, WWW 2025, Sydney, NSW, Australia, 28 April 2025- 2 May 2025*, pp. 275–292. ACM, 2025. doi: 10.1145/3696410.3714764. URL https://doi.org/10.1145/3696410.3714764.

Yiyang Zhou, Chenhang Cui, Rafael Rafailov, Chelsea Finn, and Huaxiu Yao. Aligning modalities in vision large language models via preference fine-tuning. *CoRR*, abs/2402.11411, 2024. doi: 10.48550/ARXIV.2402.11411. URL https://doi.org/10.48550/arXiv.2402.11411.

A PROTOCOL DETAILS

Table 4: Dataset statistics after content-eligibility filtering, and the number of training prefixes before and after the label-agnostic retrieval hit filter. Only hit events enter the candidate loss. Validation and test hold out one event per user; their hit counts are the covered events on which Tables 1–10 are computed.

Dataset	Users	Items	Interactions	Train prefixes	Train hits	Val hits	Test hits	Text field
MicroLens	24,873	39,562	214,905	68,412	15,666	5,820	5,742	title
MovieLens-1M	6,034	3,706	998,631	18,102	5,648	1,894	1,883	title + genres
Netflix	12,914	17,102	67,485	28,411	7,216	3,371	3,332	title + year

Filtering and split. Items without a valid image file or with an empty title are removed, then users with fewer than five remaining interactions. Per user, the last valid event is the test event, the second-to-last is validation, and training prefixes end at each of the last three training events with the full preceding history, so a user with r training events contributes $\min(3, r-1)$ prefixes (Table 4).

MicroLens histories are short: the median training history has six items, about a quarter of the users have exactly three training events and contribute two prefixes, and roughly two fifths have fewer than five history items. Netflix histories are shorter still: the median is three, about four fifths of the users have exactly three training events, and roughly nineteen in twenty have fewer than five history items. The 20-item window is therefore rarely filled on either dataset; MovieLens-1M histories fill it.

Item metadata is taken from the dataset release for MicroLens, which is the MicroLens split used by HaNoRec (25,411 users and 41,081 items before our eligibility filter), not the MicroLens-100K release, whose user sample and catalog differ. MovieLens-1M contains no images: posters were matched by title and year from a public film-poster source and frozen as a snapshot whose source, date and per-item hit list are recorded in the run manifest, and items without a poster are removed by the eligibility filter. Netflix interactions, titles, years and posters are taken as released by LLMRec (Wei et al., 2024), the same collection HaNoRec uses; because the posters post-date the interactions and the historical availability of static metadata cannot be verified, we treat static-content evaluation as a limitation.

Retrieval and percentile transform. Both channels embed the history prefix by mean-pooling item embeddings of the most recent 20 items and retrieve by cosine similarity against the catalog snapshot at the split time, excluding history items; ties are broken by item ID. The percentile transforms F_V and F_T of Section 3.2 are the empirical cumulative distribution functions of the pairwise similarities over all unordered pairs within the training candidate sets (1,000 quantile knots, linear interpolation, ties averaged), stored with the run manifest. Visual features are CLIP ViT-L/14 image embeddings (Radford et al., 2021); textual features are Sentence-BERT embeddings (Reimers & Gurevych, 2019) of the text field. The same frozen features serve retrieval and disagreement, so no additional encoder is introduced; Appendix B quantifies the coupling this creates between retrieval channel and disagreement.

Gated disagreement variant. Table 3 compares Eq. (1) with a gated alternative that keeps the disagreement bonus only for pairs that are close in at least one modality,

$$d_{ij}^{\text{gate}} = \mathbf{1}\{\max(\hat{c}_{ij}^{V}, \hat{c}_{ij}^{T}) \geq \alpha\} \cdot d_{ij}, \qquad \alpha = 0.7. \tag{7}$$

Gating sets the disagreement of the remaining pairs to zero; the kernel of Eq. (2) then still assigns them a positive baseline weight, so no comparison is deleted. The gated variant appears only in Table 3.

Auxiliary mass. The auxiliary mass of a training event is the target-row sum $\sum_{j \neq i^*} w_{i^* j}$ of the weights that enter Eq. (3). Mass-matched Global-Sym sets λ so that the mean of this quantity over training events equals that of MoDiFy-Rec; per-set mass-matched Uniform-Pair equalizes it event by event; Uniform-Pair with $w_{ij} = 1/(M-1)$ has a target-row sum of one, close to the mean row sum of MoDiFy-Rec (Table 13).

Prompt and identifiers. The prompt lists the history items (image then title), then the candidates as "[letter]: image, title", and asks for the letter of the item the user will interact with next; the

candidate logits are read at the position of the answer letter. The mapping from candidates to letters is a pre-stored label-independent permutation shared by all objectives; permuting candidates permutes the target index and the weight matrix consistently.

Relation to HaNoRec's published numbers. HaNoRec reports HR@3 and NDCG@3 over the target plus nine random negatives, whereas our main protocol scores 16 content-retrieved candidates, so absolute values are not comparable across the two protocols. Table 5 reports each number as its distance above the chance level of its own protocol (a uniformly random ranking of M candidates has $\mathrm{HR@3} = 3/M$ and $\mathrm{NDCG@3} = (1 + 1/\log_2 3 + 1/2)/M$), which is only a coarse correction for the candidate count. On the random-candidate task of Appendix C, HaNoRec lands within a point of its published distance above chance on MicroLens (35.6 against 35.4 NDCG@3 points). This comparison is the only external anchor of this paper. Under retrieval reranking the same objective sits 23.5 points above chance, because content-retrieved candidates are near neighbours of the history and therefore harder than random negatives. This comparison anchors the scale of the numbers, not the ranking of the objectives, which is established only under our shared protocol.

Table 5: Chance-normalized comparison with HaNoRec's published MicroLens numbers. Each row gives the candidate count M of its protocol, the chance level of a uniformly random ranking, the reported value, and their difference in points. The published row is copied from Wang et al. (2026b); the other rows are HaNoRec under our protocols (means over three seeds).

		HR@3			NDCG@3		
Protocol	M	Chance	Value	Above chance	Chance	Value	Above chance
HaNoRec, published (random negatives)	10	0.300	0.7360	+43.6	0.213	0.5673	+35.4
HaNoRec, random candidates (Table 15)	16	0.188	0.6045	+41.7	0.133	0.4894	+35.6
HaNoRec, retrieval reranking (Table 1)	16	0.188	0.4809	+29.3	0.133	0.3686	+23.5

Per-seed results. Table 6 lists the covered NDCG@5 of every main-table objective for each training seed and dataset; Table 1 reports their means and standard deviations, and the per-seed ranking statements of Section 4.2 are read from this table. The per-seed differences of MoDiFy-Rec from HaNoRec are +0.83/+0.90/+0.38, +0.34/+0.58/-0.54 and +0.29/-0.33/+0.57 NDCG points on MicroLens, MovieLens-1M and Netflix, and those from mass-matched Global-Sym are +0.63/+0.75/+0.65, +0.17/+0.84/-0.57 and +0.06/-0.36/+1.18 points; the test-user bootstrap intervals of Section 4.1 do not include this training variability.

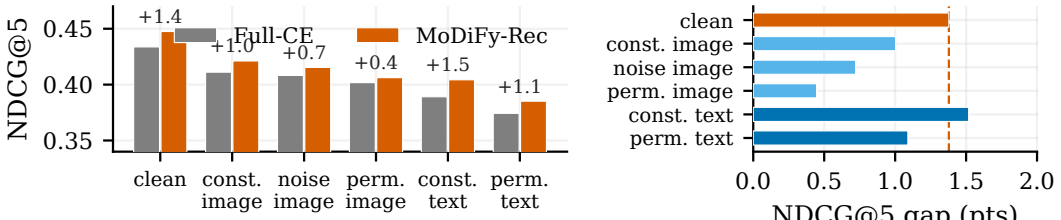
Table 6: Per-seed covered NDCG@5 of the main-table objectives. Within a seed, all objectives start from that seed's shared Full-CE initialization.

	MicroLens			MovieLens-1M			Netflix		
Objective	Seed 1	Seed 2	Seed 3	Seed 1	Seed 2	Seed 3	Seed 1	Seed 2	Seed 3
Full-CE	0.4379	0.4282	0.4351	0.4200	0.4205	0.4037	0.4610	0.4662	0.4700
Uniform-Pair	0.4423	0.4324	0.4395	0.4189	0.4254	0.4179	0.4670	0.4715	0.4720
Global-Sym (mass-matched)	0.4483	0.4347	0.4393	0.4250	0.4292	0.4190	0.4769	0.4793	0.4714
Standard-DPO	0.4363	0.4242	0.4316	0.4140	0.4196	0.4012	0.4625	0.4676	0.4669
HaNoRec	0.4462	0.4332	0.4420	0.4233	0.4318	0.4188	0.4746	0.4790	0.4774
MoDiFy-Rec (ours)	0.4546	0.4422	0.4458	0.4267	0.4376	0.4134	0.4776	0.4757	0.4831

Slices and interventions. The slice statistic is the median of d_{ij} over the $\binom{M}{2}$ pairs of a candidate set; tercile boundaries are computed on validation and frozen. Interventions replace candidate images by a mid-gray image, by pixel-wise Gaussian noise with the dataset's mean and standard deviation, or by a permutation of the images across the candidates of the same set (and analogously for texts, with a constant string "item"). History content is left intact. Because the source replication study replaces embeddings at both training and inference (Ye et al., 2026), our test-time-only interventions are a different and weaker diagnostic.

Table 7: Slice analysis on MovieLens-1M and Netflix (relative NDCG@5 change versus Full-CE in %, with 95% user-level paired bootstrap intervals; high/mid/low sizes 625/629/629 and 1,110/1,113/1,109; the last column checks the three Q1 conditions as in Table 2).

Objective	Overall	High	Mid	Low	High − Low	Q1
MovieLens-1M						
Uniform-Pair	+1.4 [+0.4, +2.5]	+1.9 [-0.1, +4.1]	+1.2 [-0.5, +3.0]	+1.2 [-0.5, +3.0]	+0.7 [-2.0, +3.5]	✗✓✗
Global-Sym (λ=1)	+1.6 [+0.5, +2.7]	+3.3 [+1.2, +5.4]	+1.0 [-0.8, +2.8]	+0.6 [-1.2, +2.4]	+2.7 [-0.1, +5.5]	✗✗✗
Global-Sym (mass-matched)	+2.3 [+1.3, +3.4]	+4.2 [+2.2, +6.3]	+1.4 [-0.3, +3.0]	+1.5 [-0.1, +3.2]	+2.7 [+0.0, +5.4]	✗✓✓
Standard-DPO	-0.8 [-1.9, +0.3]	-0.2 [-2.2, +1.9]	-1.2 [-2.9, +0.6]	-0.9 [-2.6, +0.8]	+0.7 [-2.0, +3.3]	✗✗✗
HaNoRec	+2.4 [+1.3, +3.4]	+3.0 [+1.1, +5.1]	+2.5 [+0.7, +4.5]	+1.7 [+0.1, +3.3]	+1.3 [-1.3, +4.0]	✗✓✗
MoDiFy-Rec (ours)	+2.7 [+1.6, +3.8]	+6.9 [+4.8, +9.2]	+1.3 [-0.5, +3.2]	+0.0 [-1.7, +1.7]	+6.9 [+4.3, +9.8]	✓✗✓
Netflix						
Uniform-Pair	+1.0 [+0.2, +1.8]	+1.1 [-0.2, +2.5]	+0.8 [-0.6, +2.3]	+1.0 [-0.4, +2.3]	+0.1 [-1.8, +2.1]	✗✓✗
Global-Sym (λ=1)	+1.1 [+0.3, +1.9]	+2.6 [+1.2, +4.0]	+1.0 [-0.3, +2.4]	-0.3 [-1.6, +1.0]	+2.9 [+1.0, +4.8]	✗✗✓
Global-Sym (mass-matched)	+2.2 [+1.4, +3.0]	+4.2 [+2.8, +5.6]	+1.5 [+0.1, +2.9]	+0.9 [-0.3, +2.1]	+3.3 [+1.5, +5.2]	✗✓✓
Standard-DPO	+0.0 [-0.8, +0.8]	-0.3 [-1.7, +1.1]	-0.2 [-1.6, +1.1]	+0.5 [-0.7, +1.8]	-0.8 [-2.7, +1.0]	✗✓✗
HaNoRec	+2.4 [+1.6, +3.3]	+4.0 [+2.6, +5.5]	+1.9 [+0.4, +3.5]	+1.5 [+0.2, +2.8]	+2.5 [+0.6, +4.5]	✗✓✓
MoDiFy-Rec (ours)	+2.8 [+2.0, +3.6]	+5.3 [+4.0, +6.8]	+1.7 [+0.3, +3.2]	+1.4 [+0.1, +2.8]	+3.9 [+1.9, +5.9]	✓✓✓



(a) Absolute quality under intervention

(b) Gap retained

Figure 2: Test-time modality interventions on MicroLens. Candidates, labels and slices are fixed; only candidate images (constant, Gaussian noise, or permuted across the set) or texts (constant, or permuted) are replaced at inference. (a) Covered NDCG@5 of the Full-CE and MoDiFy-Rec checkpoints with the MoDiFy-Rec−Full-CE gap in NDCG points; (b) the same gap, dashed line at the clean-input gap. A diagnostic, not a causal identification of either modality.

Table 8: Test-time modality interventions on MicroLens. Covered NDCG@5 of the Full-CE and MoDiFy-Rec checkpoints, relative change from each model's clean score (%), the MoDiFy-Rec−Full-CE gap in NDCG points, and the same three quantities on the high-disagreement slice. Gaps and relative changes are computed from unrounded scores.

	All covered events					High slice		
Intervention	Full-CE	Δ%	MoDiFy-Rec	Δ%	Gap	Full-CE	MoDiFy-Rec	Gap
Clean input	0.4337	—	0.4475	—	+1.38	0.4245	0.4530	+2.85
Constant image	0.4111	-5.2	0.4212	-5.9	+1.00	0.4000	0.4144	+1.45
Gaussian-noise image	0.4082	-5.9	0.4154	-7.2	+0.72	0.3976	0.4062	+0.87
Cross-candidate permuted image	0.4017	-7.4	0.4062	-9.2	+0.45	0.3877	0.3925	+0.49
Constant text	0.3891	-10.3	0.4043	-9.7	+1.52	0.3791	0.4015	+2.25
Cross-candidate permuted text	0.3743	-13.7	0.3852	-13.9	+1.09	0.3648	0.3812	+1.63

Table 9: Primary and secondary family tests on MicroLens: MoDiFy-Rec versus each main-table baseline on high-slice NDCG@5, relative gain in %, 95% paired bootstrap interval, raw and Holm-adjusted two-sided bootstrap p-values. The secondary family (below the rule) compares MoDiFy-Rec with the three closest methods and is Holm-corrected separately; the Row-Corrected comparison asks whether the correction of Proposition 2 changes the result; a non-significant difference there is not evidence of equivalence.

Comparison	Gain (%)	95% CI	p	p_{Holm}
MoDiFy-Rec vs. Full-CE	+6.71	[+5.42, +8.01]	$< 10^{-3}$	$< 10^{-3}$
MoDiFy-Rec vs. Uniform-Pair	+5.71	[+4.55, +6.92]	$< 10^{-3}$	$< 10^{-3}$
MoDiFy-Rec vs. Global-Sym (λ=1)	+4.10	[+2.99, +5.28]	$< 10^{-3}$	$< 10^{-3}$
MoDiFy-Rec vs. Global-Sym (mass-matched)	+2.91	[+1.85, +4.01]	$< 10^{-3}$	$< 10^{-3}$
MoDiFy-Rec vs. Row-Corrected	+0.83	[-0.20, +1.90]	0.125	0.125
MoDiFy-Rec vs. HaNoRec	+3.25	[+2.18, +4.33]	$< 10^{-3}$	$< 10^{-3}$

Table 10: Sensitivity to λ and τ on MicroLens. Validation and test covered NDCG@5 and relative change versus Full-CE (%, 95% intervals). The temperature rows use $\lambda = 1$; the strength rows use $\tau = 0.1$. τ was selected on validation and $\lambda = 1$ was fixed in advance; the validation maximum is in bold.

Setting	Val NDCG@5	Test NDCG@5	Overall	High	Low	High − Low
MoDiFy-Rec, λ=0.5	0.4481	0.4415	+1.8 [+1.1, +2.4]	+4.1 [+2.9, +5.3]	+0.1 [-1.0, +1.1]	+4.0 [+2.5, +5.6]
MoDiFy-Rec, λ=1, τ=0.1 (main)	**0.4510**	0.4475	+3.2 [+2.5, +3.8]	+6.7 [+5.4, +8.1]	+0.8 [-0.3, +1.9]	+5.9 [+4.2, +7.6]
MoDiFy-Rec, λ=2	0.4466	0.4407	+1.6 [+1.0, +2.3]	+5.3 [+4.1, +6.6]	-0.9 [-2.0, +0.2]	+6.2 [+4.6, +7.9]
MoDiFy-Rec, τ=0.05	0.4468	0.4416	+1.8 [+1.2, +2.5]	+4.9 [+3.6, +6.1]	-0.3 [-1.4, +0.8]	+5.2 [+3.5, +6.8]
MoDiFy-Rec, τ=0.2	0.4475	0.4464	+2.9 [+2.3, +3.6]	+5.9 [+4.7, +7.1]	+0.8 [-0.3, +1.9]	+5.1 [+3.4, +6.7]
MoDiFy-Rec, τ=0.5	0.4427	0.4399	+1.4 [+0.8, +2.0]	+3.0 [+1.9, +4.2]	+0.7 [-0.4, +1.9]	+2.3 [+0.7, +3.9]

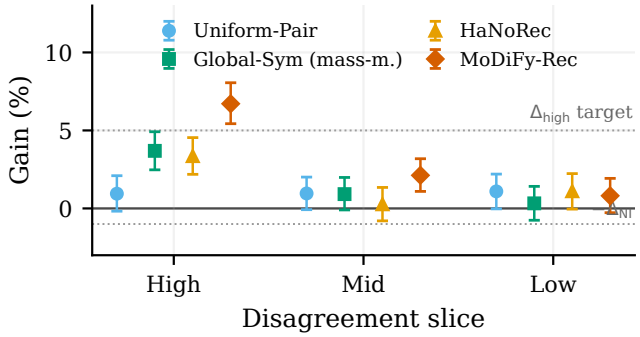
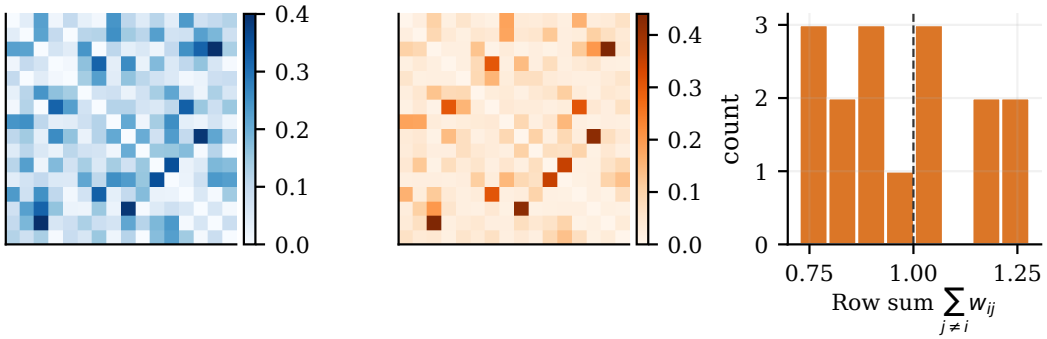
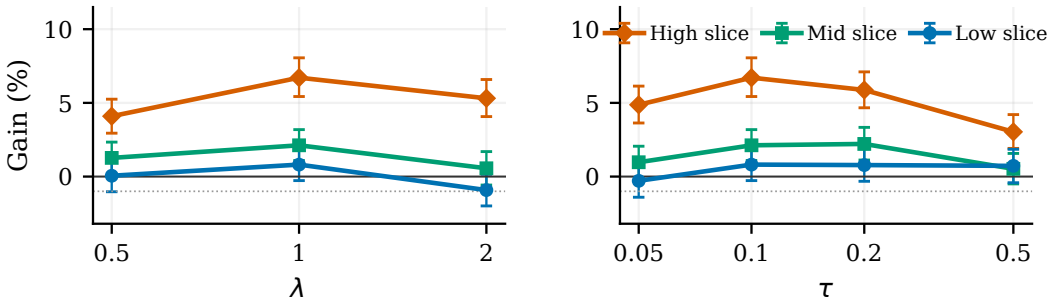


Figure 3: Gain by disagreement slice on MicroLens. Relative NDCG@5 gain over Full-CE by slice with 95% paired bootstrap intervals; dotted lines mark the pre-specified high-slice target and the non-inferiority margin.



(a) Disagreement d_{ij} (b) Symmetric weight w_{ij} (c) Row sums

Figure 4: Weight structure for one MicroLens candidate set ($M = 16$, $\tau = 0.1$). (a) Pair disagreement d_{ij} (Eq. 1); (b) the symmetric weight w_{ij} (Eq. 2); (c) row sums $\sum_{j\neq i} w_{ij}$, which cluster around one but are not exactly normalized.



(a) Auxiliary strength λ (b) Kernel temperature τ

Figure 5: Sensitivity on MicroLens. Relative NDCG@5 gain over Full-CE per slice with 95% paired bootstrap intervals as (a) the auxiliary strength λ and (b) the kernel temperature τ vary; the dotted line is the non-inferiority margin. Main-table values are $\lambda = 1$, $\tau = 0.1$.

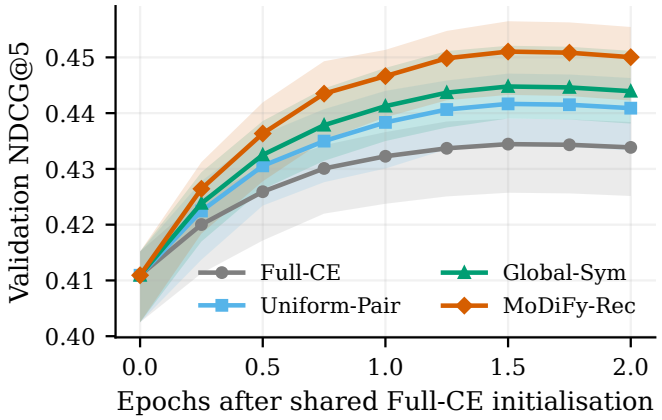


Figure 6: Validation curves on MicroLens (covered NDCG@5, evaluated every quarter epoch after the shared Full-CE initialization). Lines are means over the three seeds and bands the seed range; within a seed all objectives start from that seed's shared checkpoint. Checkpoints are selected at the best validation value with ties resolved toward the earlier one.

B RETRIEVAL-CHANNEL, WEIGHT AND DIFFICULTY DIAGNOSTICS

This appendix reports the diagnostics referred to in Sections 4.2 and 4.3. They are computed from the frozen training candidate sets and the weights of Eq. (2) at $\tau = 0.1$ (Tables 11 and 13) and from the per-user test outputs of Full-CE and MoDiFy-Rec (Tables 14 and 12); no additional training was involved. The difficulty proxy of Table 14 is the score margin of the target over the best competitor at the shared Full-CE initialization, computed before any objective was trained, and its terciles are crossed with the disagreement terciles of Section 4.1. Table 12 groups the covered test events by the channel whose top-$M/2$ list contained the target (visual only, textual only, or both), which is recorded when the candidate sets are frozen.

Table 11: Retrieval-channel composition of disagreement on MicroLens. Pairs of a candidate set are classified by the channels that retrieved their two items (both visual, both textual, or one of each). Cross-channel pairs are over-represented in the high-disagreement tail and in the auxiliary mass, but do not exhaust either.

	Within visual	Within textual	Cross-channel
Share of pairs (base rate)	23.6%	23.5%	53.0%
Mean d_{ij}	0.199	0.187	0.224
Median d_{ij}	0.198	0.186	0.224
Share among the top decile of d_{ij}	18.1%	14.0%	67.8%
Share of auxiliary mass $\sum w_{ij}$	20.5%	19.1%	60.3%

Table 12: Gain by retrieval source of the target on MicroLens. Covered test events grouped by the channel(s) that retrieved the target: number of events, share of high-disagreement sets, Full-CE covered NDCG@5, and the relative MoDiFy-Rec−Full-CE change in NDCG@5 (%, 95% user-level paired bootstrap intervals) on all events of the group and on its high-disagreement events.

Target retrieved by	n	High-slice share	Full-CE	Δ all (%)	Δ high slice (%)
Visual channel only	2,277	38%	0.4419	+3.5 [+2.5, +4.6]	+6.8 [+4.9, +8.8]
Textual channel only	2,305	35%	0.4289	+3.3 [+2.3, +4.5]	+7.1 [+5.1, +9.3]
Both channels	1,160	21%	0.4273	+2.2 [+0.7, +3.6]	+5.0 [+2.0, +8.2]

Table 13: Weight statistics on MicroLens training candidate sets ($M = 16$, $\tau = 0.1$), by disagreement slice and overall. The participation ratio $(\sum_j w_{ij})^2 / \sum_j w_{ij}^2$ is the effective number of comparisons received by a candidate.

	High slice	Mid slice	Low slice	All sets
Row sum $\sum_{j \neq i} w_{ij}$, mean (std)	0.979 (0.146)	0.978 (0.146)	0.978 (0.148)	0.978 (0.147)
Set-level median d_{ij}, mean	0.224	0.209	0.194	0.210
Cross-channel pair share per set, mean (std)	53.0% (0.4%)	53.0% (0.4%)	53.0% (0.4%)	53.0% (0.4%)
Participation ratio per row (of $M - 1 = 15$), median [IQR]	9.3 [8.1, 10.5]	9.3 [8.0, 10.4]	9.2 [7.9, 10.3]	9.3 [8.0, 10.4]
$\max w_{ij} / \min w_{ij}$ per set, median	56	53	50	53
Mass share of the three largest weights per row, median	46%	46%	46%	46%

C RANDOM-CANDIDATE SUPPLEMENTARY TASK

Table 15 reports the supplementary task. Random candidates are dissimilar to the history and to each other, so disagreement is less informative: MoDiFy-Rec gains 2.3% over Full-CE in NDCG@5 here against 3.2% under retrieval reranking, while MoDiFy-Rec remains first and Full-CE last on every metric, with the two middle baselines within seed variation of each other as in Table 1.

D COMPUTE

All runs used eight NVIDIA A100-80GB GPUs. Measured throughput was 0.58 samples/s per GPU at the visual-token budgets of Section 4.1. The shared initialization, the main-table baselines and MoDiFy-Rec on the three datasets, the ablation and sensitivity variants and all evaluations (including bootstrap resampling on CPU), each trained with 3 seeds, consumed 4,155 GPU-hours over 593 wall-clock hours. The disagreement weights are precomputed once per candidate set and cached; their cost is negligible relative to the MLLM forward pass.

Table 14: Gain by pre-treatment difficulty and disagreement on MicroLens. For each difficulty tercile (rows) and disagreement tercile (columns): number of covered test events, Full-CE covered NDCG@5, and the MoDiFy-Rec−Full-CE difference in NDCG@5 points with 95% user-level paired bootstrap interval. Difficulty terciles are defined by the target's score margin at the shared initialization.

	High disagreement			Mid disagreement			Low disagreement			All slices		
Difficulty	n	Full-CE	Δ (pts)	n	Full-CE	Δ (pts)	n	Full-CE	Δ (pts)	n	Full-CE	Δ (pts)
Hard (lowest tercile)	625	0.1808	+2.2 [+1.5, +3.0]	632	0.1701	+0.7 [+0.0, +1.4]	657	0.1935	+0.2 [-0.5, +1.0]	1,914	0.1816	+1.0 [+0.6, +1.5]
Medium	634	0.3746	+2.9 [+2.0, +3.8]	641	0.4106	+1.2 [+0.3, +2.1]	639	0.4197	+1.0 [+0.0, +1.9]	1,914	0.4017	+1.7 [+1.2, +2.2]
Easy (highest tercile)	659	0.7037	+3.4 [+2.4, +4.4]	639	0.7227	+0.9 [+0.2, +1.7]	616	0.7280	-0.1 [-0.9, +0.7]	1,914	0.7179	+1.4 [+0.9, +1.9]

Table 15: Random-candidate task on MicroLens (target plus 15 random unobserved items; every event is covered by construction). This task is easier than retrieval reranking (random negatives are far from the history) and is reported separately because its candidate distribution differs from deployment; with 16 rather than 10 candidates, absolute values sit below those reported for 10-candidate random-negative protocols (Table 5 compares them after removing the chance level). Means over three seeds; best per column in bold.

Objective	NDCG@5	Recall@5	NDCG@3	Recall@3
Full-CE	0.5405	0.7409	0.4802	0.5946
Uniform-Pair	0.5434	0.7424	0.4842	0.5983
Global-Sym (mass-matched)	0.5451	0.7445	0.4860	0.6010
HaNoRec	0.5481	0.7472	0.4894	0.6045
MoDiFy-Rec (ours)	**0.5530**	**0.7517**	**0.4941**	**0.6086**

E PROOFS

Throughout this appendix the context x is fixed and omitted from the notation. We write $p = (p_1, \dots, p_M)$ for the conditional target distribution of Section 3.5, $\mathbf{1} \in \mathbb{R}^M$ for the all-ones vector, e_i for the i-th standard basis vector, $\sigma(G) = (1 + \exp(-G))^{-1}$ for the logistic sigmoid, and

$$\pi_i(s) = \frac{\exp(s_i)}{\sum_{k=1}^{M} \exp(s_k)}, \qquad s \in \mathbb{R}^M, \tag{8}$$

for the softmax of a score vector. Three elementary identities are used repeatedly:

$$\frac{d}{dG}\,\mathrm{softplus}(G) = \sigma(G), \qquad \sigma(-G) = 1 - \sigma(G), \qquad \frac{d}{dG}\sigma(G) = \sigma(G)\,\sigma(-G). \tag{9}$$

Assumption 1 (Standing assumptions)**.** *(i) $p_i > 0$ for every candidate i and $\sum_{i=1}^{M} p_i = 1$ (if some $p_j = 0$, both risks send $s_j \to -\infty$ and the statements below hold on the support of p); (ii) the score vector $s \in \mathbb{R}^M$ is unrestricted; (iii) the weights $w_{ij} = w_{ji} > 0$ are fixed and do not depend on s.*

Taking the expectation of Eqs. (3)–(4) over the target index $i \sim p$ with x fixed gives the conditional population risks of full-candidate cross-entropy, of the auxiliary term, and of the joint objective:

$$R_{\mathrm{CE}}(s) = -\sum_{i=1}^{M} p_i \log \pi_i(s), \tag{10}$$

$$R_{\mathrm{emph}}(s) = \sum_{i=1}^{M} p_i \sum_{j \neq i} w_{ij}\ \mathrm{softplus}(s_j - s_i), \tag{11}$$

$$R_\lambda(s) = R_{\mathrm{CE}}(s) + \lambda\, R_{\mathrm{emph}}(s), \qquad \lambda \geq 0. \tag{12}$$

All three functions are finite and twice continuously differentiable on $\mathbb{R}^M$, and all three are invariant to adding a common constant to every score, because both the softmax and every score difference $s_j - s_i$ are. We write

$$S^* = \left\{ \log p + c\,\mathbf{1} :\ c \in \mathbb{R} \right\} \tag{13}$$

for the set of score vectors whose softmax equals p.

E.1 AUXILIARY LEMMAS

Lemma 1 (Pairwise decomposition). *Let $a_{ij} \geq 0$ be arbitrary coefficients indexed by ordered pairs $i \neq j$, and let $G_{ij} = s_i - s_j$. Then*

$$\sum_{i=1}^{M} p_i \sum_{j \neq i} a_{ij}\ \mathrm{softplus}(s_j - s_i) = \sum_{i<j} \Big[p_i\, a_{ij}\ \mathrm{softplus}(-G_{ij}) + p_j\, a_{ji}\ \mathrm{softplus}(G_{ij}) \Big]. \tag{14}$$

In particular, under Assumption 1(iii) the auxiliary risk of Eq. (11) equals

$$R_{\mathrm{emph}}(s) = \sum_{i<j} w_{ij}\, \phi_{ij}(G_{ij}), \qquad \phi_{ij}(G) = p_i\ \mathrm{softplus}(-G) + p_j\ \mathrm{softplus}(G). \tag{15}$$

Proof. The double sum on the left of Eq. (14) runs over all ordered pairs (i, j) with $i \neq j$. Every unordered pair $\{i, j\}$ with $i < j$ contributes exactly two ordered pairs, namely (i, j), in which i is the target and j the competitor, and (j, i), in which the roles are exchanged. Collecting these two terms and substituting $s_j - s_i = -G_{ij}$ and $s_i - s_j = G_{ij}$ gives

$$p_i\, a_{ij}\ \mathrm{softplus}(s_j - s_i) + p_j\, a_{ji}\ \mathrm{softplus}(s_i - s_j) = p_i\, a_{ij}\ \mathrm{softplus}(-G_{ij}) + p_j\, a_{ji}\ \mathrm{softplus}(G_{ij}), \tag{16}$$

and summing over $i < j$ yields Eq. (14). When $a_{ij} = a_{ji} = w_{ij}$, the common factor w_{ij} can be pulled out of the bracket,

$$\begin{aligned} p_i\, w_{ij}\ \mathrm{softplus}(-G_{ij}) + p_j\, w_{ji}\ \mathrm{softplus}(G_{ij}) &= w_{ij} \Big[p_i\ \mathrm{softplus}(-G_{ij}) + p_j\ \mathrm{softplus}(G_{ij}) \Big] \\ &= w_{ij}\, \phi_{ij}(G_{ij}), \end{aligned} \tag{17}$$

which is Eq. (15). The symmetry $w_{ij} = w_{ji}$ is used exactly in this last factorization. □

Lemma 2 (Minimizer and curvature of a pairwise term). *For constants $a, b > 0$ let $\phi(G) = a\ \mathrm{softplus}(-G) + b\ \mathrm{softplus}(G)$ on $G \in \mathbb{R}$. Then ϕ is strictly convex, its unique minimizer is*

$$G^* = \log a - \log b, \tag{18}$$

and its curvature at the minimizer is

$$\phi''(G^*) = \frac{ab}{a+b}. \tag{19}$$

For $a = p_i$ and $b = p_j$ this gives, for the function ϕ_{ij} of Lemma 1, the minimizer $G^*_{ij} = \log p_i - \log p_j$ and the curvature $\phi''_{ij}(G^*_{ij}) = p_i p_j/(p_i + p_j)$.

Proof. By the first two identities in Eq. (9), $\frac{d}{dG}\,\mathrm{softplus}(-G) = -\sigma(-G) = \sigma(G) - 1$, so

$$\phi'(G) = a\big(\sigma(G) - 1\big) + b\,\sigma(G) = (a+b)\,\sigma(G) - a. \tag{20}$$

Differentiating once more with the third identity in Eq. (9) gives

$$\phi''(G) = (a+b)\,\sigma(G)\,\sigma(-G), \tag{21}$$

which is strictly positive for every G because $a + b > 0$ and σ takes values in $(0, 1)$. Hence ϕ is strictly convex on $\mathbb{R}$. Moreover $\phi'(G) \to -a < 0$ as $G \to -\infty$ and $\phi'(G) \to b > 0$ as $G \to +\infty$, so ϕ' has a zero, and by strict convexity this zero is the unique global minimizer. Setting Eq. (20) to zero and solving,

$$\sigma(G^*) = \frac{a}{a+b} \quad \Longleftrightarrow \quad \exp(G^*) = \frac{\sigma(G^*)}{1 - \sigma(G^*)} = \frac{a}{b} \quad \Longleftrightarrow \quad G^* = \log a - \log b, \tag{22}$$

which is Eq. (18). At this point $\sigma(G^*) = a/(a+b)$ and $\sigma(-G^*) = 1 - \sigma(G^*) = b/(a+b)$, so Eq. (21) evaluates to

$$\phi''(G^*) = (a+b) \cdot \frac{a}{a+b} \cdot \frac{b}{a+b} = \frac{ab}{a+b}, \tag{23}$$

which is Eq. (19). □

Lemma 3 (Minimizers of the cross-entropy risk)**.** *Under Assumption 1(i)–(ii), the cross-entropy risk of Eq. (10) decomposes as*

$$R_{\mathrm{CE}}(s) = H(p) + \mathrm{KL}\big(p \,\|\, \pi(s)\big), \qquad H(p) = -\sum_{i=1}^{M} p_i \log p_i, \tag{24}$$

and its set of minimizers is exactly S^*.

Proof. Adding and subtracting $\sum_i p_i \log p_i$ in Eq. (10) gives

$$R_{\mathrm{CE}}(s) = -\sum_{i=1}^{M} p_i \log p_i + \sum_{i=1}^{M} p_i \log \frac{p_i}{\pi_i(s)} = H(p) + \mathrm{KL}\big(p \,\|\, \pi(s)\big), \tag{25}$$

where the Kullback–Leibler divergence is well defined because $p_i > 0$ and $\pi_i(s) > 0$ for all i. By Gibbs' inequality, $\mathrm{KL}(p \,\|\, \pi(s)) \geq 0$ with equality if and only if $\pi(s) = p$. Hence $R_{\mathrm{CE}}(s) \geq H(p)$ for all s, with equality exactly when $\pi(s) = p$. Finally, since $\pi_i(s)/\pi_j(s) = \exp(s_i - s_j)$, the condition $\pi(s) = p$ is equivalent to

$$s_i - s_j = \log p_i - \log p_j \quad \text{for all } i, j \quad \Longleftrightarrow \quad s - \log p \in \mathrm{span}\{\mathbf{1}\} \quad \Longleftrightarrow \quad s \in S^*, \tag{26}$$

where $\log p$ is finite because $p_i > 0$. The value $H(p)$ is attained on S^*, so S^* is the set of minimizers. □

E.2 Proof of Proposition 1

We first show that the auxiliary risk is minimized exactly on S^*. By Lemma 1, $R_{\mathrm{emph}}(s) = \sum_{i<j} w_{ij}\, \phi_{ij}(s_i - s_j)$, and by Lemma 2 each ϕ_{ij} is strictly convex with unique minimizer $G^*_{ij} = \log p_i - \log p_j$. Since every coefficient w_{ij} is positive, the risk is bounded below term by term,

$$R_{\mathrm{emph}}(s) = \sum_{i<j} w_{ij}\, \phi_{ij}(s_i - s_j) \;\geq\; \sum_{i<j} w_{ij}\, \phi_{ij}(G^*_{ij}) \qquad \text{for all } s \in \mathbb{R}^M, \tag{27}$$

with equality if and only if $\phi_{ij}(s_i - s_j) = \phi_{ij}(G^*_{ij})$ for every pair, that is, if and only if $s_i - s_j = G^*_{ij}$ for every pair by the uniqueness of the pairwise minimizer. The lower bound is attained: any $s \in S^*$ satisfies

$$s_i - s_j = (\log p_i + c) - (\log p_j + c) = \log p_i - \log p_j = G^*_{ij} \qquad \text{for all } i < j, \tag{28}$$

so all pairwise minima are reached simultaneously by a single score vector. Conversely, if s attains the bound then $s_i - s_j = \log p_i - \log p_j$ for all pairs, which by the equivalence displayed in the proof of Lemma 3 means $s \in S^*$. Therefore

$$\arg\min_s R_{\mathrm{emph}}(s) = S^*, \qquad \min_s R_{\mathrm{emph}}(s) = \sum_{i<j} w_{ij}\, \phi_{ij}\big(\log p_i - \log p_j\big). \tag{29}$$

By Lemma 3, the cross-entropy risk has the same set of minimizers, $\arg\min_s R_{\mathrm{CE}}(s) = S^*$.

It remains to treat the joint risk. For $\lambda = 0$ we have $R_0 = R_{\mathrm{CE}}$ and the claim is Lemma 3. For $\lambda > 0$, adding the two lower bounds gives, for every $s \in \mathbb{R}^M$,

$$\begin{aligned} R_\lambda(s) = R_{\mathrm{CE}}(s) + \lambda R_{\mathrm{emph}}(s) \;&\geq\; \min_{s'} R_{\mathrm{CE}}(s') + \lambda \min_{s'} R_{\mathrm{emph}}(s') \\ &=\; R_\lambda(s^*) \qquad \text{for any } s^* \in S^*, \end{aligned} \tag{30}$$

where the last equality holds because s^* minimizes both risks at once. Equality in Eq. (30) requires equality in both underlying bounds, i.e. $s \in S^*$ for the cross-entropy term and $s \in S^*$ for the auxiliary term. Hence $\arg\min_s R_\lambda(s) = S^*$ for every $\lambda \geq 0$.

Finally, all elements of S^* share the same score differences $s_i - s_j = \log p_i - \log p_j$ and the same candidate probabilities $\pi(s) = p$; they differ only by the common constant c, which changes neither the ranking nor the softmax. This proves the proposition. □

E.3 Derivation of Eq. (5)

We compute the Hessian of R_λ with respect to the score vector at a point $s^* \in S^*$ and evaluate its quadratic form along an arbitrary direction $u \in \mathbb{R}^M$.

For the cross-entropy term, the derivative of the log-softmax is $\partial \log \pi_i(s)/\partial s_k = \mathbf{1}\{i=k\} - \pi_k(s)$, so that, using $\sum_i p_i = 1$,

$$\frac{\partial R_{\mathrm{CE}}}{\partial s_k}(s) = -\sum_{i=1}^{M} p_i\big(\mathbf{1}\{i=k\} - \pi_k(s)\big) = \pi_k(s) - p_k. \tag{31}$$

Differentiating once more with $\partial \pi_k/\partial s_l = \pi_k(\mathbf{1}\{k=l\} - \pi_l)$ gives

$$\nabla_s^2 R_{\mathrm{CE}}(s) = \operatorname{diag}\big(\pi(s)\big) - \pi(s)\,\pi(s)^\top. \tag{32}$$

At $s^* \in S^*$ we have $\pi(s^*) = p$, and the quadratic form becomes

$$\begin{aligned} u^\top \nabla_s^2 R_{\mathrm{CE}}(s^*)\, u &= \sum_{i=1}^{M} p_i u_i^2 - \Big(\sum_{i=1}^{M} p_i u_i\Big)^2 \\ &= \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M} p_i p_j \big(u_i^2 - 2u_i u_j + u_j^2\big) \\ &= \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M} p_i p_j (u_i - u_j)^2 = \sum_{i<j} p_i p_j (u_i - u_j)^2, \end{aligned} \tag{33}$$

where the second equality expands the double sum with $\sum_j p_j = 1$, and the last equality drops the vanishing diagonal terms and uses the symmetry of the summand in (i, j).

For the auxiliary term, Lemma 1 writes R_{emph} as a sum of terms $w_{ij}\,\phi_{ij}(G_{ij})$ that depend on s only through the linear form $G_{ij} = (e_i - e_j)^\top s$. By the chain rule, the Hessian of such a term is

$$\begin{aligned} \nabla_s^2\big[w_{ij}\,\phi_{ij}(G_{ij})\big] &= w_{ij}\,\phi_{ij}''(G_{ij})\,(e_i - e_j)(e_i - e_j)^\top, \\ u^\top \nabla_s^2\big[w_{ij}\,\phi_{ij}(G_{ij})\big] u &= w_{ij}\,\phi_{ij}''(G_{ij})\,(u_i - u_j)^2. \end{aligned} \tag{34}$$

At $s^* \in S^*$ every G_{ij} equals G_{ij}^*, so Lemma 2 gives $\phi_{ij}''(G_{ij}^*) = p_i p_j/(p_i + p_j)$ and

$$u^\top \nabla_s^2 R_{\mathrm{emph}}(s^*)\, u = \sum_{i<j} w_{ij}\,\frac{p_i p_j}{p_i + p_j}\,(u_i - u_j)^2. \tag{35}$$

Adding Eq. (33) and λ times Eq. (35) yields

$$u^\top \nabla_s^2 R_\lambda(s^*)\, u = \sum_{i<j}\Big[p_i p_j + \lambda\, w_{ij}\frac{p_i p_j}{p_i + p_j}\Big](u_i - u_j)^2, \tag{36}$$

which is Eq. (5). The expression is evaluated at the shared optimum and therefore describes the risk only locally; it is a Hessian with respect to the candidate scores, not with respect to the LoRA parameters, and it says nothing about the optimization trajectory. □

E.4 Row-normalized kernels

The Row-Corrected variant of Section 4.3 replaces the symmetric weights by the row-normalized kernel $q_{ij} = W_{ij}/Z_i$, whose rows sum to one, and subtracts an offset from the score gap. The following proposition explains why an offset is needed there and not in MoDiFy-Rec, and why the closed-form offset of an earlier design is not exact once the kernel is mixed with a uniform component. Throughout, an offset δ_{ij} enters the pairwise term as $\mathrm{softplus}(s_j - s_i - \delta_{ij})$, and offsets are antisymmetric, $\delta_{ji} = -\delta_{ij}$.

Proposition 2 (Row-normalized kernel)**.** *Let Assumption 1(i)–(ii) hold, let W be symmetric with positive off-diagonal entries, $Z_i = \sum_{k\neq i} W_{ik}$ and $q_{ij} = W_{ij}/Z_i$.*

- *(a) The uncorrected risk $R_q(s) = \sum_i p_i \sum_{j\neq i} q_{ij}\ \mathrm{softplus}(s_j - s_i)$ is minimized exactly on $\{s:\ s_i = \log p_i - \log Z_i + c,\ c \in \mathbb{R}\}$.*
- *(b) With the exact offset $\delta_{ij} = \log(q_{ij}/q_{ji}) = \log Z_j - \log Z_i$, the corrected risk $\tilde{R}_q(s) = \sum_i p_i \sum_{j\neq i} q_{ij}\ \mathrm{softplus}(s_j - s_i - \delta_{ij})$ is minimized exactly on S^*.*

(c) *For the mixed kernel* $\tilde{q}_{ij} = (1-\epsilon)\, q_{ij} + \epsilon/(M-1)$ *with* $\epsilon \in (0,1)$*, the exact offset is* $\log(\tilde{q}_{ij}/\tilde{q}_{ji})$*, and it differs from* $\log Z_j - \log Z_i$ *whenever* $Z_i \neq Z_j$*; the closed-form offset* $\log Z_j - \log Z_i$ *is therefore not an exact correction for the mixed kernel.*

Proof. Lemma 1 with $a_{ij} = q_{ij}$ writes the uncorrected risk as

$$R_q(s) = \sum_{i<j} \psi_{ij}(G_{ij}), \qquad G_{ij} = s_i - s_j,$$
$$\psi_{ij}(G) = p_i\, q_{ij}\ \mathrm{softplus}(-G) + p_j\, q_{ji}\ \mathrm{softplus}(G). \tag{37}$$

Each ψ_{ij} has the form of Lemma 2 with $a = p_i q_{ij} > 0$ and $b = p_j q_{ji} > 0$, so it is strictly convex with the unique minimizer

$$G^q_{ij} = \log \frac{p_i\, q_{ij}}{p_j\, q_{ji}} = \log p_i - \log p_j + \log \frac{q_{ij}}{q_{ji}}. \tag{38}$$

Because W is symmetric, $q_{ij}/q_{ji} = (W_{ij}/Z_i)/(W_{ji}/Z_j) = Z_j/Z_i$, and Eq. (38) becomes

$$G^q_{ij} = \big(\log p_i - \log Z_i\big) - \big(\log p_j - \log Z_j\big). \tag{39}$$

The vector $s_i = \log p_i - \log Z_i + c$ attains every G^q_{ij} simultaneously, and the term-by-term lower-bound argument of Appendix E.2 (with the positive coefficients now absorbed into ψ_{ij}) shows that no other score differences do. This proves part (a): the row normalization shifts every optimal score by $-\log Z_i$, so the optimal candidate probabilities are proportional to p_i/Z_i rather than to p_i.

For part (b), the corrected risk collects, for the unordered pair $\{i, j\}$, the two terms

$$\begin{aligned} &p_i\, q_{ij}\ \mathrm{softplus}\left(s_j - s_i - \delta_{ij}\right) + p_j\, q_{ji}\ \mathrm{softplus}\left(s_i - s_j - \delta_{ji}\right) \\ &\quad = p_i\, q_{ij}\ \mathrm{softplus}\left(-(G_{ij} + \delta_{ij})\right) + p_j\, q_{ji}\ \mathrm{softplus}\left(G_{ij} + \delta_{ij}\right), \end{aligned} \tag{40}$$

where antisymmetry $\delta_{ji} = -\delta_{ij}$ was used. This is ψ_{ij} evaluated at the shifted argument $G_{ij} + \delta_{ij}$, so by Eq. (38) the unique pairwise minimizer satisfies

$$G_{ij} + \delta_{ij} = \log p_i - \log p_j + \log \frac{q_{ij}}{q_{ji}} \quad \Longleftrightarrow \quad G_{ij} = \log p_i - \log p_j \qquad \text{when } \delta_{ij} = \log \frac{q_{ij}}{q_{ji}}. \tag{41}$$

The pairwise minimizers are again attained simultaneously, now by every $s \in S^*$, and by the same lower-bound argument only there; this proves part (b).

For part (c), abbreviate $\alpha_{ij} = (1-\epsilon)W_{ij}$ and $\beta = \epsilon/(M-1) > 0$, so that $\tilde{q}_{ij} = \alpha_{ij}/Z_i + \beta$ and, by the symmetry of W, $\tilde{q}_{ji} = \alpha_{ij}/Z_j + \beta$. Then

$$\begin{aligned} \frac{\tilde{q}_{ij}}{\tilde{q}_{ji}} &= \frac{(\alpha_{ij} + \beta Z_i)/Z_i}{(\alpha_{ij} + \beta Z_j)/Z_j} = \frac{Z_j}{Z_i} \cdot \frac{\alpha_{ij} + \beta Z_i}{\alpha_{ij} + \beta Z_j}, \\ \log \frac{\tilde{q}_{ij}}{\tilde{q}_{ji}} &= \log \frac{Z_j}{Z_i} + \log \frac{\alpha_{ij} + \beta Z_i}{\alpha_{ij} + \beta Z_j}. \end{aligned} \tag{42}$$

The second term vanishes if and only if $\beta Z_i = \beta Z_j$, i.e. $Z_i = Z_j$, because $\beta > 0$. Applying part (b) to the kernel $\tilde{q}$ shows that the offset restoring S^* is $\log(\tilde{q}_{ij}/\tilde{q}_{ji})$; by Eq. (42) the closed-form offset $\log Z_j - \log Z_i$ coincides with it only for pairs with equal row sums, so it is not an exact correction for the mixed kernel. With the closed-form offset, the pairwise stationarity conditions place the optimal score differences at $\log p_i - \log p_j + \log\big[(\alpha_{ij} + \beta Z_i)/(\alpha_{ij} + \beta Z_j)\big]$, which differ from $\log p_i - \log p_j$ whenever $Z_i \neq Z_j$. □