

# MERIT: Mitigating Exposure Bias in Generative XMC for User-Interest Propensity Modeling

**Abhinav Mahajan**  
Carnegie Mellon University  
Pittsburgh, PA, USA  
abhinavm@andrew.cmu.edu

**Arindam Sarkar**  
Amazon  
Bangalore, India  
arindsar@amazon.com

**Prakash Mandayam Comar**  
Amazon  
Bangalore, India  
prakasc@amazon.com

## Abstract

Matching users to interest categories at scale is central to personalized shopping, but the task is challenging in large e-commerce platforms, where label spaces continually evolve and user-interest signals are sparse and long-tailed. Autoregressive language models are appealing because their world knowledge and semantic priors over descriptors generalize across extreme label spaces and accommodate multiple valid label assignments. Yet under teacher-forced fine-tuning, inference-time predictions become part of the conditioning context: early errors steer later outputs toward co-occurring labels, over-generating near-correlates and missing unrelated true interests. We present **MERIT**, a framework for user-interest propensity modeling that mitigates this *exposure bias* through a self-correction objective. A permutation-invariant multi-target loss over shuffled mixtures of gold and mined hard-negative labels exposes the generator to erroneous prefixes while preserving the efficiency of teacher-forced training. This training objective concentrates supervision at classification positions, yielding propensity-aligned hidden states powering a lightweight scorer for bidirectional retrieval (interests for users and users for interests). On proprietary e-commerce dataset with 250k+ interest categories, MERIT improves global recall by at least **11.9%** and average Hit@k by **6.1%**. In production A/B tests, it achieves **+0.26%** gain in user conversion.

## 1 Introduction

A key problem in personalization is predicting a user’s propensity to engage with a candidate entity such as a product, collection, or campaign. Such candidates are highly dynamic (Hu et al., 2025) (continually launched and retired) creating cold-start regimes where candidate-level learning is brittle due to lack of historical signal (Lika et al., 2014; Zhang et al., 2025). A robust alternative is to employ *semantic interest descriptors* that stay

stable across content churn, which recent generative tagging (Tang et al., 2025; Li et al., 2023; Brinkmann et al., 2024) makes feasible at scale: a model assigns interpretable descriptors (e.g., “running”, “home organization”) to entities from their text and metadata, shifting the learning target from each ephemeral candidate to a stable space that generalizes to new candidates via their tags. This is attractive at deployment, since the binding constraint on the targeting layer is training cost. Typical discriminative one-versus-all propensity models over the descriptor vocabulary scale linearly in the number of descriptors, restricting deployment to high-traffic categories and precluding vocabulary evolution without manual retraining.

This motivates *candidate-free* interest propensity modeling at extreme label scale, based on a user’s interaction history: predict a sparse set of descriptors from a large vocabulary and support bidirectional retrieval, ranking interests for a user (personalization) and users for an interest (segmentation). The setting has the hallmarks of extreme multi-label classification (XMC), i.e., a long-tailed label space, sparse positives, and ambiguous non-interactions (unobserved interests are not true negatives) with two consequences. Typical e-commerce users hold multiple, often unrelated interests, so the goal is to recover a *diverse* interest set rather than variants of one. We treat this as a retrieval problem since the predicted interests are not a final user-facing decision but an upstream retrieval interface, refined downstream by domain specific rules and eligibility constraints. As in any retrieval setting, downstream filtering is cheap and missed candidates are irrecoverable, recall is therefore the operative metric here.

Generative language models are a natural fit here, as they operate directly over textual descriptors, generalize across long-tailed spaces, and produce structured multi-label outputs that are auditable. LM finetuning further transfers semantic

priors over descriptors and entities to user-interest prediction. This is useful because descriptor co-occurrence is a tractable signal in semantic space: rather than memorizing 250k+ categories independently, the model can exploit structure among them. But this reliance on correlation is double-edged. With sparse per-descriptor supervision, the model must use co-occurrence to make predictions, but must also override it when the user history supports unrelated interests. Exposure bias sharpens this tension. Teacher forcing conditions on gold prefixes, while inference conditions on the model’s own outputs (Ranzato et al., 2015; Zhang et al., 2019). In fluent-text generation, this mismatch often appears as degraded sequence quality; in multi-label generation, it becomes a recall failure. Once decoding commits to a semantic region (e.g., “sports”), the model can over-amplify co-occurrence and emit near-correlates (“athletic gear”, “fitness equipment”) in a self-reinforcing cascade, suppressing unrelated true interests. These cascades are not incidental: training data over-represents certain descriptor co-occurrences, and pretrained models bring their own priors. They also collapse representations toward the dominant region, mirroring the single-vector bottleneck of dual encoders and obstructing calibrated cross-user propensity scoring.

We introduce **MERIT**, a framework for user-interest propensity modeling at extreme label scale (250k+ labels). Our key contributions are: (i) We identify exposure bias in generative XMC as a recall failure from co-occurrence cascades, and address it with a permutation-invariant multi-target loss over mixed gold/mined prefixes that trains recovery within a standard teacher forcing pass without per-step generation, improving recall by at least **11.9%**, (ii) show that training on this objective yields, by design, propensity-aligned hidden states and informative hard negatives that together enable the training of a lightweight scorer, improving average Hit@k by at least **6.1%**, (iii) validate MERIT at production scale via A/B tests, showing positive posterior mean conversion lift averaging +0.26%.

## 2 Background & Related Work

**User interest modeling and XMC retrieval:** CTR-style user-interest models (Zhou et al., 2018, 2019; Li et al., 2019; Qi et al., 2021) score a *given* candidate against history; we instead *predict* a descriptor set from an extreme vocabulary as a retrieval interface. XMC targets such label spaces via

tree-based (Prabhu et al., 2018; Jain et al., 2016), embedding-based (Dahiya et al., 2021; Mittal et al., 2021; Dahiya et al., 2023), and retrieval-style dual-encoder methods (You et al., 2019a; Chang et al., 2020; You et al., 2019b; Jiang et al., 2021; Jain et al., 2023; Gupta et al., 2024). Dual encoders are parameter-efficient and generalize to new labels, but hinge on informative hard negatives that are costly to mine at extreme scale. In MERIT, the generator’s own confident errors furnish such negatives for free.

**Generative and code-based XMC:** Generative XMC predicts labels autoregressively (Jung et al., 2023; Simig et al., 2022); related generative-retrieval work replaces text labels with discrete Semantic IDs (Rajput et al., 2023; Tay et al., 2022), which aid latency but sacrifice interpretability, so we retain text-level descriptors. LLM-based recommendation (Bao et al., 2023; Geng et al., 2022a; Sun et al., 2023; Geng et al., 2022b; Hou et al., 2024; Lin et al., 2024; Cui et al., 2022) via prompting lacks calibrated propensities at extreme scale (Zhao et al., 2021); we therefore use the LLM to supply training signal, delegating cross-user scoring to a distilled model.

**Exposure bias in autoregressive generation:** The gap between teacher-forced training and free-running inference (Ranzato et al., 2015; Zhang et al., 2019) has been addressed via scheduled sampling (Bengio et al., 2015), reinforcement learning (Ranzato et al., 2015), and imitation learning (Ross et al., 2011), largely in fluent generation, where the failure mode is degraded sequence quality. We study the same mismatch in *multi-label* generation, where early label errors can amplify co-occurrence priors into correlation cascades reducing coverage of true label set. MERIT addresses this with an efficient self-correction objective: mixed gold/error prefixes expose the model to its own mistakes and a multi-target teacher-forced loss learns recovery without per-step generation.

## 3 Methodology

We predict user propensities over an extremely large interest-category space  $\mathcal{L} = \{l_1, \dots, l_K\}$  that acts as a retrieval interface to dynamic entities such as products and campaigns. Each product carries several multi-token category tags (e.g., a cricket bat: "sports enthusiast", "cricket player", "fitness equipment"). For a user  $c$  purchasing at time  $t_{\text{purchase}}$ , we form one example: the **input** is the interaction

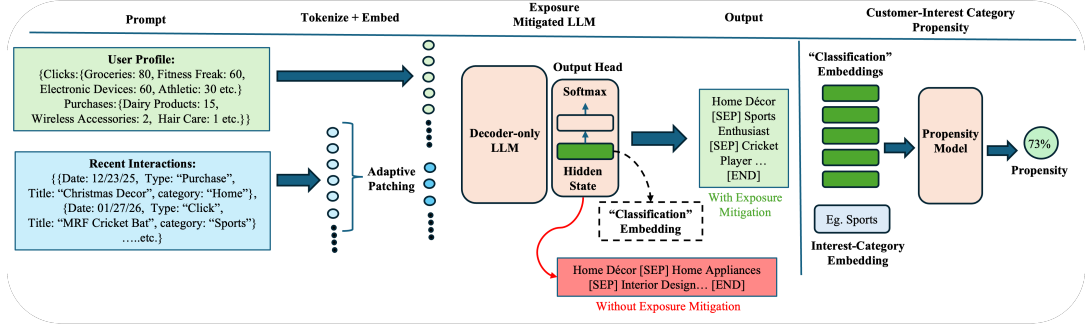


Figure 1: **MERIT inference pipeline.** User profile (aggregated interest-categories) and recent interactions (raw product data) are tokenized, embedded, and compressed via adaptive patching. The exposure-mitigated LLM autoregressively generates interest-categories. *With mitigation* (green), the model produces diverse predictions (e.g., Sports Enthusiast); *without mitigation* (red), it enters correlation loops (Home Décor → Home Appliances → Interior Design). Hidden states from the first token of the first few generated categories are extracted as enriched classification embeddings, which serve as input to the propensity model for scoring any interest category.

history  $H_c = \{(p, t)\}$  of products clicked or purchased in the preceding year, excluding the final 7 days ( $t < t_{\text{purchase}} - 7\text{d}$ ); the **output** is the tag set  $Y_c \subset \mathcal{L}$  of the purchased product. The exclusion window prevents trivial prediction from immediate pre-purchase signals and matches production settings where scores need to be pre-computed.

Finetuned LLMs are well-suited for this task, but pure generation at inference yields no calibrated scores comparable *across* users, blocking the interest→user retrieval needed for segmentation. We address this by training an exposure-mitigated generator (§3.1) with adaptive patching for long histories (§3.2), then training a lightweight propensity model on the generator’s hidden states (§3.3), reusing its own errors as hard negatives.

### 3.1 Exposure-Mitigation for Self-Correction

Standard teacher forcing conditions on gold prefixes during training but on the model’s own predictions at inference. This exposure bias makes the model exploit prefix correlations rather than classify independently: once an early token commits to a semantic region (e.g., "sports"), later outputs cluster on correlated variants ("athletic gear", "fitness equipment") instead of diverse, unrelated interests, directly harming recall. We mitigate this with a self-correction objective that teaches the model to recover from its own errors, in two steps.

**Step 1: Offline hard-negative mining.** We first train an initial generator with standard cross-entropy, then run independent temperature sampling over the training set to mine plausible but incorrect labels  $\hat{Y}_{\text{filtered}}$ , augmenting the golden tag set into the shuffled sequence,  $\mathcal{L}_c = Y_c \cup \hat{Y}_{\text{filtered}}$ .

Because this inference pass is performed offline, it is far cheaper than online sampling methods that generate at every training step. We drop any mined negative whose Cohere similarity to a gold label exceeds 0.6, intentionally below our  $\tau=0.7$  soft-match evaluation threshold so that near-synonyms and paraphrases of a user’s true interests are not injected into the prefix as negatives (see Limitations for a discussion of residual false-negative risk).

#### Step 2: Exposure-mitigated teacher forcing.

For each user we shuffle  $\mathcal{L}_c$  and concatenate its labels with a special [SEP] token into a target sequence. We retain the efficiency of teacher forcing by scoring the full sequence (prompt + shuffled target) in a *single forward pass* with no generation; the only change is to the per-token loss, which we modify to admit *multiple valid targets* while never rewarding a mined negative as a label choice, so the model is exposed to incorrect labels in the prefix without learning to reproduce them. At each position  $t$ , we define the *prefix* as the tokens consumed so far within the current label (i.e., since the last [SEP]); the valid set  $\mathcal{V}_t$  depends on this prefix:

- *Empty prefix* (a *classification* position):  $\mathcal{V}_t$  is the first tokens of every not-yet-completed golden label, or [END] if none remain. Pooling over these first tokens makes the loss **permutation-invariant** so that the model is not penalized for ordering golden labels differently.
- *Prefix matches golden label(s)*:  $\mathcal{V}_t$  is their next tokens, covering shared prefixes (e.g., "sports enthusiast" / "sports gear").
- *Prefix matches only a hard negative* (e.g., "kitchen" from "kitchen appliances"):  $\mathcal{V}_t$  com-

pletes that negative, enforcing **label coherence** so even incorrect labels stay well-formed rather than switching mid-phrase.

Whenever the prefix completes a label, [SEP] joins  $\mathcal{V}_t$ . The loss pools probability mass over  $\mathcal{V}_t$ :

$$\ell_t = -\log \left( \frac{\sum_{v \in \mathcal{V}_t} \exp(z_t^{(v)})}{\sum_{v'} \exp(z_t^{(v')})} \right). \quad (1)$$

Algo. 1 gives the full procedure. We **strongly encourage** the reader to go through the intricate example dry-run in **Appendix A**.

---

#### Algorithm 1 Self-Correction Loss Computation

---

**Require:** Golden labels  $Y_c$ , mined hard negatives  $\hat{Y}_{\text{filtered}}$

- 1:  $\mathcal{L}_c \leftarrow Y_c \cup \hat{Y}_{\text{filtered}}$  ▷ Augmented label set
- 2:  $L \leftarrow \text{Shuffle}(\mathcal{L}_c)$
- 3:  $S \leftarrow \text{Tokenize}(L \oplus [\text{SEP}])$  ▷ Target sequence
- 4:  $X \leftarrow \text{prompt} \oplus S$
- 5:  $\{z_t\}_{t=1}^T \leftarrow \text{Model}(X)$  ▷ Target Logits, single pass
- 6:  $\text{prefix} \leftarrow []$ ,  $\text{completed} \leftarrow \{\}$
- 7: **for**  $t = 1$  to  $T$  **do**
- 8:   **if**  $\text{prefix} = []$  **then** ▷ Starting a new label
- 9:      $\mathcal{V}_t \leftarrow \{\text{first token of } \ell \mid \ell \in Y_c \setminus \text{completed}\}$
- 10:    **if**  $Y_c \setminus \text{completed} = \emptyset$  **then**
- 11:      $\mathcal{V}_t \leftarrow \{[\text{END}]\}$
- 12:   **else** ▷ Continuing a label
- 13:      $G \leftarrow \{\ell \in Y_c \setminus \text{completed} : \ell \text{ has prefix } \text{prefix}\}$
- 14:     **if**  $G \neq \emptyset$  **then** ▷ Golden match
- 15:        $\mathcal{V}_t \leftarrow \text{all next tokens after prefix in labels}$
- 16:     **else** ▷ Hard negative match
- 17:        $\mathcal{V}_t \leftarrow \text{next token continuing the hard negative}$
- 18:     **if**  $\text{prefix} \in \mathcal{L}_c$  **then** ▷ Prefix forms a complete label
- 19:        $\mathcal{V}_t \leftarrow \mathcal{V}_t \cup \{[\text{SEP}]\}$
- 20:    Compute  $\ell_t = -\log \left( \frac{\sum_{v \in \mathcal{V}_t} \exp(z_t^{(v)})}{\sum_{v'} \exp(z_t^{(v')})} \right)$
- 21:    **if**  $y_t = [\text{SEP}]$  **then**
- 22:       $\text{prefix} \leftarrow []$  ▷ Reset for next label
- 23:    **else**
- 24:       $\text{prefix} \leftarrow \text{prefix} \oplus [y_t]$  ▷ Append token
- 25:      **if**  $\text{prefix} \in Y_c$  and  $\text{prefix} \notin \text{completed}$  **then**
- 26:        $\text{completed} \leftarrow \text{completed} \cup \{\text{prefix}\}$
- 27: **return**  $\frac{1}{T} \sum_{t=1}^T \ell_t$

---

**Training dynamics and efficiency.** Completion losses vanish once the model learns the vocabulary, so loss concentrates at classification positions, where the model cannot rely on prefix correlations and must classify, again if necessary, from the user’s history alone, precisely what suppresses exposure bias. As the negatives are precomputed once offline, training stays as cheap as standard teacher forcing, with no generation in the loop. This single mining pass also yields two assets we reuse for the propensity model (§3.3): (1) the per-example negatives, which serve directly as informative negatives

for its training, and (2) the hidden states at classification positions, which are *propensity-aligned* by virtue of training, they encode the user’s standing interest independent of the generated prefix.

### 3.2 Adaptive Patching for Compression

Tokenized histories routinely exceed the context budget, forcing a choice between truncating history and losing signal. We adapt the patching strategy of PatchRec (Liao et al., 2025) into a *selective*, recency-aware scheme: rather than compressing the sequence uniformly, we pool token embeddings only within older, less-predictive interaction windows (non-overlapping, window/stride = 10, 10× compression) while leaving recent interactions at full resolution. This fits a full year of history within a budget that would otherwise hold only the most recent months, preserving fine-grained signal where it matters most. We introduce patching through a four-phase curriculum (Appendix G) that anneals from full to fully-patched sequences, letting the model master the prediction task before adapting to compressed inputs. While the accuracy gain is modest (Table 10, appendix), the contribution is making extended histories deployable within fixed inference time token budget limits.

### 3.3 Propensity Model

Generator ranks categories at user level, but yields no scores comparable *across* users, so it cannot rank users for a given interest (needed for segmentation and merchandising usecases). We therefore train a lightweight propensity model that outputs cross-user-comparable scores, enabling bidirectional retrieval (interests for users, users for interests) by leveraging the propensity-aligned hidden states and hard negatives surfaced by the exposure-mitigated generator (§3.1).

**Architecture.** Self-correction concentrates supervision at classification positions (§3.1), so the first-token hidden state of each generated category is propensity-aligned and largely prefix-independent. We pass these states for the first  $k_g$  generated categories through a small trainable transformer ( $\approx 230\text{M}$  params, significantly smaller than the 3.8B generator, which is not trained at this stage) encoder to form a user representation; to score a (user, category) pair, we concatenate this representation with the category’s pretrained Cohere (Cohere, 2023) embedding and pass it through a shared MLP trained with binary cross-entropy. Concatenation rather than a bare dot product lets

the two representations interact through the network, avoiding the single-vector bottleneck we set out to escape. User and category embeddings are each precomputed once and cached, so scoring any pair is a cheap MLP forward pass.

**Smart negatives.** A generator’s confident-but-incorrect predictions are informative hard negatives for free, and we exploit this at *both* pipeline stages. Step 1 negatives from the initial vanilla generator train the exposure-mitigation objective; to supervise the propensity model, we then mine from the *trained* mitigated generator, whose greater diversity surfaces fresh plausible-but-incorrect labels and thus harder negatives than the vanilla model yields. We train the scorer on these alongside popular and random negatives.

## 4 Experiments

We evaluate MERIT on a proprietary e-commerce dataset of anonymized user interactions (500k training users, 100k disjoint test users), where each product is annotated with multiple interest categories from a vocabulary of 291k entries spanning concrete descriptors (e.g., “Exercise Equipment”) and intent-level tags (e.g., “Wellness Lifestyle”). Inputs and targets follow the formulation of §3; the year-long history with a 7-day exclusion window matches production settings where scores are computed ahead of purchase. We assess three tasks: generative classification (§4.1), propensity ranking (§4.2), and a production A/B test (§4.3).

**Setup, Baselines and Metrics.** MERIT uses phi-3.5-mini (3.8B) as the base generator, fine-tuned with our exposure-mitigated objective and adaptive patching (refer Appendix G); following (Jung et al., 2023) we manage the label space via hierarchical generation, organizing the 291k leaf categories into 30k semantic clusters and predicting clusters before leaves. The propensity model is a lightweight 230M-parameter encoder trained on hidden states from the phi generator (*frozen* at this stage), using the first-token hidden states of the first  $k_g=5$  generated categories per user. Full architecture, hard-negative mining and training details are deferred to Appendix H.

We compare against representative generative and dual-encoder XMC methods: **Trivial**, predicting user’s 5 most frequently purchased categories; **XLGen**, cluster-guided label generation (Jung et al., 2023); **GROOV**, open-vocabulary generative XMC (Simig et al., 2022); **PatchRec**, LLM-

based recommendation with patching (Liao et al., 2025); and a dual encoder (same phi model used) in two loss configurations, **DEXML-DS** (DecoupledSoftmax) and **DEXML-STk** (SoftTop- $k$ ) as proposed by (Gupta et al., 2024). Per internal disclosure policy, we report relative gain rather than absolute numbers: For generative classification, we report **Recall/Precision/F1** (soft matching via Cohere cosine similarity  $\geq 0.7$ , calibrated on manually judged boundary pairs) absolute improvement over the Trivial baseline (Table 1); for propensity ranking, we report **Hit@ $k$**  (the fraction of an interest cluster’s true-positive users falling in its top- $k$ ) as improvement over DEXML-STk (Table 2). We validate this soft-matching protocol directly with human judgments and a threshold-sensitivity sweep in Appendix B: at  $\tau=0.7$ , human annotators judged 96.2% of a 230-pair boundary sample as at least loosely semantically equivalent (65.4% strictly equivalent), and MERIT’s recall advantage over every baseline holds under all matching criteria we tested, including exact match.

### 4.1 Generative Classification and Ablations

Table 1 reports generative-classification performance and ablates the training objective; we show five common business verticals in e-commerce here, with all 20 in Appendix D (Tables 6 - 8). *Global* pools predictions over the entire test set rather than averaging the shown verticals. MERIT improves Global recall by +14.1 over Trivial and by at least +11.9 over the strongest baseline (the conservative figure reported in the abstract), with consistent F1 gains. Dual encoders rank the full label space and need a top- $k$  cutoff to yield predictions. We tuned for F1 on a control set, this gives  $k=5$  (the value at which DEXML baselines are reported), and even so they perform poorly. Raising  $k$  to 100 still trails MERIT’s recall by +1.8 despite severely low precision, pointing to embedding collapse: an overloaded space favoring one dominant category and its neighbors, the single-vector analogue of the correlation cascade, rather than output budget. Precision dips slightly in some verticals, most notably Apparel, where GROOV edges our F1/precision despite our  $3.5\times$  higher recall. This is an acceptable trade-off since recall is the operative retrieval metric: missed interests are unrecoverable downstream, while over-generation is cheaply filtered. This is also supported by §4.3.

The ablation isolates the gain’s source: *Permutation invariance* (similar to (Simig et al.,

Table 1: Generative classification performance (**Precision, Recall, F1**) and training-objective ablation. Absolute improvement over the trivial baseline across product verticals; The lower block ablates different training-objectives.

| Method                                     | Global       |             |             | Beauty       |             |             | Apparel      |             |             | Electronics  |             |             | Sports       |             |             | Toys        |             |             |
|--------------------------------------------|--------------|-------------|-------------|--------------|-------------|-------------|--------------|-------------|-------------|--------------|-------------|-------------|--------------|-------------|-------------|-------------|-------------|-------------|
|                                            | R            | F1          | P           | R            | F1          | P           | R            | F1          | P           | R            | F1          | P           | R            | F1          | P           | R           | F1          | P           |
| <i>Baselines</i>                           |              |             |             |              |             |             |              |             |             |              |             |             |              |             |             |             |             |             |
| XLGen                                      | +2.1         | +2.0        | +1.9        | +2.8         | +2.2        | +1.4        | +3.5         | +4.0        | +4.7        | +1.1         | +0.9        | +0.5        | +1.7         | +1.6        | +1.4        | +0.5        | +0.4        | +0.3        |
| GROOV                                      | +1.8         | +2.3        | +2.9        | +2.3         | +2.4        | +2.4        | +3.9         | <b>+5.3</b> | <b>+7.5</b> | +0.2         | +0.3        | +0.6        | +2.2         | +2.6        | <b>+3.3</b> | +2.0        | <b>+2.3</b> | <b>+2.7</b> |
| PatchRec                                   | +0.2         | +1.2        | +3.1        | -2.3         | -2.3        | -2.1        | -0.2         | +1.3        | +5.1        | -1.0         | -0.9        | -0.6        | -1.1         | -0.6        | +0.7        | -1.2        | -1.1        | -0.7        |
| DEXML-DS (top_k=5)                         | +1.2         | +2.2        | <b>+3.8</b> | +2.6         | +3.1        | <b>+3.7</b> | +1.8         | +2.6        | +3.9        | +1.6         | +2.6        | <b>+4.5</b> | -0.7         | -0.5        | -0.1        | -0.3        | -0.3        | -0.2        |
| DEXML-STk (top_k=5)                        | -0.9         | -4.1        | -5.2        | +4.1         | -3.1        | -5.7        | -1.8         | -4.2        | -5.1        | -3.2         | -4.4        | -5.4        | -0.9         | -4.4        | -6.0        | -1.5        | -2.2        | -2.5        |
| <i>Ablation: training objective (ours)</i> |              |             |             |              |             |             |              |             |             |              |             |             |              |             |             |             |             |             |
| Standard Teacher Forcing                   | +0.2         | +0.5        | <b>+1.0</b> | -0.6         | -0.9        | -1.4        | -0.0         | +0.8        | +2.5        | -1.0         | -1.0        | -1.0        | +1.3         | +1.7        | +2.4        | +0.3        | +0.4        | +0.6        |
| Permutation invariance                     | +3.0         | +1.6        | +0.4        | +4.3         | +1.2        | -1.3        | -0.7         | -0.9        | -1.0        | +0.5         | -0.6        | -1.8        | +2.0         | +0.6        | -0.8        | -0.2        | -0.6        | -1.0        |
| Cross-entropy invariance                   | +3.4         | +1.9        | +0.6        | +3.9         | +0.6        | -2.0        | -0.4         | -0.7        | -1.2        | +1.5         | +0.3        | -1.1        | +1.8         | +0.4        | -1.1        | -0.0        | -0.4        | -0.9        |
| <b>MERIT</b>                               | <b>+14.1</b> | <b>+3.4</b> | -0.2        | <b>+15.5</b> | <b>+3.2</b> | -1.4        | <b>+13.8</b> | +4.5        | +0.9        | <b>+11.5</b> | <b>+2.8</b> | -0.7        | <b>+16.2</b> | <b>+5.1</b> | +0.5        | <b>+7.3</b> | +1.9        | +0.1        |

Table 2: Propensity ranking performance (Hit@k). Results show absolute improvement over baseline: DEXML-STk

| Method              | Beauty       |              |              | Apparel     |             |              | Electronics |             |              | Sports      |             |              | Toys        |             |              |
|---------------------|--------------|--------------|--------------|-------------|-------------|--------------|-------------|-------------|--------------|-------------|-------------|--------------|-------------|-------------|--------------|
|                     | @500         | @5k          | @20k         | @500        | @5k         | @20k         | @500        | @5k         | @20k         | @500        | @5k         | @20k         | @500        | @5k         | @20k         |
| DEXML-DS            | +7.8         | +6.8         | +5.2         | +0.4        | +4.4        | +7.5         | -2.4        | +2.9        | +4.7         | -0.4        | -0.9        | +1.1         | -1.6        | -3.3        | -3.0         |
| <b>MERIT (230M)</b> | <b>+20.8</b> | <b>+12.6</b> | <b>+21.3</b> | <b>+1.4</b> | <b>+9.3</b> | <b>+19.7</b> | <b>+3.4</b> | <b>+7.1</b> | <b>+14.0</b> | <b>+3.0</b> | <b>+7.8</b> | <b>+15.8</b> | <b>+4.2</b> | <b>+7.4</b> | <b>+15.8</b> |

2022)) and *Cross-entropy Invariance* (the same, with cross-entropy loss instead) each add only modest recall over Standard Teacher Forcing compared to our objective’s +13.9 recall (same model/data/architecture). To further reinforce the objective’s benefit, MERIT still outperforms aggregating four independent temperature-sampled outputs from the Standard Teacher Forcing model, in both recall (+6.75) and precision (+1.76), ruling out the alternate explanation that gains merely stem from diversity but rather from recovery from erroneous prefixes.

## 4.2 Propensity Model based Ranking

Ranking users for individual fine-grained categories is statistically unstable: most categories have very few true-positive users. We therefore evaluate at a much coarser product-category granularity, also the decision-relevant level for campaign targeting. We aggregate leaf propensities as  $P(g | u) = \sum_c P(g | c) P(c | u)$ , where  $P(g | c)$  is the empirical likelihood that category  $c$  implies product-category  $g$  (precomputed offline) and  $P(c | u)$  comes from the propensity model. Table 2 compares against the same 3.8B phi-3.5-mini dual encoder (Gupta et al., 2024). Full per-vertical results across all 20 verticals and every Hit@k threshold are in Appendix E (Table 9).

Averaged over all  $k$  and verticals, MERIT’s 230M encoder trained on *frozen* generator em-

beddings improves Hit@k by +6.1 (figure reported in abstract) over DEXML-DS despite far fewer parameters, which we attribute to its less-collapsed, propensity-aligned embeddings and the mined negatives from the generator. Crucially, its scores are comparable across users, enabling both user→category (Table 1) and category→user retrieval (Table 2), showing our exposure mitigated generator unlocks bidirectional ranking. Appendix C reports a full sensitivity sweep over  $k_g$  (the number of generated categories pooled into the user representation) and we see that the ranking gains above are not sensitive to this choice.

## 4.3 Production A/B Test

We tested MERIT in online A/B tests selecting users for targeted merchandising campaigns from their predicted propensities. Across 3 one-week campaigns in different verticals (0.6M - 3.5M customers per treatment arm), users were randomly assigned to Treatment or Control, with T segments based on user - campaign propensity and C segments configured by domain experts. Treatments were matched in size, creative, and delivery. Lift was evaluated using a Bayesian, covariate-adjusted online-experiment framework (Kamalbash and Eugster, 2020; Deng, 2015), with the posterior probability of positive lift,  $P(\text{lift} > 0)$ , as the decision metric. All campaigns cleared our internal shipping threshold of  $(P(\text{lift} > 0) \geq 0.90)$

with per-campaign lift of: Personal Care +0.23%, Toys +0.19%, and Sports +0.37% (+0.26% avg.). Detailed credible intervals are reported in App. F (Table 11). Deployment is two-stage: user embeddings are extracted periodically and stored, and interest→user scoring reduces to lightweight MLP lookup over precomputed category embeddings, making the system well-suited to our usecase.

## 5 Conclusion

We presented MERIT, an efficient exposure-mitigated training framework for generative extreme multi-label classification in a user-interest retrieval setting. A single-pass self-correction objective teaches generators to recover from erroneous prefixes without the per-step cost of RL or scheduled sampling, yielding propensity-aligned embeddings that power a lightweight scorer for offline customer segmentation ahead of time. It outperforms strong generative and dual-encoder baselines and delivers meaningful lift in production settings. Since exposure bias may arise in any teacher-forced generator operating over large, structured label spaces, the proposed method offers a general mechanism for improving robustness in generative XMC and multi-label generation tasks.

## 6 Limitations

MERIT trades a portion of precision to optimize for recall gains across all verticals, by design: recall is the operative metric for an upstream retrieval interface, but the raw generated set is noisier and may require downstream filtering before use in precision-sensitive applications.

Our evaluation is on a single proprietary e-commerce dataset, and per internal disclosure policy we report relative rather than absolute metrics, limiting direct comparison to public XMC benchmarks and other generative-XMC literature. Extending MERIT to public XMC benchmarks is left to future work.

Because our target (gold labels) reflect only categories corresponding to a user’s immediate next purchase, mined negatives from §3.1 can occasionally be unobserved true interests; training and evaluating over multiple future purchases rather than a single next-purchase target could yield more stable propensities, and is left to future work

Finally, our current deployment is two-stage and offline: user embeddings are extracted and scored in batch, so the system supports segmentation and

merchandising use cases. Real-time, on-demand scoring while has its benefits has substantial cost implications as well, and a thorough analysis is left to future work.

## Ethical Considerations

MERIT is trained on anonymized user-interaction logs, and interest-category descriptors are derived from product metadata rather than from personal or demographic attributes; we do not model or predict sensitive attributes (e.g., health, financial, or demographic status) as interest categories. Predicted propensities are not shown directly to customers: they feed a downstream filtering rule and eligibility layer (§1) before influencing any customer-facing surface, which also bounds the impact of MERIT’s recall-oriented, higher-noise predictions.

## Acknowledgments

We thank the anonymous reviewers and area chair for their constructive feedback, which improved this paper.

## References

- Keqin Bao, Jizhi Zhang, Yang Zhang, Wenjie Wang, Fuli Feng, and Xiangnan He. 2023. [Tallrec: An effective and efficient tuning framework to align large language model with recommendation](#). In *Proceedings of the 17th ACM Conference on Recommender Systems, RecSys ’23*, page 1007–1014, New York, NY, USA. Association for Computing Machinery.
- Samy Bengio, Oriol Vinyals, Navdeep Jaitly, and Noam Shazeer. 2015. Scheduled sampling for sequence prediction with recurrent neural networks. In *Advances in Neural Information Processing Systems*, volume 28.
- Alexander Brinkmann, Roei Shraga, and Christian Bizer. 2024. [Extractgpt: Exploring the potential of large language models for product attribute value extraction](#). In *Information Integration and Web Intelligence: 26th International Conference, IiWAS 2024, Bratislava, Slovak Republic, December 2–4, 2024, Proceedings, Part I*, page 38–52, Berlin, Heidelberg. Springer-Verlag.
- Wei-Cheng Chang, Hsiang-Fu Yu, Kai Zhong, Yiming Yang, and Inderjit S Dhillon. 2020. Taming pre-trained transformers for extreme multi-label text classification. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 3163–3171.
- Cohere. 2023. Introducing Embed v3. <https://cohere.com/blog/introducing-embed-v3>. Accessed: 2026-06-17.

- Zeyu Cui, Jianxin Ma, Chang Zhou, Jingren Zhou, and Hongxia Yang. 2022. M6-rec: Generative pretrained language models are open-ended recommender systems. In *arXiv preprint arXiv:2205.08084*.
- Kunal Dahiya, Nilesh Gupta, Deepak Saini, Akshay Soni, Yajun Wang, Kushal Dave, Jian Jiao, Gururaj K, Prasenjit Dey, Amit Singh, Deepesh Hada, Vidit Jain, Bhawna Paliwal, Anshul Mittal, Sonu Mehta, Ramachandran Ramjee, Sumeet Agarwal, Purushottam Kar, and Manik Varma. 2023. Ngame: Negative mining-aware mini-batching for extreme classification. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, pages 258–266.
- Kunal Dahiya, Deepak Saini, Anshul Mittal, Ankush Shaw, Kushal Dave, Akshay Soni, Himanshu Jain, Sumeet Agarwal, and Manik Varma. 2021. Deepxml: A deep extreme multi-label learning framework applied to short text documents. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, pages 31–39.
- Alex Deng. 2015. [Objective bayesian two sample hypothesis testing for online controlled experiments](#). In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15 Companion, page 923–928, New York, NY, USA. Association for Computing Machinery.
- Shijie Geng, Shuchang Liu, Zuohui Fu, Yingqiang Ge, and Yongfeng Zhang. 2022a. [Recommendation as language processing \(rlp\): A unified pretrain, personalized prompt & predict paradigm \(p5\)](#). In *Proceedings of the 16th ACM Conference on Recommender Systems*, RecSys '22, page 299–315, New York, NY, USA. Association for Computing Machinery.
- Shijie Geng, Shuchang Liu, Zuohui Fu, Yingqiang Ge, and Yongfeng Zhang. 2022b. Recommendation as language processing (rlp): A unified pretrain, personalized prompt & predict paradigm (p5). In *Proceedings of the 16th ACM conference on recommender systems*, pages 299–315.
- Nilesh Gupta, Fnu Devvrit, Ankit Singh Rawat, Sri-nadh Bhojanapalli, Prateek Jain, and Inderjit Dhillon. 2024. Dual-encoders for extreme multi-label classification. In *International Conference on Learning Representations*, volume 2024, pages 52858–52884.
- Yupeng Hou, Junjie Zhang, Zihan Lin, Hongyu Lu, Ruobing Xie, Julian McAuley, and Wayne Xin Zhao. 2024. Large language models are zero-shot rankers for recommender systems. In *European conference on information retrieval*, pages 364–381. Springer.
- Yunan Hu, Mark Thornburg, Mario Garcia Armas, Vito Ostuni, Anne Cocos, Kriti Kohli, Christoph Kofler, and Rob Saltiel. 2025. [Cold starting a new content type: A case study with netflix live](#). In *Proceedings of the Nineteenth ACM Conference on Recommender Systems*, RecSys '25, page 927–930, New York, NY, USA. Association for Computing Machinery.
- Himanshu Jain, Yashoteja Prabhu, and Manik Varma. 2016. Extreme multi-label loss functions for recommendation, tagging, ranking & other missing label applications. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 935–944.
- Vidit Jain, Jatin Prakash, Deepak Saini, Jian Jiao, Ramachandran Ramjee, and Manik Varma. 2023. Re-nee: End-to-end training of extreme classification models. *Proceedings of Machine Learning and Systems*, 5:646–665.
- Ting Jiang, Deqing Wang, Leilei Sun, Huayi Yang, Zhengyang Zhao, and Fuzhen Zhuang. 2021. Lightxml: Transformer with dynamic negative sampling for high-performance extreme multi-label text classification. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 7987–7994.
- Taehee Jung, Joo-Kyung Kim, Sungjin Lee, and Dongyeop Kang. 2023. Cluster-guided label generation in extreme multi-label classification. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 1670–1685.
- Shafi Kamalbasha and Manuel J. A. Eugster. 2020. [Bayesian a/b testing for business decisions](#). Preprint, arXiv:2003.02769.
- Chao Li, Zhiyuan Liu, Mengmeng Wu, Yuchi Xu, Huan Zhao, Pipei Huang, Guoliang Kang, Qiwei Chen, Wei Li, and Dik Lun Lee. 2019. [Multi-interest network with dynamic routing for recommendation at tmall](#). In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, CIKM '19, page 2615–2623, New York, NY, USA. Association for Computing Machinery.
- Chen Li, Yixiao Ge, Jiayong Mao, Dian Li, and Ying Shan. 2023. [Taggpt: Large language models are zero-shot multimodal taggers](#). *ArXiv*, abs/2304.03022.
- Jiayi Liao, Ruobing Xie, Sihang Li, Xiang Wang, Xingwu Sun, Zhanhui Kang, and Xiangnan He. 2025. Multi-grained patch training for efficient llm-based recommendation. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1572–1581.
- Blerina Lika, Kostas Kolomvatsos, and Stathes Hadjiefthymiades. 2014. [Facing the cold start problem in recommender systems](#). *Expert Systems with Applications*, 41(4, Part 2):2065–2073.
- Jianghao Lin, Rong Shan, Chenxu Zhu, Kounianhua Du, Bo Chen, Shigang Quan, Ruiming Tang, Yong Yu, and Weinan Zhang. 2024. Rella: Retrieval-enhanced large language models for lifelong sequential behavior comprehension in recommendation. In *Proceedings of the ACM Web Conference 2024*, pages 3497–3508.

- Anshul Mittal, Kunal Dahiya, Sheshansh Agrawal, Deepak Saini, Sumeet Agarwal, Purushottam Kar, and Manik Varma. 2021. Decaf: Deep extreme classification with label features. In *Proceedings of the 14th ACM international conference on web search and data mining*, pages 49–57.
- Yashoteja Prabhu, Anil Kag, Shrutendra Harsola, Rahul Agrawal, and Manik Varma. 2018. Parabel: Partitioned label trees for extreme classification with application to dynamic search advertising. In *Proceedings of the 2018 World Wide Web Conference*, pages 993–1002.
- Tao Qi, Fangzhao Wu, Chuhan Wu, Peiru Yang, Yang Yu, Xing Xie, and Yongfeng Huang. 2021. [HieRec: Hierarchical user interest modeling for personalized news recommendation](#). In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 5446–5456, Online. Association for Computational Linguistics.
- Shashank Rajput, Nikhil Mehta, Anima Singh, Raghunandan Keshavan, Trung Vu, Lukasz Heldt, Lichan Hong, Yi Tay, Vinh Q. Tran, Jonah Samost, Maciej Kula, Ed H. Chi, and Maheswaran Sathiamoorthy. 2023. Recommender systems with generative retrieval. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23, Red Hook, NY, USA. Curran Associates Inc.
- Marc’Aurelio Ranzato, Sumit Chopra, Michael Auli, and Wojciech Zaremba. 2015. Sequence level training with recurrent neural networks. *arXiv preprint arXiv:1511.06732*.
- Stéphane Ross, Geoffrey Gordon, and Drew Bagnell. 2011. A reduction of imitation learning and structured prediction to no-regret online learning. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 627–635. JMLR Workshop and Conference Proceedings.
- Daniel Simig, Fabio Petroni, Pouya Yanki, Kashyap Popat, Christina Du, Sebastian Riedel, and Majid Yazdani. 2022. Open vocabulary extreme classification using generative models. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 1561–1583.
- Weiwei Sun, Lingyong Yan, Zheng Chen, Shuaiqiang Wang, Haichao Zhu, Pengjie Ren, Zhumin Chen, Dawei Yin, Maarten Rijke, and Zhaochun Ren. 2023. [Learning to tokenize for generative retrieval](#). In *Advances in Neural Information Processing Systems*, volume 36, pages 46345–46361. Curran Associates, Inc.
- Ruiming Tang, Chenxu Zhu, Bo Chen, Weipeng Zhang, Menghui Zhu, Xinyi Dai, and Huifeng Guo. 2025. [Llm4tag: Automatic tagging system for information retrieval via large language models](#). In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2*, KDD '25, page 4882–4890, New York, NY, USA. Association for Computing Machinery.
- Yi Tay, Vinh Q. Tran, Mostafa Dehghani, Jianmo Ni, Dara Bahri, Harsh Mehta, Zhen Qin, Kai Hui, Zhe Zhao, Jai Gupta, Tal Schuster, William W. Cohen, and Donald Metzler. 2022. Transformer memory as a differentiable search index. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, Red Hook, NY, USA. Curran Associates Inc.
- Ronghui You, Zihan Zhang, Ziyi Wang, Suyang Dai, Hiroshi Mamitsuka, and Shanfeng Zhu. 2019a. [AttentionXML: label tree-based attention-aware deep model for high-performance extreme multi-label text classification](#). Curran Associates Inc., Red Hook, NY, USA.
- Ronghui You, Zihan Zhang, Ziyi Wang, Suyang Dai, Hiroshi Mamitsuka, and Shanfeng Zhu. 2019b. Attentionxml: Label tree-based attention-aware deep model for high-performance extreme multi-label text classification. *Advances in neural information processing systems*, 32.
- Weizhi Zhang, Yuanchen Bei, Liangwei Yang, Henry Peng Zou, Peilin Zhou, Aiwei Liu, Yinghui Li, Hao Chen, Jianling Wang, Yu Wang, Feiran Huang, Sheng Zhou, Jiajun Bu, Allen Lin, James Caverlee, Fakhri Karray, Irwin King, and Philip S. Yu. 2025. [Cold-start recommendation towards the era of large language models \(llms\): A comprehensive survey and roadmap](#). *CoRR*, abs/2501.01945.
- Wen Zhang, Yang Feng, Fandong Meng, Di You, and Qun Liu. 2019. [Bridging the gap between training and inference for neural machine translation](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4334–4343, Florence, Italy. Association for Computational Linguistics.
- Zihao Zhao, Eric Wallace, Shi Feng, Dan Klein, and Sameer Singh. 2021. [Calibrate before use: Improving few-shot performance of language models](#). In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 12697–12706. PMLR.
- Guorui Zhou, Na Mou, Ying Fan, Qi Pi, Weijie Bian, Chang Zhou, Xiaoqiang Zhu, and Kun Gai. 2019. [Deep interest evolution network for click-through rate prediction](#). In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence*, AAAI'19/IAAI'19/EAAI'19. AAAI Press.
- Guorui Zhou, Xiaoqiang Zhu, Chenru Song, Ying Fan, Han Zhu, Xiao Ma, Yanghui Yan, Junqi Jin, Han Li,

and Kun Gai. 2018. [Deep interest network for click-through rate prediction](#). In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '18, page 1059–1068, New York, NY, USA. Association for Computing Machinery.

## Appendix

### A Detailed Walkthrough of Exposure Mitigation Training

To illustrate the exposure mitigation training procedure (Algorithm 1), we present a complete example showing how the self-correction objective operates at each token position. This walkthrough uses all variables from the algorithm to clarify their roles.

#### A.1 Setup

**User Context:** Consider a user with profile frequent sports purchases, occasional home items. Let's say, they purchased a Cricket Bat. Therefore,

**Golden Labels** (ground truth interest categories for the purchased product):

$$Y_c = \{\text{"Sports"}, \text{"Sports Enthusiast"}, \text{"Cricket Fan"}, \text{"Sports Gear"}\}$$

**Mined Hard Negatives** (incorrect but plausible predictions from stage-1 model):

$$\hat{Y}_{\text{filtered}} = \{\text{"Home Decor"}, \text{"Milk"}, \text{"Mobile Phone"}\}$$

**Augmented Label Set:**

$$\mathcal{L}_c = Y_c \cup \hat{Y}_{\text{filtered}}$$

**Tokenization** (simplified for clarity):

- "Sports" → [Sports]
- "Sports Enthusiast" → [Sports, Enthusiast]
- "Cricket Fan" → [Cricket, Fan]
- "Sports Gear" → [Sports, Gear]
- "Home Decor" → [Home, Decor]
- "Milk" → [Milk]
- "Mobile Phone" → [Mobile, Phone]

**Shuffled Target Sequence** (after random shuffle of  $\mathcal{L}_c$ ):

$S = [\text{Home}, \text{Decor}, [\text{SEP}], \text{Sports}, [\text{SEP}], \text{Milk}, [\text{SEP}], \text{Cricket}, \text{Fan}, [\text{SEP}], \text{Sports}, \text{Enthusiast}, [\text{SEP}], \text{Mobile}, \text{Phone}, [\text{SEP}], \text{Sports}, \text{Gear}, [\text{END}]]$

#### A.2 Forward Pass with Loss Computation

**Important:** We perform a *single forward pass* through the model with the input prompt concatenated with the shuffled target sequence  $S$ . This produces next-token logits  $z_1, z_2, \dots, z_{19}$  for each position. We then compute the loss at each position according to our exposure mitigation rules. The model does *not* autoregressively produce tokens during training, this is close to teacher forcing, just with modified loss computation that decides which target tokens probabilities should be penalized.

**Input to model:**

[User Profile] [Recent Interactions]  $S$

**Model output:** Logits  $z_1, z_2, \dots, z_{19}$  representing next-token probability distributions at each position.

We now walk through each position  $t$  in the target sequence  $S$ , showing how we compute the loss  $\ell_t$  on the corresponding logits  $z_t$ .

**Position  $t = 1$ : Token  $y_1 = \text{"Home"}$ , Logits  $z_1$  State:**

- prefix = [] (empty, starting new label)
- completed = {} (no golden labels completed yet)
- $Y_c \setminus \text{completed} = \{\text{Sports}, \text{Sports Enthusiast}, \text{Cricket Fan}, \text{Sports Gear}\}$

**Valid targets  $\mathcal{V}_1$ :** Since prefix is empty, we are at a *classification position*. Valid targets are first tokens of all uncompleted golden labels:

$$\mathcal{V}_1 = \{\text{Sports}, \text{Cricket}\}$$

Note: "Sports" appears as the first token of three labels (Sports, Sports Enthusiast, Sports Gear), and "Cricket" is the first token of Cricket Fan.

**Loss computation on  $z_1$ :**

$$\ell_1 = -\log \left( \frac{\exp(z_1^{\text{Sports}}) + \exp(z_1^{\text{Cricket}})}{\sum_{v'} \exp(z_1^{(v')})} \right)$$

**Discussion:** This is a classification position where we supervise the model to predict correct interest categories based on user context. Although the actual next token in the sequence is "Home" (incorrect), we only reward the logits corresponding to "Sports" and "Cricket" (golden first tokens).

Note, the loss is permutation invariant in the sense that we are not penalizing the model if it is giving high probability to either Sports or Cricket.

**State update:** prefix  $\leftarrow$  [Home]

**Position  $t = 2$ : Token  $y_2 = \text{"Decor"}$ , Logits  $z_2$**

**State:**

- prefix = [Home]
- completed = {}
- $Y_c \setminus \text{completed} = \{\text{Sports, Sports Enthusiast, Cricket Fan, Sports Gear}\}$

**Valid targets  $\mathcal{V}_2$ :** Prefix is non-empty. Check if it matches any golden labels:

$$G = \{\ell \in Y_c \setminus \text{completed} : \ell \text{ starts with "Home"}\} = \emptyset$$

No golden labels start with "Home". This prefix matches the hard negative "Home Decor". For *label coherence*, we complete it:

$$\mathcal{V}_2 = \{\text{Decor}\}$$

**Loss computation on  $z_2$ :**

$$\ell_2 = -\log \left( \frac{\exp(z_2^{\text{Decor}})}{\sum_{v'} \exp(z_2^{(v')})} \right)$$

**Discussion:** Once the prefix becomes "Home" (incorrect), we enforce completion of the full phrase "Home Decor" rather than allowing the loss to reward switching mid-label. This prevents malformed outputs and teaches the model to produce well-formed categories even when the prefix corresponds to an error.

**State update:** prefix  $\leftarrow$  [Home, Decor]

**Position  $t = 3$ : Token  $y_3 = \text{[SEP]}$ , Logits  $z_3$**

**State:**

- prefix = [Home, Decor]
- Prefix forms complete label "Home Decor"  $\in \mathcal{L}_c$
- completed = {} ("Home Decor" is a hard negative, not golden)

**Valid targets  $\mathcal{V}_3$ :** The prefix completes a label (the hard negative), so separator is valid:

$$\mathcal{V}_3 = \{[\text{SEP}]\}$$

**Loss computation on  $z_3$ :**

$$\ell_3 = -\log \left( \frac{\exp(z_3^{[\text{SEP}]})}{\sum_{v'} \exp(z_3^{(v')})} \right)$$

**State update:** Since  $y_3 = [\text{SEP}]$ , reset prefix  $\leftarrow$  []

**Position  $t = 4$ : Token  $y_4 = \text{"Sports"}$ , Logits  $z_4$**

**State:**

- prefix = [] (reset after separator)
- completed = {}
- $Y_c \setminus \text{completed} = \{\text{Sports, Sports Enthusiast, Cricket Fan, Sports Gear}\}$

**Valid targets  $\mathcal{V}_4$ :** Prefix empty, so this is a *classification position*:

$$\mathcal{V}_4 = \{\text{Sports, Cricket}\}$$

**Loss computation on  $z_4$ :**

$$\ell_4 = -\log \left( \frac{\exp(z_4^{\text{Sports}}) + \exp(z_4^{\text{Cricket}})}{\sum_{v'} \exp(z_4^{(v')})} \right)$$

**Discussion:** This is another classification position, but now the consumed prefix contains "Home Decor" (incorrect). The model must learn to make independent classification decisions despite having seen errors in the prefix. This is the core of *exposure mitigation*: training the model to recover from mistakes that appear in the prefix.

**State update:** prefix  $\leftarrow$  [Sports]

**Position  $t = 5$ : Token  $y_5 = \text{[SEP]}$ , Logits  $z_5$**

**State:**

- prefix = [Sports]
- Prefix "Sports" is a complete golden label AND partial match for others
- completed = {}

**Valid targets  $\mathcal{V}_5$ :** Prefix "Sports" matches multiple golden labels:

$$G = \{\text{Sports, Sports Enthusiast, Sports Gear}\}$$

Valid continuations include tokens that extend any of these labels:

- "Sports" is already complete  $\rightarrow$  [SEP] is valid
- "Sports Enthusiast" needs "Enthusiast"
- "Sports Gear" needs "Gear"

Therefore:

$$\mathcal{V}_5 = \{[\text{SEP}], \text{Enthusiast, Gear}\}$$

$$\ell_5 = -\log \left( \left( \exp(z_5^{[\text{SEP}]}) + \exp(z_5^{\text{Enthusiast}}) + \exp(z_5^{\text{Gear}}) \right) / \sum_{v'} \exp(z_5^{(v')}) \right)$$

**Discussion:** This position demonstrates *permutation invariance* and robustness to edge cases. We pool probability mass over all valid options: treating "Sports" as complete, or extending it to "Sports Enthusiast" or "Sports Gear". This avoids penalizing any particular ordering choice.

**State update:** Since  $y_5 = [\text{SEP}]$ ,  $\text{prefix} \leftarrow []$  and  $\text{completed} \leftarrow \{\text{Sports}\}$

**Position  $t = 6$ : Token  $y_6 = \text{"Milk"}$ , Logits  $z_6$**   
**State:**

- $\text{prefix} = []$
- $\text{completed} = \{\text{Sports}\}$
- $Y_c \setminus \text{completed} = \{\text{Sports Enthusiast}, \text{Cricket Fan}, \text{Sports Gear}\}$

**Valid targets  $\mathcal{V}_6$ :** Classification position, first tokens of remaining golden labels:

$$\mathcal{V}_6 = \{\text{Sports}, \text{Cricket}\}$$

**Loss computation on  $z_6$ :**

$$\ell_6 = -\log \left( \frac{\exp(z_6^{\text{Sports}}) + \exp(z_6^{\text{Cricket}})}{\sum_{v'} \exp(z_6^{(v')})} \right)$$

**Discussion:** Another classification position after the prefix has consumed both correct ("Sports") and incorrect ("Home Decor") labels. The loss trains the model to classify independently of prefix correlations.

**State update:**  $\text{prefix} \leftarrow [\text{Milk}]$

**Position  $t = 7$ : Token  $y_7 = [\text{SEP}]$ , Logits  $z_7$**   
**State:**

- $\text{prefix} = [\text{Milk}]$
- "Milk" is a complete hard negative label

**Valid targets  $\mathcal{V}_7$ :**

$$\mathcal{V}_7 = \{[\text{SEP}]\}$$

**Loss computation on  $z_7$ :**

$$\ell_7 = -\log \left( \frac{\exp(z_7^{[\text{SEP}]})}{\sum_{v'} \exp(z_7^{(v')})} \right)$$

**State update:**  $\text{prefix} \leftarrow []$

**Position  $t = 8$ : Token  $y_8 = \text{"Cricket"}$ , Logits  $z_8$**   
**State:**

- $\text{prefix} = []$
- $\text{completed} = \{\text{Sports}\}$
- $Y_c \setminus \text{completed} = \{\text{Sports Enthusiast}, \text{Cricket Fan}, \text{Sports Gear}\}$

**Valid targets  $\mathcal{V}_8$ :**

$$\mathcal{V}_8 = \{\text{Sports}, \text{Cricket}\}$$

**Loss computation on  $z_8$ :**

$$\ell_8 = -\log \left( \frac{\exp(z_8^{\text{Sports}}) + \exp(z_8^{\text{Cricket}})}{\sum_{v'} \exp(z_8^{(v')})} \right)$$

**State update:**  $\text{prefix} \leftarrow [\text{Cricket}]$

**Position  $t = 9$ : Token  $y_9 = \text{"Fan"}$ , Logits  $z_9$**   
**State:**

- $\text{prefix} = [\text{Cricket}]$
- $\text{completed} = \{\text{Sports}\}$

**Valid targets  $\mathcal{V}_9$ :** Prefix "Cricket" matches exactly one golden label "Cricket Fan":

$$G = \{\text{Cricket Fan}\} \Rightarrow \mathcal{V}_9 = \{\text{Fan}\}$$

**Loss computation on  $z_9$ :**

$$\ell_9 = -\log \left( \frac{\exp(z_9^{\text{Fan}})}{\sum_{v'} \exp(z_9^{(v')})} \right)$$

**State update:**  $\text{prefix} \leftarrow [\text{Cricket}, \text{Fan}]$

**Position  $t = 10$ : Token  $y_{10} = [\text{SEP}]$ , Logits  $z_{10}$**   
**State:**

- $\text{prefix} = [\text{Cricket}, \text{Fan}]$
- Prefix completes golden label "Cricket Fan"

**Valid targets  $\mathcal{V}_{10}$ :**

$$\mathcal{V}_{10} = \{[\text{SEP}]\}$$

**Loss computation on  $z_{10}$ :**

$$\ell_{10} = -\log \left( \frac{\exp(z_{10}^{[\text{SEP}]})}{\sum_{v'} \exp(z_{10}^{(v')})} \right)$$

**State update:**  $\text{prefix} \leftarrow []$ ,  $\text{completed} \leftarrow \{\text{Sports}, \text{Cricket Fan}\}$

**Position  $t = 11$ : Token  $y_{11} = \text{"Sports"}$ , Logits  $z_{11}$**   
**State:**

- prefix = []
- completed = {Sports, Cricket Fan}
- $Y_c \setminus \text{completed} = \{\text{Sports Enthusiast}, \text{Sports Gear}\}$

**Valid targets  $\mathcal{V}_{11}$ :** Both remaining golden labels start with "Sports":

$$\mathcal{V}_{11} = \{\text{Sports}\}$$

**Loss computation on  $z_{11}$ :**

$$\ell_{11} = -\log \left( \frac{\exp(z_{11}^{\text{Sports}})}{\sum_{v'} \exp(z_{11}^{(v')})} \right)$$

**State update:** prefix  $\leftarrow [\text{Sports}]$

**Position  $t = 12$ : Token  $y_{12} = \text{"Enthusiast"}$ , Logits  $z_{12}$**

**State:**

- prefix = [Sports]
- completed = {Sports, Cricket Fan}
- Remaining: {Sports Enthusiast, Sports Gear}

**Valid targets  $\mathcal{V}_{12}$ :** Prefix "Sports" matches both remaining labels:

$$G = \{\text{Sports Enthusiast}, \text{Sports Gear}\} \\ \Rightarrow \mathcal{V}_{12} = \{\text{Enthusiast}, \text{Gear}\}$$

**Loss computation on  $z_{12}$ :**

$$\ell_{12} = -\log \left( \frac{\exp(z_{12}^{\text{Enthusiast}}) + \exp(z_{12}^{\text{Gear}})}{\sum_{v'} \exp(z_{12}^{(v')})} \right)$$

**Discussion:** Permutation invariance, the loss pools probability over both "Enthusiast" and "Gear" since either continuation is valid.

**State update:** prefix  $\leftarrow [\text{Sports}, \text{Enthusiast}]$

**Position  $t = 13$ : Token  $y_{13} = [\text{SEP}]$ , Logits  $z_{13}$**

**State:**

- prefix = [Sports, Enthusiast]
- Prefix completes "Sports Enthusiast"

**Valid targets  $\mathcal{V}_{13}$ :**

$$\mathcal{V}_{13} = \{[\text{SEP}]\}$$

**Loss computation on  $z_{13}$ :**

$$\ell_{13} = -\log \left( \frac{\exp(z_{13}^{[\text{SEP}]})}{\sum_{v'} \exp(z_{13}^{(v')})} \right)$$

**State update:** prefix  $\leftarrow []$ ,  
completed  $\leftarrow \{\text{Sports}, \text{Cricket Fan}, \text{Sports Enthusiast}\}$

**Position  $t = 14$ : Token  $y_{14} = \text{"Mobile"}$ , Logits  $z_{14}$**

**State:**

- prefix = [] (reset after separator)
- completed = {Sports, Cricket Fan, Sports Enthusiast}
- $Y_c \setminus \text{completed} = \{\text{Sports Gear}\}$

**Valid targets  $\mathcal{V}_{14}$ :** Classification position, only one golden label remains:

$$\mathcal{V}_{14} = \{\text{Sports}\}$$

Only "Sports Gear" remains, which starts with "Sports".

**Loss computation on  $z_{14}$ :**

$$\ell_{14} = -\log \left( \frac{\exp(z_{14}^{\text{Sports}})}{\sum_{v'} \exp(z_{14}^{(v')})} \right)$$

**Discussion:** Another classification position demonstrating exposure mitigation.

**State update:** prefix  $\leftarrow [\text{Mobile}]$

**Position  $t = 15$ : Token  $y_{15} = \text{"Phone"}$ , Logits  $z_{15}$**

**State:**

- prefix = [Mobile]
- completed = {Sports, Cricket Fan, Sports Enthusiast}

**Valid targets  $\mathcal{V}_{15}$ :** Prefix "Mobile" matches hard negative "Mobile Phone":

$$G = \emptyset \Rightarrow \mathcal{V}_{15} = \{\text{Phone}\}$$

No golden labels match, so complete the hard negative for coherence.

**Loss computation on  $z_{15}$ :**

$$\ell_{15} = -\log \left( \frac{\exp(z_{15}^{\text{Phone}})}{\sum_{v'} \exp(z_{15}^{(v')})} \right)$$

**State update:** prefix  $\leftarrow [\text{Mobile}, \text{Phone}]$

**Position  $t = 16$ : Token  $y_{16} = [\text{SEP}]$ , Logits  $z_{16}$**

**State:**

- prefix = [Mobile, Phone]

- "Mobile Phone" is complete

**Valid targets  $\mathcal{V}_{16}$ :**

$$\mathcal{V}_{16} = \{[\text{SEP}]\}$$

**Loss computation on  $z_{16}$ :**

$$\ell_{16} = -\log \left( \frac{\exp(z_{16}^{[\text{SEP}]})}{\sum_{v'} \exp(z_{16}^{(v')})} \right)$$

**State update:** prefix  $\leftarrow []$

**Position  $t = 17$ : Token  $y_{17} = \text{"Sports"}$ , Logits**

$z_{17}$

**State:**

- prefix = []
- completed = {Sports, Cricket Fan, Sports Enthusiast}
- $Y_c \setminus \text{completed} = \{\text{Sports Gear}\}$

**Valid targets  $\mathcal{V}_{17}$ :** Only one golden label remains:

$$\mathcal{V}_{17} = \{\text{Sports}\}$$

**Loss computation on  $z_{17}$ :**

$$\ell_{17} = -\log \left( \frac{\exp(z_{17}^{\text{Sports}})}{\sum_{v'} \exp(z_{17}^{(v')})} \right)$$

**State update:** prefix  $\leftarrow [\text{Sports}]$

**Position  $t = 18$ : Token  $y_{18} = \text{"Gear"}$ , Logits**

$z_{18}$

**State:**

- prefix = [Sports]
- completed = {Sports, Cricket Fan, Sports Enthusiast}
- Remaining: {Sports Gear}

**Valid targets  $\mathcal{V}_{18}$ :** Prefix "Sports" matches the last remaining golden label:

$$G = \{\text{Sports Gear}\} \Rightarrow \mathcal{V}_{18} = \{\text{Gear}\}$$

**Loss computation on  $z_{18}$ :**

$$\ell_{18} = -\log \left( \frac{\exp(z_{18}^{\text{Gear}})}{\sum_{v'} \exp(z_{18}^{(v')})} \right)$$

**State update:** prefix  $\leftarrow [\text{Sports}, \text{Gear}]$

**Position  $t = 19$ : Token  $y_{19} = [\text{END}]$ , Logits  $z_{19}$**

**State:**

- prefix = [Sports, Gear]

- completed = {Sports, Cricket Fan, Sports Enthusiast}

- All golden labels will be complete after this position

**Valid targets  $\mathcal{V}_{19}$ :** "Sports Gear" is complete. Since all golden labels are now completed, only the end token is valid:

$$\mathcal{V}_{19} = \{[\text{END}]\}$$

**Loss computation on  $z_{19}$ :**

$$\ell_{19} = -\log \left( \frac{\exp(z_{19}^{[\text{END}]})}{\sum_{v'} \exp(z_{19}^{(v')})} \right)$$

**State update:**

completed  $\leftarrow \{\text{Sports, Cricket Fan, Sports Enthusiast, Sports Gear}\}$   
(all golden labels complete)

**Final Loss:**

$$\mathcal{L} = \frac{1}{19} \sum_{t=1}^{19} \ell_t$$

### A.3 Key Insights

**Exposure mitigation:** Classification positions ( $t=1,4,6,8,11,14,17$ ) occur after the prefix has consumed both correct and incorrect labels. The model learns to make independent classifications based on user context rather than following prefix correlations. Crucially, this means hidden states at classification positions encode genuine user-interest propensities independent of the prefix, explaining why the Propensity Model achieves strong ranking performance despite using representations extracted from a generative task. These representations are propensity-aligned by design.

**Permutation invariance:** Positions like  $t=5,12$  pool probability mass over multiple valid continuations, avoiding penalties for arbitrary ordering choices.

**Label coherence:** Positions like  $t=2,7$  enforce completion of hard negatives once started, preventing malformed outputs.

**Training dynamics:** Early in training, completion losses ( $t=2,9,12$ ) are high as the model learns the label vocabulary. These quickly decrease. Classification losses ( $t=1,4,6,8,11$ ) remain challenging throughout training as the model learns which interests match the user profile from 250k+ candidates.

This natural concentration of learning signal at classification positions is exactly what suppresses exposure bias.

## B Threshold Sensitivity and Human Agreement

To validate the  $\tau=0.7$  soft-matching protocol used throughout the paper, we (1) re-evaluate all methods across a range of thresholds and under normalized Exact Match, and (2) collect human judgments on a sample of boundary pairs.

**Threshold sensitivity.** Table 3 is from an independent re-evaluation run conducted for this analysis; absolute values at  $\tau=0.70$  therefore differ marginally (within 0.6pp) from Table 1’s soft-matching numbers due to ordinary run-to-run variance, with Table 1 the authoritative source for the paper’s headline results. Method rankings and conclusions are unchanged. The table reports absolute improvement over the Trivial baseline, pooled globally across verticals, at  $\tau \in \{0.70, 0.80, 0.90\}$  and under normalized Exact Match.  $\tau$ -AUC is the normalized trapezoidal area under the improvement-versus-threshold curve over  $\tau \in [0.70, 0.90]$ . MERIT’s recall advantage holds under every criterion, including Exact Match, and relative rankings across methods are unchanged across  $\tau$ . Some baselines edge out MERIT on precision at stricter thresholds, which is acceptable for the retrieval setting of this work: recall is our operating metric for the generative model, as downstream filtering is cheap and missed candidates are irrecoverable. Propensity ranking (Table 2, main paper) and the A/B campaigns are unaffected by this choice, since they aggregate leaf propensities via the structural likelihood in §4.2, not soft-matching.

**Human agreement.** We sampled 230 interest-category pairs directly around the  $\tau=0.7$  decision boundary to evaluate the hardest edge cases, rather than an uninformative random sample. Human annotators evaluating semantic equivalence judged 96.2% of pairs accepted at  $\tau=0.7$  as at least loosely equivalent (similar purchase intent), with 65.4% judged as strictly equivalent; errors at the boundary were concentrated on near-synonymy rather than completely unrelated categories.

## C Sensitivity to $k_g$

$k_g$  is the number of first-token hidden states from generated categories that are pooled into the propensity model’s user representation (§3.3);

the paper uses  $k_g=5$ . Table 4 sweeps  $k_g \in \{1, 3, 5, 7, 10\}$  across all 20 product verticals and every Hit@ $k$  cutoff, reporting the fraction of vertical-cutoff cells where MERIT beats both DEXML-DS and DEXML-STk simultaneously; every setting beats both baselines in 89–94% of the 100 vertical-cutoff cells, so the choice is not fragile. Table 5 reports the mean Hit@ $k$  improvement over DEXML-STk for representative settings:  $k_g=3$  is marginally strongest overall (+15.0),  $k_g=5$  (used throughout the paper) is close behind (+13.2), and  $k_g=10$  is somewhat lower (+12.4), all far ahead of the strongest baseline, DEXML-DS (mean gain +5.2).

## D Full Per-Vertical Generative-Classification Results

Table 1 reports five representative verticals; Tables 6–8 give all 20 verticals plus the Global aggregate, split across three tables for readability, with Recall/F1/Precision kept together per vertical.

## E Full Per-Vertical Propensity-Ranking Results

Table 2 reports five verticals at three operating points; Table 9 gives all 20 verticals across every Hit@ $k$  threshold, for both DEXML-DS and MERIT’s 230M propensity model, each expressed relative to DEXML-STk. Averaged over the 20 verticals, MERIT improves Hit@5k by +9.0 points over DEXML-STk and +4.3 over DEXML-DS.

## F Production A/B Test Details

Table 11 reports per-campaign A/B-test results, complementing the summary in §4.3. Each campaign ran for one week, with users randomly assigned (50:50) to treatment (propensity-model-derived segments) or control (expert-configured segments). Treatment and control arms were matched on segment size, creative, and delivery, so segment quality is the only differing factor. Lift is evaluated with a Bayesian covariate-adjusted framework; the primary decision metric is the posterior probability of positive conversion lift,  $P(\text{lift} > 0)$ , rather than exclusion of zero from the 95% credible interval. Our experimentation pipeline considers a treatment shippable when  $P(\text{lift} > 0) \geq 0.90$ , corresponding to 9-to-1 posterior odds in favor of positive impact. All three campaigns clear this pre-specified decision threshold. Two campaigns have 95% credible intervals

Table 3: Generative classification performance (Precision, Recall, F1) under different prediction vs. ground-truth matching criteria ( $\tau$  is semantic matching threshold). Absolute improvement over the trivial baseline aggregated over all product verticals is reported.  $\tau$ -AUC is the normalized trapezoidal area under the improvement-versus-threshold curve over  $\tau \in [0.70, 0.90]$ .

| Method    | $\tau=0.70$  |             |             | $\tau=0.80$  |             |             | $\tau=0.90$ |             |             | $\tau=\text{exact}$ |             |             | $\tau\text{-AUC}$ |             |             |
|-----------|--------------|-------------|-------------|--------------|-------------|-------------|-------------|-------------|-------------|---------------------|-------------|-------------|-------------------|-------------|-------------|
|           | $\Delta R$   | $\Delta F1$ | $\Delta P$  | $\Delta R$   | $\Delta F1$ | $\Delta P$  | $\Delta R$  | $\Delta F1$ | $\Delta P$  | $\Delta R$          | $\Delta F1$ | $\Delta P$  | $\Delta R$        | $\Delta F1$ | $\Delta P$  |
| Trivial   | +0.0         | +0.0        | +0.0        | +0.0         | +0.0        | +0.0        | +0.0        | +0.0        | +0.0        | +0.0                | +0.0        | +0.0        | +0.0              | +0.0        | +0.0        |
| XLGen     | +2.0         | +2.0        | +1.9        | +1.2         | +1.0        | +0.7        | +0.9        | +0.7        | +0.3        | +0.8                | +0.7        | +0.5        | +1.3              | +1.2        | +0.9        |
| GROOV     | +1.8         | +2.3        | +3.1        | +1.2         | +1.5        | +1.9        | +1.0        | +1.2        | +1.5        | +1.1                | <b>+1.2</b> | <b>+1.5</b> | +1.3              | +1.6        | +2.1        |
| PatchRec  | -0.1         | +0.8        | +2.5        | -0.7         | -0.2        | +0.8        | -0.8        | -0.4        | +0.4        | +0.1                | +0.4        | +1.1        | -0.6              | +0.0        | +1.1        |
| DEXML-DS  | +1.1         | +2.0        | <b>+3.7</b> | +0.7         | +1.5        | <b>+2.9</b> | +0.7        | <b>+1.3</b> | <b>+2.6</b> | -0.1                | -0.1        | -0.1        | +0.8              | +1.6        | <b>+3.0</b> |
| DEXML-STk | -1.3         | -4.3        | -5.4        | -2.5         | -4.2        | -5.1        | -3.1        | -4.3        | -5.1        | -1.6                | -2.1        | -2.5        | -2.3              | -4.2        | -5.2        |
| MERIT     | <b>+14.1</b> | <b>+3.4</b> | -0.3        | <b>+10.3</b> | <b>+1.6</b> | -1.4        | <b>+9.1</b> | +1.0        | -1.7        | <b>+5.5</b>         | +0.9        | -0.6        | <b>+11.0</b>      | <b>+1.9</b> | -1.2        |

Table 4: Sensitivity to  $k_g$ : fraction of the 20 product verticals, per Hit@ $k$  cutoff, where the given  $k_g$  beats both DEXML-DS and DEXML-STk simultaneously. **Overall** aggregates all 100 vertical-cutoff cells.

| $k_g$     | @0.5k | @1k   | @5k   | @10k  | @20k  | Overall |
|-----------|-------|-------|-------|-------|-------|---------|
| 1         | 16/20 | 17/20 | 18/20 | 20/20 | 20/20 | 91%     |
| 3         | 18/20 | 18/20 | 19/20 | 19/20 | 20/20 | 94%     |
| 5 (paper) | 17/20 | 17/20 | 19/20 | 19/20 | 19/20 | 91%     |
| 7         | 16/20 | 16/20 | 19/20 | 19/20 | 19/20 | 89%     |
| 10        | 17/20 | 17/20 | 18/20 | 19/20 | 19/20 | 90%     |

that overlap zero, while Campaign 3’s interval lies entirely above zero; we therefore interpret the on-line results as positive Bayesian decision evidence across campaigns, with strongest individual evidence in Campaign 3.

## G Adaptive Patching Details

**Compression strategy.** We apply non-overlapping window pooling (window size  $W = 10$ , stride  $S = 10$ ) to token embeddings in older interaction windows, achieving  $10\times$  compression in those segments. This fits substantially more history, often the full year of interactions, within the same context budget that would otherwise hold only the most recent months. Unlike aggressive session-level compression, this preserves finer temporal granularity while prioritizing recent interactions, which remain uncompressed to retain the strongest predictive signal.

**Curriculum learning.** Training on patched sequences from the start degrades performance, as the model must simultaneously learn the prediction task and adapt to compressed representations. We therefore introduce patching gradually over four phases: Phase 1 (0-25%): full sequences only; Phase 2 (25-50%): patching applied with linearly increasing probability  $p(t)$ ; Phase 3 (50-

75%): patching applied consistently to older interaction windows; Phase 4 (75-100%): training exclusively on patched sequences. This curriculum lets the model first master the prediction task on uncompressed data before adapting to compressed inputs.

**Results.** Table 10 reports the sequence-compression ablation. Patching the full year of history into the 8k budget yields a small but consistent recall gain over using only the most recent interactions (+0.5 global recall over 4k-recent-only), confirming that the added history helps and that compression does not degrade the retained recent signal.

## H Implementation Details and Hyperparameters

**Generator.** MERIT uses phi-3.5-mini (3.8B) as the base generator, fine-tuned with our exposure-mitigated objective and adaptive patching. Following (Jung et al., 2023), we manage the label space via hierarchical generation: the 291k leaf categories are organized into 30k semantic clusters, and the model predicts clusters before leaf-level categories.

**Propensity model.** The propensity model is a 4-layer transformer encoder followed by a 3-layer

Table 5: Mean Hit@ $k$  improvement (absolute, averaged across all 20 verticals) over DEXML-STk, for representative  $k_g$  settings; DEXML-DS shown for reference.

| Method                 | @0.5k        | @1k          | @5k          | @10k         | @20k         | Mean         |
|------------------------|--------------|--------------|--------------|--------------|--------------|--------------|
| DEXML-DS               | +5.5         | +5.3         | +4.7         | +5.3         | +4.9         | +5.2         |
| MERIT, $k_g=3$         | <b>+14.3</b> | <b>+12.0</b> | <b>+12.5</b> | <b>+16.4</b> | <b>+19.8</b> | <b>+15.0</b> |
| MERIT, $k_g=5$ (paper) | +11.6        | +10.3        | +11.1        | +14.9        | +18.2        | +13.2        |
| MERIT, $k_g=10$        | +11.4        | +10.0        | +10.5        | +13.9        | +16.3        | +12.4        |

Table 6: Extended generative-classification results, part A (absolute pp improvement over the trivial baseline). Companion to Table 1; baseline short-names are defined in §4.

| Method                                     | Global       |             |             | Beauty       |             |             | Apparel      |             |             | Wireless     |              |              | Drugstore    |             |             | Kitchen      |             |             | Home         |              |              |
|--------------------------------------------|--------------|-------------|-------------|--------------|-------------|-------------|--------------|-------------|-------------|--------------|--------------|--------------|--------------|-------------|-------------|--------------|-------------|-------------|--------------|--------------|--------------|
|                                            | R            | F1          | P           | R            | F1          | P           | R            | F1          | P           | R            | F1           | P            | R            | F1          | P           | R            | F1          | P           | R            | F1           | P            |
| <i>Baselines</i>                           |              |             |             |              |             |             |              |             |             |              |              |              |              |             |             |              |             |             |              |              |              |
| XLGen                                      | +2.1         | +2.0        | +1.9        | +2.8         | +2.2        | +1.4        | +3.5         | +4.0        | +4.7        | +6.9         | +5.8         | +4.4         | +4.3         | +5.0        | +5.8        | -3.3         | -3.6        | -3.9        | +0.9         | +0.7         | +0.5         |
| GROOV                                      | +1.8         | +2.3        | +2.9        | +2.3         | +2.4        | +2.4        | +3.9         | <b>+5.3</b> | <b>+7.5</b> | +3.3         | +3.2         | +2.9         | +3.4         | +4.6        | <b>+6.4</b> | -3.5         | -3.4        | -3.2        | +1.8         | +2.0         | +2.2         |
| PatchRec                                   | +0.2         | +1.2        | +3.1        | -2.3         | -2.3        | -2.1        | -0.2         | +1.2        | +5.1        | +4.5         | +5.1         | +5.9         | -2.3         | -1.9        | -1.0        | -4.3         | -3.4        | <b>-1.8</b> | +8.3         | +11.4        | <b>+17.0</b> |
| DEXML-DS                                   | +1.2         | +2.2        | <b>+3.8</b> | +2.6         | +3.1        | <b>+3.7</b> | +1.8         | +2.6        | +3.9        | +11.5        | <b>+15.4</b> | <b>+22.6</b> | +1.8         | +2.7        | +4.0        | -3.8         | -3.7        | -3.6        | -0.8         | -0.6         | -0.3         |
| DEXML-STk                                  | -0.9         | -4.1        | -5.2        | +4.1         | -3.1        | -5.7        | -1.8         | -4.2        | -5.1        | -5.3         | -6.1         | -7.2         | +1.3         | -1.9        | -2.9        | -4.2         | -7.6        | -8.6        | +1.1         | -1.9         | -2.7         |
| <i>Ablation: training objective (ours)</i> |              |             |             |              |             |             |              |             |             |              |              |              |              |             |             |              |             |             |              |              |              |
| Standard Teacher Forcing                   | +0.2         | +0.5        | +1.0        | -0.6         | -0.9        | -1.4        | -0.0         | +0.8        | +2.5        | +4.0         | +3.8         | +3.3         | +1.4         | +2.2        | +3.4        | -4.1         | -3.8        | -3.5        | +3.1         | +3.5         | +4.2         |
| Permutation invariance                     | +3.0         | +1.6        | +0.4        | +4.3         | +1.2        | -1.3        | -0.7         | -0.9        | -1.0        | +13.5        | +7.9         | +4.0         | -0.2         | -0.4        | -0.7        | -0.3         | -1.1        | -1.9        | +12.3        | +10.5        | +8.9         |
| Cross-entropy Invariance                   | +3.4         | +1.9        | +0.6        | +3.9         | +0.6        | -2.0        | -0.4         | -0.7        | -1.2        | +15.9        | +9.4         | +4.9         | +0.7         | +0.2        | -0.2        | -0.5         | -1.3        | -2.1        | +12.4        | <b>+11.9</b> | +11.5        |
| <b>MERIT (1-beam)</b>                      | <b>+14.1</b> | <b>+3.4</b> | -0.2        | <b>+15.5</b> | <b>+3.2</b> | -1.4        | <b>+13.8</b> | +4.5        | +0.9        | <b>+22.6</b> | +5.5         | +0.1         | <b>+15.9</b> | <b>+6.5</b> | +3.0        | <b>+11.4</b> | <b>-0.8</b> | -4.3        | <b>+13.7</b> | +3.7         | +0.9         |

Table 7: Extended generative-classification results, part B (absolute pp improvement over the trivial baseline). Companion to Table 1; baseline short-names are defined in §4.

| Method                                     | PC           |             |             | Grocery      |             |             | Shoes        |             |             | Book        |             |             | Wireless Acc. |             |             | Electronics  |             |             | Home Improv. |              |              |
|--------------------------------------------|--------------|-------------|-------------|--------------|-------------|-------------|--------------|-------------|-------------|-------------|-------------|-------------|---------------|-------------|-------------|--------------|-------------|-------------|--------------|--------------|--------------|
|                                            | R            | F1          | P           | R            | F1          | P           | R            | F1          | P           | R           | F1          | P           | R             | F1          | P           | R            | F1          | P           | R            | F1           | P            |
| <i>Baselines</i>                           |              |             |             |              |             |             |              |             |             |             |             |             |               |             |             |              |             |             |              |              |              |
| XLGen                                      | +0.8         | +0.6        | +0.2        | +2.7         | +1.8        | +0.3        | +1.3         | +1.3        | +1.1        | +0.3        | +0.2        | +0.1        | +1.4          | +1.1        | +0.8        | +1.1         | +0.9        | +0.5        | +1.4         | +1.3         | +1.3         |
| GROOV                                      | +0.3         | +0.5        | +0.8        | +1.8         | +1.7        | <b>+1.3</b> | +2.2         | +3.1        | <b>+4.9</b> | +3.5        | <b>+4.1</b> | <b>+4.8</b> | +0.7          | +0.8        | +1.0        | +0.2         | +0.3        | +0.6        | +0.9         | +1.3         | +1.7         |
| PatchRec                                   | -2.1         | -2.1        | -1.8        | -3.8         | -4.3        | -4.6        | -1.4         | -0.5        | +3.2        | -1.8        | -1.1        | +0.0        | -1.5          | -1.4        | -0.9        | -1.0         | -0.9        | -0.6        | +12.5        | <b>+15.3</b> | <b>+19.2</b> |
| DEXML-DS                                   | +1.9         | <b>+3.7</b> | <b>+8.5</b> | -0.0         | -0.8        | -2.2        | -0.5         | -0.5        | -0.6        | -3.1        | -2.8        | -2.4        | +1.1          | +2.3        | <b>+5.0</b> | +1.6         | +2.6        | <b>+4.5</b> | +0.2         | +0.7         | +1.2         |
| DEXML-STk                                  | -3.9         | -5.5        | -6.8        | -3.5         | -6.8        | -9.5        | +2.9         | -5.0        | -8.5        | -4.8        | -5.0        | -5.1        | -3.6          | -4.3        | -4.9        | -3.2         | -4.4        | -5.4        | +0.3         | -2.6         | -3.0         |
| <i>Ablation: training objective (ours)</i> |              |             |             |              |             |             |              |             |             |             |             |             |               |             |             |              |             |             |              |              |              |
| Standard Teacher Forcing                   | -0.4         | -0.3        | -0.3        | -3.2         | -4.1        | -5.5        | -3.2         | -3.7        | -4.3        | +0.0        | +0.4        | +0.8        | -0.3          | -0.2        | +0.0        | -1.0         | -1.0        | -1.0        | +2.5         | +3.1         | +3.7         |
| Permutation invariance                     | +1.5         | -0.5        | -2.3        | -3.7         | -5.4        | -7.8        | -2.3         | -3.8        | -6.0        | +0.4        | -1.1        | -1.9        | +2.0          | +1.0        | +0.0        | +0.5         | -0.6        | -1.8        | +16.2        | +12.6        | +10.2        |
| Cross-entropy Invariance                   | +2.1         | -0.3        | -2.4        | -3.0         | -5.1        | -7.8        | -2.4         | -4.1        | -6.5        | +2.3        | -0.0        | -1.2        | +4.0          | +2.8        | +1.6        | +1.5         | +0.3        | -1.1        | +14.8        | +11.7        | +9.6         |
| <b>MERIT (1-beam)</b>                      | <b>+13.2</b> | +2.7        | -1.4        | <b>+14.7</b> | <b>+2.8</b> | -3.0        | <b>+13.6</b> | <b>+3.6</b> | -2.1        | <b>+8.4</b> | +0.5        | -1.5        | <b>+12.9</b>  | <b>+3.3</b> | +0.0        | <b>+11.5</b> | <b>+2.8</b> | -0.7        | <b>+16.8</b> | +4.2         | +1.2         |

Table 8: Extended generative-classification results, part C (absolute pp improvement over the trivial baseline). Companion to Table 1; baseline short-names are defined in §4.

| Method                                     | Automotive   |              |              | Baby Product |             |              | Sports       |             |             | Toys        |             |             | BISS         |             |             | Lawn & Garden |             |             | Personal Care |             |             |
|--------------------------------------------|--------------|--------------|--------------|--------------|-------------|--------------|--------------|-------------|-------------|-------------|-------------|-------------|--------------|-------------|-------------|---------------|-------------|-------------|---------------|-------------|-------------|
|                                            | R            | F1           | P            | R            | F1          | P            | R            | F1          | P           | R           | F1          | P           | R            | F1          | P           | R             | F1          | P           | R             | F1          | P           |
| <i>Baselines</i>                           |              |              |              |              |             |              |              |             |             |             |             |             |              |             |             |               |             |             |               |             |             |
| XLGen                                      | +1.1         | +1.0         | +0.9         | +5.3         | +5.9        | +6.6         | +1.7         | +1.6        | +1.4        | +0.5        | +0.4        | +0.3        | +0.9         | +0.8        | +0.7        | +2.2          | +2.3        | +2.5        | +2.1          | +1.8        | +1.5        |
| GROOV                                      | +0.8         | +1.1         | +1.3         | -0.4         | +0.6        | +1.9         | +2.2         | +2.6        | <b>+3.3</b> | +2.0        | <b>+2.3</b> | <b>+2.7</b> | +0.6         | +0.9        | +1.3        | +0.5          | +1.0        | +1.7        | +3.1          | +3.5        | <b>+4.0</b> |
| PatchRec                                   | +11.3        | <b>+14.3</b> | <b>+18.7</b> | -4.4         | -3.9        | -3.2         | -1.1         | -0.6        | +0.7        | -1.2        | -1.1        | -0.7        | +3.0         | +4.4        | <b>+6.1</b> | +4.3          | +6.0        | <b>+8.7</b> | -2.3          | -2.4        | -2.3        |
| DEXML-DS                                   | +3.3         | +4.7         | +6.6         | +3.4         | <b>+6.1</b> | <b>+10.6</b> | -0.7         | -0.5        | -0.1        | -0.3        | -0.2        | -0.2        | -0.3         | +0.1        | +0.5        | +1.6          | +2.2        | +3.2        | +3.2          | +3.6        | <b>+4.0</b> |
| DEXML-STk                                  | +1.1         | -2.3         | -2.9         | +7.7         | -3.0        | -5.2         | -0.9         | -4.4        | -6.0        | -1.5        | -2.2        | -2.5        | -1.6         | -3.4        | -3.7        | -1.3          | -3.4        | -4.0        | +0.3          | -2.3        | -3.1        |
| <i>Ablation: training objective (ours)</i> |              |              |              |              |             |              |              |             |             |             |             |             |              |             |             |               |             |             |               |             |             |
| Standard Teacher Forcing                   | +1.1         | +1.3         | +1.6         | +2.0         | +3.2        | +5.0         | +1.3         | +1.7        | +2.4        | +0.3        | +0.4        | +0.6        | +0.6         | +1.1        | +1.7        | +3.1          | +3.8        | +4.7        | +0.2          | +0.3        | +0.4        |
| Permutation invariance                     | +2.5         | +1.3         | +0.6         | +7.4         | +5.9        | +4.7         | +2.0         | +0.6        | -0.8        | -0.2        | -0.6        | -1.0        | +9.4         | <b>+7.0</b> | +5.4        | +5.6          | +3.4        | +1.9        | -0.8          | -1.4        | -1.9        |
| Cross-entropy Invariance                   | +10.2        | +8.1         | +6.6         | -3.5         | -4.6        | -5.3         | +1.8         | +0.4        | -1.1        | -0.0        | -0.4        | -0.9        | +4.1         | +2.1        | +0.9        | +8.4          | <b>+6.3</b> | +4.7        | +0.1          | -0.6        | -1.2        |
| <b>MERIT (1-beam)</b>                      | <b>+11.5</b> | +2.4         | +0.1         | <b>+14.6</b> | +3.1        | -0.8         | <b>+16.2</b> | <b>+5.1</b> | +0.5        | <b>+7.3</b> | +1.9        | +0.1        | <b>+12.2</b> | +2.5        | +0.1        | <b>+13.7</b>  | +4.0        | +0.9        | <b>+14.7</b>  | <b>+4.3</b> | +1.2        |

MLP (230M parameters total), trained on the first-token hidden states ( $k_g=5$ ) extracted from the frozen generator. For categories we use pretrained Cohere (Cohere, 2023) embeddings; the propen-

sity score is the MLP output over the concatenated (user, category) representation.

**Optimization.** We use AdamW (learning rate  $2 \times 10^{-5}$ ), batch size 32, and mixed precision (fp16) on

Table 9: Extended propensity-ranking results (Hit@ $k$ , absolute pp improvement over the DEXML-STk baseline) across all 20 verticals and all  $k$ . Companion to Table 2; short-names are defined in §4. Both method blocks are expressed relative to DEXML-STk; MERIT uses the 230M propensity model on frozen generator embeddings.

| Vertical      | DEXML-DS     |              |              |              |              | MERIT (230M) |              |              |              |              |
|---------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
|               | @0.5k        | @1k          | @5k          | @10k         | @20k         | @0.5k        | @1k          | @5k          | @10k         | @20k         |
| Beauty        | +3.8         | +4.5         | +6.1         | +6.5         | +5.0         | <b>+20.8</b> | <b>+18.1</b> | <b>+12.6</b> | <b>+16.8</b> | <b>+21.3</b> |
| Apparel       | -4.8         | -0.5         | +2.6         | +4.8         | +5.4         | <b>+1.4</b>  | <b>+2.9</b>  | <b>+9.3</b>  | <b>+15.5</b> | <b>+19.7</b> |
| Wireless      | <b>+64.6</b> | <b>+53.6</b> | <b>+22.1</b> | +25.3        | +22.5        | +58.4        | +51.5        | +21.5        | <b>+25.9</b> | <b>+27.7</b> |
| Drugstore     | +3.2         | +2.4         | -0.5         | -1.0         | -1.5         | <b>+12.4</b> | <b>+13.0</b> | <b>+9.2</b>  | <b>+12.6</b> | <b>+15.2</b> |
| Kitchen       | -2.4         | -2.4         | -2.5         | -3.1         | -2.0         | <b>+0.0</b>  | <b>+0.0</b>  | <b>+0.6</b>  | <b>+0.7</b>  | <b>+3.1</b>  |
| Home          | -2.8         | -2.0         | -1.6         | -2.0         | -0.5         | <b>+6.0</b>  | <b>+3.2</b>  | <b>+5.2</b>  | <b>+7.0</b>  | <b>+12.2</b> |
| PC            | -2.4         | -0.8         | +4.7         | +5.8         | +8.3         | <b>+6.0</b>  | <b>+5.4</b>  | <b>+6.2</b>  | <b>+9.1</b>  | <b>+13.6</b> |
| Grocery       | <b>+18.2</b> | <b>+17.8</b> | +13.2        | +14.0        | +11.9        | +11.6        | +14.3        | <b>+21.0</b> | <b>+28.6</b> | <b>+28.2</b> |
| Shoes         | -1.0         | -0.8         | +2.0         | +5.4         | +7.3         | <b>+17.0</b> | <b>+16.0</b> | <b>+12.8</b> | <b>+16.4</b> | <b>+21.2</b> |
| Book          | -3.6         | -3.2         | -1.5         | -1.7         | -3.5         | <b>+16.4</b> | <b>+15.1</b> | <b>+14.3</b> | <b>+18.2</b> | <b>+20.2</b> |
| Wireless Acc. | +2.2         | +3.6         | +4.9         | +5.7         | +8.1         | <b>+4.8</b>  | <b>+4.9</b>  | <b>+6.3</b>  | <b>+9.3</b>  | <b>+15.3</b> |
| Electronics   | -2.0         | -0.7         | +2.9         | +3.7         | +4.7         | <b>+3.4</b>  | <b>+2.7</b>  | <b>+7.1</b>  | <b>+8.4</b>  | <b>+14.0</b> |
| Home Improv.  | +5.0         | +3.4         | +6.3         | +8.1         | +6.0         | <b>+7.4</b>  | <b>+5.6</b>  | <b>+10.1</b> | <b>+15.4</b> | <b>+19.6</b> |
| Automotive    | <b>+2.0</b>  | <b>+4.9</b>  | <b>+7.5</b>  | <b>+8.6</b>  | +9.6         | +0.8         | +0.0         | +1.0         | +5.9         | <b>+15.9</b> |
| Baby Product  | <b>+30.2</b> | <b>+23.1</b> | <b>+24.3</b> | <b>+23.7</b> | <b>+16.4</b> | -0.8         | +0.6         | +4.9         | +8.8         | +12.2        |
| Sports        | -0.4         | -0.4         | -0.9         | +0.9         | +1.1         | <b>+3.0</b>  | <b>+3.5</b>  | <b>+7.8</b>  | <b>+12.6</b> | <b>+15.8</b> |
| Toys          | -0.8         | -0.7         | -3.3         | -4.7         | -3.5         | <b>+4.2</b>  | <b>+3.2</b>  | <b>+7.4</b>  | <b>+10.3</b> | <b>+15.8</b> |
| BISS          | <b>+1.0</b>  | <b>+2.5</b>  | <b>+4.6</b>  | +3.6         | +2.0         | -0.6         | -0.4         | +2.3         | <b>+4.3</b>  | <b>+8.6</b>  |
| Lawn & Garden | +0.4         | +1.4         | +6.1         | +7.5         | +6.7         | <b>+11.2</b> | <b>+9.1</b>  | <b>+19.1</b> | <b>+24.6</b> | <b>+25.9</b> |
| Personal Care | +0.2         | +0.1         | -2.8         | -4.2         | -5.7         | <b>+2.2</b>  | <b>+1.2</b>  | <b>+1.9</b>  | <b>+3.3</b>  | <b>+8.1</b>  |

Table 10: Sequence compression ablation. Results show absolute improvement over trivial baseline. Adaptive patching enables fitting full user histories within fixed context budgets.

| Context Strategy                 | Global       |             |             | Beauty       |             |             | Apparel      |             |             | Electronics  |             |             | Kitchen      |             |             | Toys        |             |             |
|----------------------------------|--------------|-------------|-------------|--------------|-------------|-------------|--------------|-------------|-------------|--------------|-------------|-------------|--------------|-------------|-------------|-------------|-------------|-------------|
|                                  | R            | F1          | P           | R            | F1          | P           | R            | F1          | P           | R            | F1          | P           | R            | F1          | P           | R           | F1          | P           |
| 4k recent only                   | +13.6        | +3.5        | -0.1        | +14.9        | +3.3        | -1.3        | +13.2        | +4.5        | +1.0        | +11.3        | +2.9        | -0.6        | +11.0        | -0.7        | -4.1        | +7.0        | +2.0        | +0.2        |
| 8k recent only                   | +13.9        | +3.5        | -0.2        | +15.4        | +3.4        | -1.3        | +13.6        | +4.5        | +0.9        | +11.2        | +2.7        | -0.7        | +11.1        | -0.8        | -4.2        | +7.2        | +1.9        | +0.2        |
| <b>8k full history (patched)</b> | <b>+14.1</b> | <b>+3.4</b> | <b>-0.2</b> | <b>+15.5</b> | <b>+3.2</b> | <b>-1.5</b> | <b>+13.8</b> | <b>+4.5</b> | <b>+0.9</b> | <b>+11.5</b> | <b>+2.8</b> | <b>-0.7</b> | <b>+11.4</b> | <b>-0.8</b> | <b>-4.3</b> | <b>+7.3</b> | <b>+1.9</b> | <b>+0.1</b> |

Table 11: Per-campaign A/B-test results. Lift is the relative gain in purchase conversions of treatment over control. CI denotes the 95% credible interval;  $P(\text{lift}>0)$  is the posterior probability of positive lift under our Bayesian covariate-adjusted framework.

| Campaign      | Lift   | 95% CI           | $P(\text{lift}>0)$ |
|---------------|--------|------------------|--------------------|
| Personal Care | +0.23% | (-0.12%, +0.57%) | 0.90               |
| Toys          | +0.19% | (-0.06%, +0.43%) | 0.93               |
| Sports        | +0.37% | (+0.02%, +0.72%) | 0.98               |
| Average       | +0.26% | —                | —                  |

8×A100 GPUs. Generator fine-tuning takes ~48h; the propensity model takes ~3h.

**Hard-negative mining.** We draw 4 independent temperature samples per training example (recall plateaus beyond 4). For the propensity model we

mine from the *trained* mitigated generator and train against these alongside popular and random negatives.

**Evaluation.** For generative classification we report Recall, Precision, and F1 with soft matching via Cohere cosine similarity  $\geq 0.7$ , calibrated on a small set of manually judged boundary pairs. For propensity ranking we report Hit@ $k$  when ranking users for an interest cluster, aggregating leaf propensities to product-category granularity via the aggregation in section 4.2. All numbers are absolute percentage-point improvements over a trivial baseline that predicts a user’s 5 most frequently purchased categories.