Tencent | GPR

Unified Value Alignment for Generative Recommendation in Industrial Advertising

Xinxun Zhang [1*], Yuling Xiong [2*], Jiale Zhou [3*], Zhengkai Guo [2], Zhennan Pang [2], Junbang Huo [2], Jingwen Wang [2], Xuyang Sun [2], Enming Zhang [2], Jiaguang Jin [2], Changping Wang [2], Yi Li [2], Jun Zhang [2†], Xiao Yan [1], Jiawei Jiang [1†], Jie Jiang [2†],
[1] Wuhan University, China; [2] Tencent Inc., China; [3] Peking University, China
{xxzhangstu,jiawei.jiang,yanxiaosunny}@whu.edu.cn,
jlzhou25@stu.pku.edu.cn,{whitnyxiong,neoxzhang,zeus}@tencent.com

Abstract

Generative Recommendation (GR) reformulates recommendation as a next-token generation problem and has shown promise in industrial applications. However, extending GR to industrial advertising is non-trivial because the system must optimize not only user interest but also commercial value. Existing GR pipelines remain largely semantics-centric, making it difficult to align value signals across tokenization, decoding, and online serving. To address this issue, we propose UniVA, a **Uni**fied **V**alue **A**lignment framework for advertising recommendation. We first introduce a Commercial SID tokenizer that injects value-related attributes into SID construction, yielding value-discriminative item representations. We then develop a Generation-as-Ranking SID Decoder jointly optimized by supervised learning and eCPM-aware reinforcement learning, which fuses value scores into next-item SID generation to perform generation and ranking in one decoding process. Finally, we design a value-guided personalized beam search that reuses generation-as-ranking logits as online value guidance and applies a personalized trie tree to constrain decoding to request-valid SID paths. Experiments on the Tencent WeChat Channels advertising platform show that UniVA achieves a 37.04% improvement in offline Hit Rate@100 over the baseline and a 1.5% GMV lift in online A/B tests.

CCS Concepts

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

Keywords

Generative Recommendation, Advertising, Commercial Value Modeling, Semantic ID

1 Introduction

Driven by the generative modeling capability of large language models (LLMs)[1, 19, 31], recommendation systems are shifting from multi-stage deep-learning pipelines to end-to-end generative architectures [13, 14]. Generative Recommendation (GR) formulates recommendation as a next-token prediction task by assigning each item a discrete Semantic ID (SID) and training an autoregressive model to generate SID sequences conditioned on a user's interaction history [5, 23, 26, 29]. This formulation unifies user sequential modeling and item retrieval within a single framework, and has shown strong performance across search, e-commerce, and content recommendation [13, 21, 26].

Advertising recommendation is a core monetization task for internet platforms and is inherently multi-objective [12]. Unlike many other recommendation settings, it must jointly consider user satisfaction and monetization outcomes [16]. In practice, the system needs to balance user interest with advertiser-side signals such as bid, return on investment (ROI), and expected cost per mille (eCPM), so as to improve platform revenue without harming user experience [27, 28]. These requirements introduce a value-modeling problem that goes beyond semantic matching [22, 23, 26].

The success of GR across various industrial domains naturally raises the question of whether advertising can similarly benefit from this paradigm. Recent studies have shown encouraging progress in this direction. For example, GPR [26] introduces a unified tokenization scheme and model architecture for end-to-end generative advertising recommendation, while subsequent work further incorporates auction-aware objectives and business-side supervision [22, 23, 28]. However, existing approaches still optimize commercial objectives in a fragmented manner across different stages of the generative pipeline. Tokenization mainly preserves semantic similarity, autoregressive decoding remains dominated by generation likelihood, and online serving often depends on additional value ranking or filtering modules.

Commercial value signals are therefore inconsistently modeled throughout SID construction, generation, and online serving, resulting in a mismatch between semantic generation and business objectives. We refer to this problem as *value inconsistency* in generative advertising systems, which mainly manifests in the following three aspects:

❶ **Value-insensitive SID tokenization.** Existing RQ-based tokenization pipelines are primarily designed to preserve multimodal semantic similarity, but they do not explicitly model the commercial heterogeneity inherent in advertising [15, 18]. As a result, semantically similar advertisements may still correspond to substantially different monetization utilities, while being mapped into nearby SID paths. Such representations provide limited commercial discriminability within the token space itself, making downstream

---

[*] These authors contributed equally to this work
[†] Corresponding author

Conference acronym 'XX, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

generation semantically coherent but insufficiently aligned with business objectives.

❷ **Semantic-dominated SID decoding.** Existing GR methods mainly optimize commercial value at the training-objective level, while autoregressive SID decoding is still dominated by semantic likelihood and sequence consistency. This limitation is particularly critical in advertising recommendation because decoding directly determines which SID trajectories survive during token expansion. Once commercially promising paths are pruned at early decoding steps, subsequent objectives can no longer recover them. As a result, commercial value cannot directly influence autoregressive decoding decisions, causing commercially promising SID trajectories to be prematurely pruned during token expansion.

❸ **Value-unaware online serving.** Even with value-aware training, online serving is still largely dominated by semantic beam expansion and heuristic filtering. Existing beam search procedures do not explicitly incorporate commercial value during SID trajectory selection, causing commercially promising paths to be overlooked during online decoding. Moreover, expanding over the full SID space wastes substantial computation on invalid candidates that violate inventory constraints or targeting rules, while introducing additional online value modules further increases serving latency and system complexity.

Our Solution. The key insight of this work is that commercial value should not be treated as an auxiliary signal introduced after generation, but instead should be consistently embedded into SID construction, autoregressive decoding, and online serving within a unified generative framework. To this end, we propose **UniVA**, a **Uni**fied **V**alue **A**lignment framework for generative advertising recommendation, which propagates commercial value signals throughout tokenization, decoding, and online serving:

- **Commercial SID Tokenization.** UniVA disentangles semantic organization and commercial refinement within hierarchical SID construction. Upper SID levels preserve semantic locality through RQ-based partitioning, while the final SID level specializes in commercial discrimination through a classify-then-bin strategy that integrates structured commercial attributes and bid discretization into value-aware token assignment. This design preserves semantic organization while improving commercial coherence among SID paths.
- **Generation-as-Ranking SID Decoder.** UniVA introduces a dual-head Generation-as-Ranking SID Decoder that jointly models generative likelihood and commercial value. A generation head performs autoregressive SID prediction, while a value head estimates token-level commercial preferences within the same SID space. The two outputs are fused during decoding, enabling each autoregressive decision to jointly consider semantic relevance and monetization objectives. To achieve unified optimization, UniVA jointly trains the generation head with supervised learning and the value head with eCPM-aware reinforcement learning, allowing commercial value signals to directly shape SID generation trajectories during autoregressive decoding.
- **Value-Guided Personalized Beam Search.** UniVA further proposes a value-guided personalized beam search for online serving. A request-specific valid-path trie tree first constrains beam expansion to feasible SID trajectories that satisfy inventory constraints and targeting rules. During decoding, beam scores are directly computed from the fused generation-ranking logits, allowing commercial value signals to participate in online SID trajectory selection. This design substantially reduces invalid search space while enabling lightweight value-aware online serving without introducing additional value ranking modules.

We conduct a comprehensive evaluation of UniVA on the Tencent WeChat Channels advertising platform. Experimental results demonstrate UniVA achieves a 37.04% relative improvement in offline HR@100. Further value analyses show that UniVA also improves value-oriented task substantially, with ValueHR@100 and wNDCG@100 increasing by 37.01% and 26.20%, respectively, while the bid distribution variance and range within each SID path are reduced by about one order of magnitude. Finally, online A/B testing demonstrates a 1.5% GMV lift, confirming the practical value of unified value alignment in production.

- We formulate the value inconsistency problem in generative advertising systems and identify three key limitations spanning tokenization, autoregressive decoding, and online serving..
- We propose a commercial SID tokenizer that injects value-related attributes into hierarchical SID construction, improving commercial discriminability while preserving semantic organization.
- We propose a Generation-as-Ranking SID Decoder that unifies supervised learning and eCPM-aware reinforcement learning, enabling value signals to participate directly in SID generation.
- We propose a value-guided personalized beam search that combines a personalized trie tree for valid-path pruning and fused generation-ranking logits for value-aware online serving.

2 Preliminaries

Semantic ID. GR formulates recommendation as a sequence generation problem. Given a user u, context c, and historical item sequence $\mathbf{i}_{1:T} = (i_1, i_2, \ldots, i_T)$, the model directly predicts the next target instead of ranking over a candidate set. Each item is mapped to a discrete SID sequence $s_i = \Phi(i) = \{s_i^1, s_i^2, \ldots, s_i^L\}$ of length L, and recommendation is performed by autoregressively generating SID tokens. A widely used implementation is RQ, which discretizes an item embedding into hierarchical codebook indices. Given an item embedding $\mathbf{Z}_i$ with $\mathbf{r}_i^1 = \mathbf{Z}_i$, RQ selects $s_i^l = \arg\min_k \|\mathbf{r}_i^l - \mathbf{c}_k^l\|_2^2, \mathbf{r}_i^{l+1} = \mathbf{r}_i^l - \mathbf{c}_{s_i^l}^l$, where $C^l = \{\mathbf{c}_1^l, \mathbf{c}_2^l, \ldots, \mathbf{c}_K^l\}$ denotes the level-l codebook and $\mathbf{r}_i^l$ denotes the residual representation at level l.

Advertisement Attributes. Each advertisement item i is associated with heterogeneous features $x_i = (x_i^s, x_i^c)$. The semantic attributes are defined as $x_i^s = (x_i^{\text{text}}, x_i^{\text{img}}, x_i^{\text{video}})$, which denote the textual, image, and video signals of the advertisement. The commercial attributes are defined as $x_i^c = (x_i^o, x_i^r, x_i^{\text{ind}}, x_i^b)$, corresponding to optimization goal, ROI target, industry, and bid. We focus on these four attributes because they are most closely related to bidding and commercial value.

Objective. Given the historical interaction sequence $\mathbf{x}_{1:T}$, UniVA maps each ad into an SID sequence $s_i = \Phi(x_i)$ and predicts the next SID autoregressively as $p_\theta(s_{T+1} \mid \mathbf{x}_{1:T}, u, c) = \prod_{l=1}^{L} p_\theta(s_{T+1}^l \mid s_{T+1}^{<l}, \mathbf{x}_{1:T}, u, c)$. where u denotes the user and c denotes the request

context. Unlike generic GR, the advertising setting requires the model to optimize both user relevance and commercial return.

3 Methodology

The overall framework of UniVA is shown in Figure 1. First, a commercial SID tokenizer injects commercial attributes and bid information into the final SID layer, making the token space value-discriminative. Second, the Generation-as-Ranking SID Decoder combines generation scores and token-level value scores within the same decoding process. Supervised learning is used to learn stable SID generation, while eCPM-aware reinforcement learning introduces downstream value guidance. Third, a value-guided personalized beam search restricts online expansion to valid personalized branches and directly reuses the fused generation-ranking logits as the online scoring signal, improving serving efficiency while reducing invalid SID-path exploration. The following sections describe each component in detail.

3.1 Commercial SID Tokenization

Existing SID construction methods primarily encode semantic characteristics, which leaves the token space insufficiently discriminative for commercial value. To address this limitation, UniVA adopts a semantic-commercial hybrid SID structure with an explicit Commercial SID:

$$(s_i^1, \dots, s_i^{L-1}) = \Phi_{\text{sem}}(x_i^s), \qquad s_i^L = \Phi_{\text{com}}(x_i^c), \tag{1}$$

where Φ_{sem} reuses the RQ-Kmeans+ semantic tokenizer [26] to preserve the semantic organization of the upper SID levels, and Φ_{com} maps structured commercial attributes into a discrete value-aware token at the final level. In this way, UniVA preserves the coarse-to-fine semantic hierarchy of GR while explicitly injecting commercial discriminability into the SID path. The construction of Φ_{com} consists of two steps: attribute space compression, followed by composition-key construction with equi-frequency bid binning.

Attribute Space Compression. The raw commercial attribute space is too sparse to support reliable token construction directly. In particular, optimization goals and industry categories exhibit substantial long-tail behavior, and directly taking their Cartesian product would lead to severe vocabulary explosion and unstable fine-grained statistics. UniVA therefore first compresses the structured commercial attributes before constructing the final commercial token. Formally, we define

$$x_i^{o'} = \phi_O(x_i^o), \qquad x_i^{r'} = \phi_{\mathcal{R}}(x_i^r), \qquad x_i^{\text{ind}'} = \phi_{\mathcal{I}}(x_i^{\text{ind}}), \tag{2}$$

where ϕ_O, $\phi_{\mathcal{R}}$, and $\phi_{\mathcal{I}}$ denote the compression operators for optimization goal, ROI, and industry, respectively. For optimization goal, UniVA retains the values covering 99% of the data and clusters the remaining tail values according to the similarity of their bid distributions, resulting in 25 categories. For ROI, UniVA retains the values covering 99% of the data and merges the remaining tail into one fallback class, resulting in 8 categories. For industry, UniVA retains the top 9 first-level industries covering 75% of the data and merges the remaining tail into one fallback class, resulting in 10 categories. This step improves statistical robustness while keeping the downstream token vocabulary under control.

Value-Aware Discretization. After attribute compression, UniVA constructs a composition key for each advertisement from the compressed business attributes:

$$k_i = (x_i^{o'}, x_i^{r'}, x_i^{\text{ind}'}) \in \mathcal{K} \subseteq O \times \mathcal{R} \times \mathcal{I}. \tag{3}$$

Each key k_i represents a local commercial context in which ads share similar structured business conditions. For each composition key $k \in \mathcal{K}$, the associated bid set is

$$\mathcal{B}_k = \{x_i^b \mid k_i = k\}. \tag{4}$$

We then apply a classify-then-bin strategy: ads are first grouped by composition key, and the bids within each key are partitioned into n_k equi-frequency bins. The bin count is chosen according to the sample volume of key k and is bounded by $n_{\max}$ and $n_{\min}$. Under the vocabulary budget constraint $\sum_{k \in \mathcal{K}} n_k \leq V$, we select $\{n_k\}$ to maximize the weighted entropy of the discretized bid distribution:

$$H_k = -\sum_{j=1}^{n_k} p_j^{(k)} \log p_j^{(k)}, \quad H = \sum_{k \in \mathcal{K}} w_k H_k, \tag{5}$$

where $p_j^{(k)}$ is the sample proportion of the j-th bin under key k, and w_k is the sample proportion of key k in the samples. Weighted entropy encourages a more balanced sample allocation across bins and therefore yields more stable bid discretization. This design assigns finer bid resolution to dense commercial contexts while keeping sparse contexts compact and robust.

After the binning scheme is fixed, UniVA assigns a global commercial token ID to each key-bin pair and defines the final commercial SID as

$$s_i^L = \Phi_{\text{com}}(x_i^c) = \psi(k_i, x_i^b), \tag{6}$$

where $\psi(\cdot)$ maps the compressed key and bid value to the corresponding global bin ID. For unseen keys at inference time, UniVA falls back to global bid discretization. As a result, ads sharing the same full SID path become more consistent in both semantic content and commercial value, yielding SID paths with stronger semantic and commercial coherence.

3.2 Generation-as-Ranking SID Decoder

Commercial SID makes the token space value-discriminative, while UniVA further injects commercial value into SID decoding through a Generation-as-Ranking SID Decoder. Following GPR, UniVA adopts the same unified input schema and HSTU encoder backbone [25]. The input sequence contains four token groups: User Token (U), Organic Token (O), Environment Token (E), and Item Token (I), which respectively encode user attributes and preferences, users' organic-content behaviors, request context, and historical ad interactions. The encoder processes these tokens and produces user-conditioned hidden states denoted by $h = \text{Enc}(U, O, E, I)$.

Context-Conditioned SID Decoding. Conditioned on h, the decoder autoregressively generates the target SID. At decoding step t, the current SID hidden state first performs fully visible

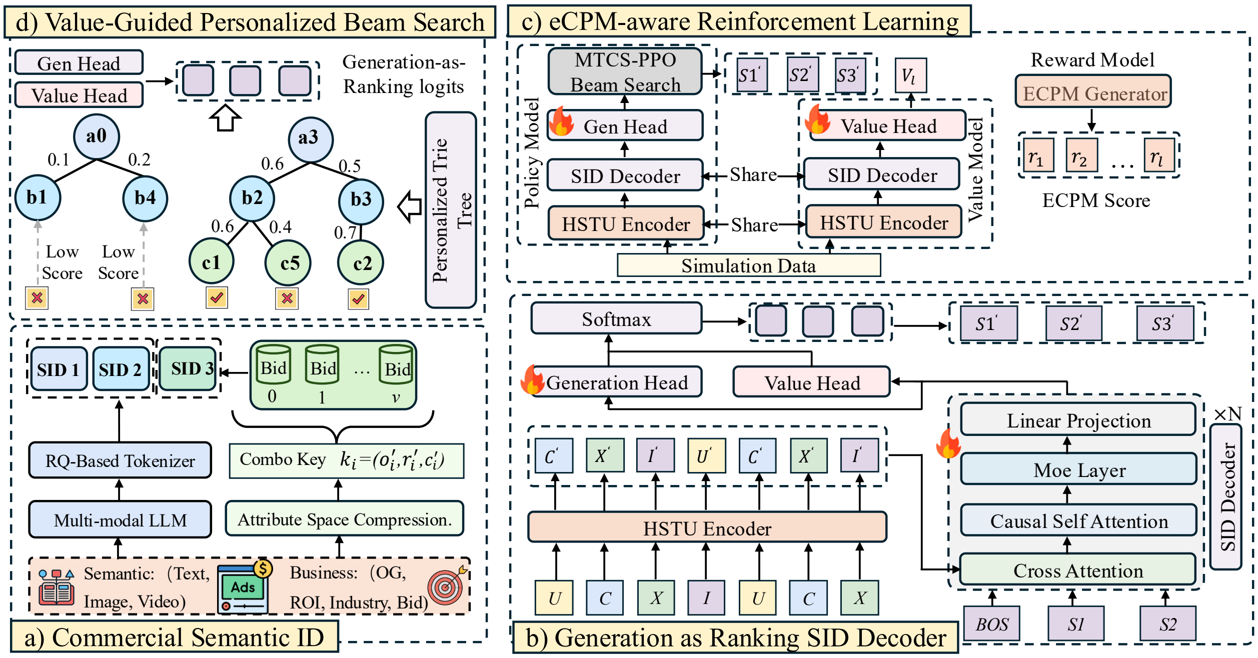

Figure 1: The framework of UniVA

cross-attention over the encoder states and then applies causal self-attention:

$$\tilde{z}^{(t)} = \mathrm{CrossAttn}(Q = z^{(t)}, K = h, V = h), \quad \hat{z}^{(t)} = \mathrm{SelfAttn}(\tilde{z}^{(t)}), \tag{7}$$

Cross-attention injects request-aware user context into the current decoding state, and self-attention organizes this refreshed state together with previously generated SID tokens. As a result, each next-token decision is grounded in both user intent and SID-prefix dependency. Here the SID hidden states serve as queries and the encoder output serves as keys and values.

Scalable SID Decoder. To further enhance decoder capacity, UniVA combines sparse Mixture-of-Experts (MoE) [3] and Mixture-of-Recursions (MoR) [2] for conditional specialization and recursive depth scaling. For MoE, UniVA uses N routed experts and activates only the top-K experts for each token, so that diverse advertising patterns can be decomposed and learned by different experts. The MoE transformation is formulated as

$$g(\hat{z}^{(t)}) = \mathrm{Softmax}(W_r \hat{z}^{(t)}), \tag{8}$$

$$z^{(t+1)} = E_0(\hat{z}^{(t)}) + \sum_{m \in \mathrm{TopK}(g(\hat{z}^{(t)}),K)} g_m(\hat{z}^{(t)}) E_m(\hat{z}^{(t)}). \tag{9}$$

Here W_r is the routing matrix, $g(\hat{z}^{(t)})$ is the routing distribution over the N routed experts, and $g_m(\hat{z}^{(t)})$ is the weight assigned to expert E_m. E_0 denotes the always-activated shared expert that captures common transformations, while the routed experts $\{E_m\}$ specialize in context-dependent advertising patterns such as different product categories, bidding strategies, and user-intent regimes. $\mathrm{TopK}(\cdot, K)$ selects the K most relevant routed experts for the current token, enabling conditional computation and fine-grained specialization.

To avoid router collapse, where a small subset of experts is repeatedly over-selected while others receive insufficient training, UniVA further adopts dynamic load balancing. Specifically, the router maintains historical expert-load statistics and adjusts the routing bias accordingly: overloaded experts receive lower bias, while underutilized experts receive higher bias before top-K selection. This encourages a more even distribution of training signals across experts, improves training stability, and reduces expert-capacity waste.

For depth scaling, we adopt Mixture-of-Recursions (MoR) [2] to recursively reuse a shared middle block:

$$h^{(0)} = \ell_{\mathrm{in}}(x), \qquad h^{(r)} = \ell_{\mathrm{mid}}(h^{(r-1)}), \qquad y = \ell_{\mathrm{out}}(h^{(R)}). \tag{10}$$

Here ℓ_{in}, ℓ_{mid}, and ℓ_{out} denote the input block, shared middle block, and output block, respectively, and R denotes the number of recursive rounds. By repeatedly applying the same middle transformation, MoR increases effective depth through iterative refinement without proportional parameter growth. Together, MoE expands width through conditional specialization and MoR increases effective depth, providing a stronger backbone for the SID decoder.

Dual-Head Generation-as-Ranking. On top of the shared decoder trunk, UniVA introduces two output heads to realize generation-as-ranking during SID decoding. Let $z^{(l)}$ denote the decoder hidden state at SID level l after the shared trunk. The generation head and value head first produce vocabulary-level generation scores and value scores, which are then fused and normalized to form the final generation-as-ranking distribution:

$$o_{\text{gen}}^{(l)} = f_{\text{gen}}(z^{(l)}), \quad o_{\text{value}}^{(l)} = f_{\text{value}}(z^{(l)}) \tag{11}$$

$$\tilde{\pi}_\theta(\cdot \mid s_{<l}, h) = \text{Softmax}(\text{Fuse}(o_{\text{gen}}^{(l)}, o_{\text{value}}^{(l)})) \tag{12}$$

Here $f_{\text{gen}}(\cdot)$ and $f_{\text{value}}(\cdot)$ denote the two output heads, and $o_{\text{gen}}^{(l)}$ and $o_{\text{value}}^{(l)}$ are their corresponding vocabulary-level outputs at SID level l. In our implementation, $\text{Fuse}(\cdot,\cdot)$ is instantiated as element-wise summation. This dual-head design makes generation and ranking happen within the same decoding process: the generation head preserves sequential SID generation ability, while the value head injects token-level commercial preference into each next-token decision. As a result, UniVA realizes true generation-as-ranking instead of following a generate-then-rerank pipeline, thereby avoiding the extra online cost introduced by an additional post-generation ranking stage.

SID Decoder Learning Objective. UniVA first uses supervised learning to establish stable SID generation behavior. The supervised objective is

$$\mathcal{L}_{\text{SL}} = -\sum_{(u,c,\mathbf{x}_{1:T},s_{T+1})\in\mathcal{D}_{\text{SL}}} \sum_{l=1}^{L} \log p_\theta(s_{T+1}^{l} \mid s_{T+1}^{<l}, h). \tag{13}$$

Here $\mathcal{D}_{\text{SL}}$ denotes the supervised training set and $p_\theta(\cdot \mid s_{T+1}^{<l}, h)$ denotes the next-SID generation distribution produced by the generation head. In this stage, the shared decoder trunk and generation head are optimized by the supervised objective. However, supervised learning alone does not provide explicit value supervision for token selection. UniVA therefore introduces the subsequent RL stage to incorporate downstream value signals into decoding.

3.3 eCPM-aware Reinforcement Learning

Supervised learning stabilizes SID generation but does not directly optimize commercial return. UniVA therefore introduces an eCPM-aware reinforcement learning stage. Concretely, the generation head learned in the SL stage is directly reused as the RL policy head, while the value head is optimized as the critic to estimate commercial value. The reward is produced by an online eCPM generator, and the decoder is later trained through iterative SL-RL optimization so that stable SID generation and commercial value alignment are learned within the same model.

Simulation-Based Value Optimization. Directly querying the production ranking service for each sampled SID path is prohibitively expensive. UniVA therefore follows the simulation-based post-training paradigm of GPR [26] and builds a high-fidelity offline simulator from recent production snapshots. The simulator reproduces candidate inventory, feature pipelines, business constraints, and the downstream ranking stack, providing scalable offline reward evaluation under serving-time constraints. RL training data are obtained by simulation sampling over recorded online requests, so the sampling strategy directly affects the quality ceiling of policy learning under limited simulation budget. UniVA therefore replaces the previous fixed 5% sampling with adaptive sampling up to full traffic based on historical learning difficulty and prediction entropy, and further augments replayed requests with simulated future requests derived from the user's latest state.

During RL, the encoder first produces the context state h, and the policy head defines the token-generation policy $\tilde{\pi}_\theta(\cdot \mid s_{<l}, h)$. Trajectories are collected by beam search and MCTS-PPO [10], where MCTS-PPO additionally reuses the value head as a node evaluator:

$$\begin{aligned}\mathcal{Y}(h) &= \mathcal{Y}_{\text{beam}}\big(\tilde{\pi}_\theta(\cdot \mid h)\big) \\ &\quad \cup\ \mathcal{Y}_{\text{mcts-ppo}}\big(\tilde{\pi}_\theta(\cdot \mid h), V_\theta(\cdot \mid h)\big) = \{y^{(1)}, \dots, y^{(K)}\},\end{aligned} \tag{14}$$

where $y^{(k)} = (a_1^{(k)}, \dots, a_L^{(k)})$ is a complete SID path and $V_\theta(\cdot \mid h)$ denotes the value estimates for intermediate prefixes. Beam search provides stable high-probability rollouts, while MCTS-PPO performs structured exploration over SID prefixes and helps discover promising but low-probability paths. For a search node n with action set $\mathcal{A}(n)$, MCTS-PPO selects actions by

$$a^\star = \arg\max_{a\in\mathcal{A}(n)} \left(\bar{Q}(n,a) + c\sqrt{\frac{\log N(n)}{1+N(n,a)}} \right), \tag{15}$$

where $\bar{Q}(n,a)$, $N(n)$, and $N(n,a)$ denote the running action value, node visit count, and edge visit count, respectively. Online serving, by contrast, uses only beam search for efficiency. Each sampled path is then resolved to a concrete ad and assigned an estimated eCPM reward:

$$R_{\text{eCPM}}^{(k)} = g_{\text{eCPM}}\big(h, y^{(k)}\big), \tag{16}$$

where $g_{\text{eCPM}}(\cdot)$ is produced by the copied production pCTR/pCVR models. This keeps the RL target consistent with the serving-time monetization objective. UniVA further normalizes rewards within each request to reduce scale variation across traffic contexts:

$$\bar{R}^{(k)} = \frac{R_{\text{eCPM}}^{(k)} - \mu_R(h)}{\sigma_R(h) + \epsilon_r}. \tag{17}$$

This normalization makes the policy updates depend more on relative value differences among sampled candidates under the same request.

Advantage Estimation and Losses. For a sampled SID path $y = (a_1, a_2, \dots, a_L)$ with normalized terminal reward $\bar{R}$, UniVA applies PPO-style GAE to obtain token-level advantages A_l. Let a_l denote the selected token at level l, then

$$v_l = o_{\text{value}}^{(l)}[a_l], \qquad \hat{G}_l = A_l + v_l. \tag{18}$$

Here A_l is defined relative to the current value baseline v_l, so $\hat{G}_l$ is the corresponding return target. In this way, the value head learns to predict token-level future value under the current decoding policy.

Let $\tilde{\pi}_{\text{ref}}$ denote a lagged reference policy with the same architecture as the current model and periodically synchronized parameters. The PPO ratio and clipped objective are

$$\rho_l = \frac{\tilde{\pi}_\theta(a_l \mid s_{<l}, h)}{\tilde{\pi}_{\text{ref}}(a_l \mid s_{<l}, h)}, \mathcal{L}_{\text{PPO}} = -\mathbb{E}\left[\min\left(\rho_l A_l, \text{clip}(\rho_l, 1-\epsilon, 1+\epsilon) A_l\right)\right]. \tag{19}$$

The value head is optimized by

$$\mathcal{L}_{\text{value}} = \mathbb{E}\left[\left(v_l - \hat{G}_l\right)^2\right]. \tag{20}$$

The overall RL objective is

$$\mathcal{L}_{\text{RL}} = \mathcal{L}_{\text{PPO}} + \lambda_v \mathcal{L}_{\text{value}}, \tag{21}$$

where λ_v balances policy optimization and value regression. This objective aligns SID decoding with high-eCPM paths while learning token-level value estimates.

3.4 Joint Optimization

Supervised learning is used to establish stable SID generation, while reinforcement learning introduces commercial value supervision through simulator-generated reranking rewards. UniVA combines the two stages through collaborative iterative training so that SID generation and value estimation are aligned within the same decoder. Concretely, for a supervised batch UniVA updates the shared decoder and generation head with $\mathcal{L}_{\text{SL}}$; for an RL batch it updates the fused generation-ranking policy and value head with $\mathcal{L}_{\text{RL}}$. The overall training objective is

$$\mathcal{L}_{\text{train}} = \mathbb{I}_{\text{SL}}\mathcal{L}_{\text{SL}} + \mathbb{I}_{\text{RL}}, \tag{22}$$

where $\mathbb{I}_{\text{SL}}$ and $\mathbb{I}_{\text{RL}}$ indicate whether the current batch is a supervised batch or an RL batch. By alternating SL and RL batches, UniVA progressively aligns SID generation and value estimation within the same decoder, so that decoding is guided toward commercially valuable paths. This produces a closed loop between training-time generation-as-ranking and serving-time.

3.5 Value-Guided Personalized Beam Search

UniVA keeps online serving consistent with the generation-as-ranking design under the unified value-alignment objective. The central principle of this stage is to let commercial value directly participate in beam expansion, so that online SID trajectory selection is governed by joint relevance and monetization preference rather than by semantic likelihood alone. To make such value-guided decoding both valid and efficient under request-time constraints, UniVA further introduces a personalized valid-path trie that restricts beam expansion to feasible SID trajectories.

A global valid-path trie tree is first constructed over feasible SID paths in the candidate inventory. For an incoming request, serving constraints such as targeting, availability, and creative rules are applied to derive a personalized subtree $\mathcal{T}_u$:

$$\mathcal{T}_u = \Gamma(u)(\mathcal{T}). \tag{23}$$

Given a SID prefix $s_{<l}$, the valid next-token set is defined as

$$\mathcal{V}(s_{<l};\mathcal{T}_u) = \{s_l \in \mathcal{S}_l \mid s_{\le l} = (s_{<l}, s_l) \in \mathcal{P}(\mathcal{T}_u)\}, \tag{24}$$

where $\mathcal{S}_l$ denotes the SID vocabulary at level l and $\mathcal{P}(\mathcal{T}_u)$ denotes the set of valid prefixes in the personalized trie. Conditioned on the user state, UniVA performs beam search only over valid continuations in $\mathcal{V}(s_{<l};\mathcal{T}_u)$, which substantially reduces invalid-path expansion and concentrates decoding budget on request-valid candidates.

Within this constrained search space, UniVA further injects value guidance into online SID trajectory selection. At decoding step l, the dual-head decoder produces generation scores and value scores for valid candidate tokens, and their fused output is directly used as the beam-expansion signal. The cumulative beam score of a valid SID prefix is then defined as

$$\text{Score}(s_{\le l}) = \sum_{t=1}^{l} \text{Fuse}(o_{\text{gen}}^{(t)}, o_{\text{value}}^{(t)})[s_t], \qquad \text{s.t. } s_{\le l} \in \mathcal{P}(\mathcal{T}_u). \tag{25}$$

By directly reusing the fused generation-ranking logits, UniVA allows commercial value signals to participate in beam competition throughout online decoding, rather than relying on semantic likelihood alone. As a result, beam search preserves SID prefixes favored jointly by user relevance and monetization value, reducing the risk that commercially promising trajectories are pruned at early decoding steps. The personalized trie and value-guided beam scoring therefore play complementary roles: the former enforces request-specific validity and reduces the invalid search space, while the latter enables lightweight value-aware online serving without introducing an additional value-ranking module. Consequently, online serving preserves a single-pass generation-as-ranking process and remains fully aligned with the end-to-end value objective.

4 Experiments

4.1 Experimental Setup

Datasets and Baselines. Following GPR [26], we build the offline dataset from a large-scale Tencent advertising corpus that mixes ads with organic media such as short videos, social feeds, and news, so that evaluation reflects realistic mixed-traffic contexts. Each sample contains session-level behaviors together with item-level multimodal features, including textual metadata and visual signals from covers or sampled frames. We remove near-duplicate instances, rebalance category distributions to reduce sampling bias, and split the corpus into 80% training and 20% testing. To focus on SID-level effectiveness, we use GPR as the main system baseline and a vanilla decoder-only Transformer as the decoder baseline, and then progressively introduce Commercial SID together with different SID-decoder designs to compare their contributions.

Implementation Details. UniVA uses a three-level SID structure with a codebook size of 2048. The SID decoder contains 4 layers with embedding dimension 256. For Commercial SID construction, the adaptive bid-binning hyperparameters are set to $n_{\max} = 25$ and $n_{\min} = 3$, and the final binning configuration is selected by grid search to maximize the weighted entropy H under the vocabulary budget of 2048. For the sparse MoE backbone, UniVA uses 64 total experts and activates the top 16 experts for each token; the hidden dimension of each expert is set to 128. In training, UniVA optimizes the model with Adam using a learning rate of 0.001, and the batch size is 16. The input sequence length is 2048.

Evaluation Metrics. For offline evaluation, we report Hit Rate at K (HR@K). To assess value modeling on the GMV-weighted next-conversion set, we further report two value-oriented metrics:

$$\text{ValueHR@K} = \frac{\sum_{t=1}^{T} \text{gmv}_{i_t} \cdot \mathbb{I}(i_t \in R_t^K)}{\sum_{t=1}^{T} \text{gmv}_{i_t}},$$

$$\text{wNDCG@K} = \frac{\sum_{t=1}^{T} w_t \cdot \text{NDCG}_t\text{@}K}{\sum_{t=1}^{T} w_t}, \qquad w_t = \log_{10}(1 + \text{gmv}_{i_t}).$$

where T is the number of evaluation requests, i_t is the ground-truth converted item of request t, gmv_{i_t} is its GMV, and R_t^K is the model's Top-K candidate set. ValueHR@K measures how much conversion value is covered by Top-K retrieval, while wNDCG@K measures value-weighted ranking quality by emphasizing whether high-value requests are ranked earlier. For online evaluation, we report GMV

Table 1: Offline next interacted item prediction performance comparison of UniVA and its SID-level variants on the industrial advertising benchmark. Parameters and FLOPs report only the SID decoder cost, excluding the encoder. ΔHR@100 denotes the relative improvement over the base GPR + SID Decoder.

Model	Parameters	FLOPs	ΔHR@100
Base			
GPR+SID Decoder	3M	4.1G	+ 0.0%
SID Design			
+ Commercial SID	3M	4.1G	+5.78%
+ (layer2-layer4)	7M	7.1G	+6.10%
+ MOR	5M	7.1G	+13.56%
+ Sparse MOE	60M	8.5G	+18.40%
UniVA (Full)	80M	23.2G	+37.04%

and GMV(normal). GMV is the standard gross merchandise volume metric and directly reflects business return, while GMV(normal) excludes ROI ads.

4.2 Overall Performance

As summarized in Table 1, UniVA consistently improves next interacted item prediction performance performance over the base GPR + SID Decoder, and the reported parameters and FLOPs include only the SID decoder. Commercial SID improves HR@100 by 5.78% without increasing decoder parameters or computation. This indicates that Commercial SID introduces a stronger value-structured bias into SID construction: the grouping criterion moves from pure content homogeneity to joint content-value homogeneity, making each SID path more commercially coherent and giving the model a clearer learning signal.

Table 1 also reveals a scaling trend for the SID decoder. As decoder capacity increases, the relative HR@100 gain rises from 6.10% for the deeper decoder to 13.56% for MoR and further to 18.40% for Sparse MoE. This suggests that SID decoding in industrial advertising benefits directly from stronger modeling capacity: MoR improves effective depth through recursive refinement, while Sparse MoE increases conditional capacity through expert specialization. Together, these results show that SID decoding exhibits clear scaling behavior under more expressive decoder backbones.

Full UniVA achieves the largest gain, reaching a 37.04% relative improvement over the base model. Beyond decoder scaling, this improvement comes from eCPM-aware reinforcement learning and joint optimization, which expose the decoder to downstream value signals instead of only supervised next-SID targets. By aligning token selection with monetization-aware rewards, UniVA better preserves high-value prefixes and completes more commercially favorable SID paths. These results show that Commercial SID, decoder scaling, and value-aware joint optimization are complementary.

4.3 Value Alignment Performance

To verify whether UniVA can capture commercial value, we conduct a value analysis experiment under different SID designs on the GMV-weighted next-conversion set. Figure 2 presents the value-oriented evaluation under different SID designs on the GMV-weighted next-conversion set. 2*2048 SID + CSID performs best on most cutoffs, achieving the highest ValueHR@10/32/50 and wNDCG@32/50. At K=100, it reaches 0.0677 on ValueHR@K and 0.0554 on wNDCG@K, clearly outperforming the two three-level SID variants, while the semantic-only SID settings remain slightly stronger only at the most restrictive positions such as wNDCG@1 and wNDCG@10.

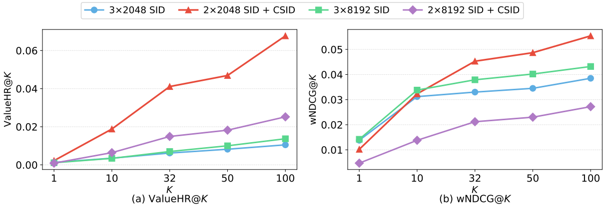

Figure 2: Offline value analysis under four SID designs on the GMV-weighted next-conversion set.

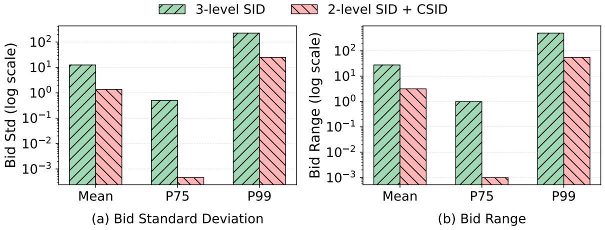

Figure 3: Path-level bid-dispersion statistics for 3-level SID and 2-level SID + Commercial SID. Each subfigure reports Mean, P75, and P99 over complete SID paths, and the y-axis is shown in log scale.

This pattern is consistent with the role of Commercial SID. The three-level SID settings mainly preserve semantic similarity, which is helpful at very small cutoffs, but they do not explicitly separate high-value and low-value ads in the token space. After introducing Commercial SID, value-consistent ads are more likely to share coherent SID paths, enabling the decoder to retrieve and rank high-value candidates more effectively. The comparison between 2*2048 SID + CSID and 2*8192 SID + CSID further suggests that a moderate codebook is more suitable for value modeling, because an overly large vocabulary disperses data and weakens stable commercial grouping. Overall, these results show that UniVA improves not only prediction accuracy but also value capture.

4.4 More Insights

Commercial SID Quality Analysis. Figure 3 compares path-level bid dispersion before and after introducing Commercial SID. Relative to 3-level SID, 2-level SID + CSID consistently reduces both bid standard deviation and bid range across Mean, P75, and P99. Under the log-scale view, most statistics decrease by about one order of magnitude, with especially clear compression in the middle and tail. This indicates that items assigned to the same full SID path become much more consistent in commercial value, rather than mixing ads with widely different bid levels under the same semantic

Table 2: HR@K comparison under different SID codebook sizes. All values are reported in percentage points (%).

SID Configuration	HR@1	HR@10	HR@32	HR@50	HR@100
3*2048 SID	0.09	0.72	1.60	2.15	3.23
3*8192 SID	0.10	0.83	2.06	2.77	4.03
2*2048 SID + CSID	**0.14**	**1.02**	**2.17**	**2.84**	**4.20**
2*8192 SID + CSID	0.09	0.92	1.98	2.63	3.84

path. Consequently, Commercial SID provides a cleaner structural basis for value-aware decoding and reduces unstable high-variance paths.

Commercial SID Strategy Analysis. Figure 4 compares different Commercial SID construction strategies. Classify-then-Bin combined with Equal-frequency achieves the highest weighted entropy while keeping the vocabulary size closest to the target budget of 2048, with $H = 7.487$ and $V = 1939$. Direct Binning ignores structured commercial attributes and therefore mixes heterogeneous ads before bid discretization, leading to coarse and less balanced partitions. Cluster-then-Bin improves bid-distribution grouping, but the resulting clusters are less stable and often trade vocabulary efficiency for limited entropy gain. Among in-bin strategies, Equal-width is sensitive to long-tail bid distributions, while Clustering tends to consume more vocabulary without consistent benefit. Overall, the results support the choice of Classify-then-Bin with Equal-frequency as the most balanced strategy for Commercial SID construction.

Codebook Size Analysis. Table 2 presents HR@K under different semantic codebook sizes and Commercial SID settings. 2*2048 SID + CSID performs best across all cutoffs from HR@1 to HR@100, with relative improvements of 55.56%, 41.67%, 35.63%, 32.09%, and 30.03% respectively. Under the 8192 setting, however, 2*8192 SID + CSID remains below 3*8192 SID across all cutoffs. This difference is consistent with the codebook design: the Commercial SID vocabulary is fixed at 2048, so it matches the 2048 semantic setting more naturally, while replacing one semantic level under 8192 introduces a stronger vocabulary mismatch and reduces the benefit of finer semantic partitioning. Overall, these results show that Commercial SID is most effective when paired with a moderate semantic codebook, where explicit value modeling complements semantic structure without introducing excessive semantic fragmentation.

4.5 Online A/B Test

Personalized Beam Search. Under the same beam width of 300, personalized trie-based beam search produces 300 valid SID paths, whereas beam search without the trie produces only 48. In other words, without the personalized trie tree, beam search recovers only 16% of the valid paths obtained with trie-constrained search. This shows that the trie tree can filter invalid branches before expansion, so beam capacity is concentrated on feasible paths instead of being wasted on invalid ones. As a result, UniVA saves online search resources while generating substantially more valid ad creatives under the same decoding budget.

Table 3: Online A/B test results on Tencent WeChat Channels advertising traffic from March 7 to March 11, 2026, over 5% traffic. All lifts denote relative improvement against the production baseline.

Online Version	GMV Lift	GMV(normal) Lift
v1 w/o Generation-as-Ranking	+1.03%	+1.17%
v2 with Generation-as-Ranking	**+1.50%**	**+1.42%**

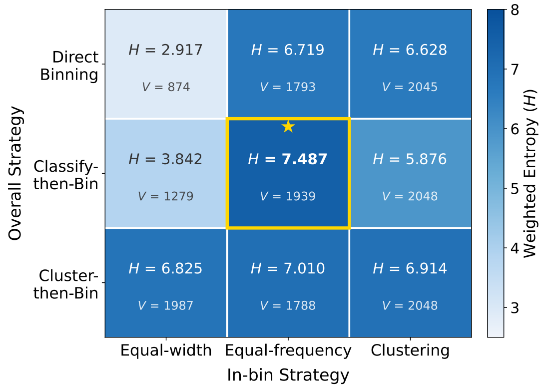

Figure 4: Commercial SID strategy comparison across three overall strategies and three in-bin strategies. Each cell corresponds to one strategy combination and reports weighted entropy H and vocabulary size V.

Online GMV Results. Table 3 presents the online A/B results on Tencent WeChat Channels advertising traffic. Even without generation-as-ranking, the online system already achieves positive gains, with GMV lift of 1.03% and GMV(normal) lift of 1.17%. After adopting generation-as-ranking, the gains further improve to 1.50% and 1.42%, respectively. The first online version already verifies that Commercial SID together with the SID decoder can improve business value in production, even before introducing the full generation-as-ranking design. This indicates that value-aware SID construction and value-enhanced SID decoding are already helpful for retrieving more commercially suitable ads. After further introducing generation-as-ranking, candidate generation and value guidance are unified within the same decoding process, so the SID decoder can use value signals earlier and more consistently during online search. Overall, the stronger GMV gains show that joint optimization on the SID decoder leads to better value capture and transfers effectively to real monetization improvement.

5 Related Work.

Generative Recommendation. GR reformulates recommendation as autoregressive generation over discrete item identifiers, thereby replacing hand-crafted multi-stage ranking stacks with a unified next-token prediction paradigm [4, 18, 24, 25, 29]. Recent advances show that this line is evolving quickly from proof-of-concept models to industrial-scale systems, with active exploration of one-model training, large-catalog generation, decoding acceleration, and scalable serving architectures [9, 21, 30]. The same trend has also reached advertising, where generative pipelines are increasingly studied as an alternative to conventional cascaded systems [23, 26, 28]. These efforts mainly establish GR as a viable system paradigm. In contrast, UniVA focuses on a more advertising-specific question: how to make commercial value participate consistently throughout the GR pipeline rather than improving generation quality, model scale, or decoding speed alone.

Semantic ID. Semantic ID is a key enabler of large-scale generative recommendation because it converts items in massive catalogs into compact discrete token sequences. Early work learns vector-quantized item identifiers from semantic embeddings [6, 7], while later studies improve the SID pipeline through learnable tokenization, large-catalog training, and parallel generation of long semantic-ID sequences [8, 17, 20]. More recent work further moves toward end-to-end SID generation in industrial advertising environments [11]. Although these studies substantially improve semantic fidelity, token efficiency, and generation scalability, they are still primarily semantics-driven: the SID space is designed to preserve item similarity or facilitate generation, rather than to explicitly expose monetization differences.

Value Modeling in Recommendation. Value-aware recommendation has long recognized that commercial systems should optimize monetization-related utility in addition to user relevance, leading to research on profit-aware learning, multi-objective recommendation, and value alignment in industrial scenarios [15, 16]. In advertising, recent generative systems have also started to incorporate auction signals, eCPM-oriented supervision, and business-aware objectives into model training and serving [22, 23, 26–28]. However, in most prior work, value remains a downstream target, an auxiliary objective, or a later-stage ranking signal applied after candidate generation. UniVA differs by treating value as a pipeline-wide modeling principle.

6 Conclusion

In this paper, we first identify the core challenge of value modeling for generative recommendation in industrial advertising: the model must capture not only semantic relevance but also commercial value, while keeping tokenization, decoding, and online serving aligned under real serving constraints. To address this issue, we propose UniVA, a unified value alignment framework that introduces Commercial SID to build a value-discriminative token space, a Co-Generation-Ranking SID Decoder to inject value directly into SID generation through joint supervised and eCPM-aware reinforcement learning, and a Value-Guided Personalized Beam Search to combine personalized valid-path filtering with single-pass generation-as-ranking during online serving. Experimental results show that UniVA consistently improves offline downstream prediction performance, strengthens value modeling under value-oriented metrics, produces more commercially coherent SID distributions, and improves serving-time validity and efficiency through personalized beam search, while online A/B testing further confirms that these offline gains transfer to real business value in production.

References

[1] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774* (2023).

[2] Sangmin Bae, Yujin Kim, Reza Bayat, Sungnyun Kim, Jiyoun Ha, Tal Schuster, Adam Fisch, Hrayr Harutyunyan, Ziwei Ji, Aaron Courville, et al. 2025. Mixture-of-recursions: Learning dynamic recursive depths for adaptive token-level computation. *arXiv preprint arXiv:2507.10524* (2025).

[3] Damai Dai, Chengqi Deng, Chenggang Zhao, RX Xu, Huazuo Gao, Deli Chen, Jiashi Li, Wangding Zeng, Xingkai Yu, Yu Wu, et al. 2024. Deepseekmoe: Towards ultimate expert specialization in mixture-of-experts language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 1280–1297.

[4] Ruidong Han, Bin Yin, Shangyu Chen, He Jiang, Fei Jiang, Xiang Li, Chi Ma, Mincong Huang, Xiaoguang Li, Chunzhen Jing, et al. 2025. Mtgr: Industrial-scale generative recommendation framework in meituan. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management*. 5731–5738.

[5] Xuegang Hao, Ming Zhang, Alex Li, Xiangyu Qian, Zhi Ma, Yanlong Zang, Shijie Yang, Zhongxuan Han, Xiaolong Ma, Jinguang Liu, et al. 2025. OxygenREC: An Instruction-Following Generative Framework for E-commerce Recommendation. *arXiv preprint arXiv:2512.22386* (2025).

[6] Yupeng Hou, Zhankui He, Julian McAuley, and Wayne Xin Zhao. 2023. Learning vector-quantized item representation for transferable sequential recommenders. In *Proceedings of the ACM Web Conference 2023*. 1162–1171.

[7] Yupeng Hou, Zhankui He, Julian McAuley, and Wayne Xin Zhao. 2023. Learning vector-quantized item representation for transferable sequential recommenders. In *Proceedings of the ACM Web Conference 2023*. 1162–1171.

[8] Yupeng Hou, Jiacheng Li, Ashley Shin, Jinsung Jeon, Abhishek Santhanam, Wei Shao, Kaveh Hassani, Ning Yao, and Julian McAuley. 2025. Generating long semantic ids in parallel for recommendation. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*. 956–966.

[9] Yanhua Huang, Yuqi Chen, Xiong Cao, Rui Yang, Mingliang Qi, Yinghao Zhu, Qingchang Han, Yaowei Liu, Zhaoyu Liu, Xuefeng Yao, et al. 2025. Towards Large-scale Generative Ranking. *arXiv preprint arXiv:2505.04180* (2025).

[10] Jie Jiang, Yangru Huang, Zeyu Wang, Changping Wang, Yuling Xiong, Jun Zhang, and Huan Yu. 2026. Spend Search Where It Pays: Value-Guided Structured Sampling and Optimization for Generative Recommendation. *arXiv preprint arXiv:2602.10699* (2026).

[11] Jie Jiang, Xinxun Zhang, Enming Zhang, Yuling Xiong, Jun Zhang, Jingwen Wang, Huan Yu, Yuxiang Wang, Hao Wang, Xiao Yan, et al. 2026. End-to-End Semantic ID Generation for Generative Advertisement Recommendation. *arXiv preprint arXiv:2602.10445* (2026).

[12] Yuchen Jiang, Qi Li, Han Zhu, Jinbei Yu, Jin Li, Ziru Xu, Huihui Dong, and Bo Zheng. 2022. Adaptive domain interest network for multi-domain recommendation. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. 3212–3221.

[13] Xiaopeng Li, Bo Chen, Junda She, Shiteng Cao, You Wang, Qinlin Jia, Haiying He, Zheli Zhou, Zhao Liu, Ji Liu, et al. 2025. A survey of generative recommendation from a tri-decoupled perspective: Tokenization, architecture, and optimization. (2025).

[14] Yongqi Li, Xinyu Lin, Wenjie Wang, Fuli Feng, Liang Pang, Wenjie Li, Liqiang Nie, Xiangnan He, and Tat-Seng Chua. 2024. A survey of generative search and recommendation in the era of large language models. *arXiv preprint arXiv:2404.16924* (2024).

[15] Xinchen Luo, Jiangxia Cao, Tianyu Sun, Jinkai Yu, Rui Huang, Wei Yuan, Hezheng Lin, Yichen Zheng, Shiyao Wang, Qigen Hu, et al. 2025. Qarm: Quantitative alignment multi-modal recommendation at kuaishou. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management*. 5915–5922.

[16] Changhua Pei, Xinru Yang, Qing Cui, Xiao Lin, Fei Sun, Peng Jiang, Wenwu Ou, and Yongfeng Zhang. 2019. Value-aware Recommendation based on Reinforcement Profit Maximization. In *The World Wide Web Conference (WWW '19)*. Association for Computing Machinery, New York, NY, USA, 3123–3129.

[17] Aleksandr V Petrov and Craig Macdonald. 2024. RecJPQ: training large-catalogue sequential recommenders. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*. 538–547.

[18] Shashank Rajput, Nikhil Mehta, Anima Singh, Raghunandan Hulikal Keshavan, Trung Vu, Lukasz Heldt, Lichan Hong, Yi Tay, Vinh Tran, Jonah Samost, et al. 2023. Recommender systems with generative retrieval. *Advances in Neural Information Processing Systems* 36 (2023), 10299–10315.

[19] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288* (2023).

[20] Wenjie Wang, Honghui Bao, Xinyu Lin, Jizhi Zhang, Yongqi Li, Fuli Feng, See-Kiong Ng, and Tat-Seng Chua. 2024. Learnable item tokenization for generative recommendation. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*. 2400–2409.

[21] Yejing Wang, Shengyu Zhou, Jinyu Lu, Ziwei Liu, Langming Liu, Maolin Wang, Wenlin Zhang, Feng Li, Wenbo Su, Pengjie Wang, et al. 2025. NEZHA: A Zero-sacrifice and Hyperspeed Decoding Architecture for Generative Recommendations. *arXiv preprint arXiv:2511.18793* (2025).

[22] Lee Xiong, Zhirong Chen, Rahul Mayuranath, Shangran Qiu, Arda Ozdemir, Lu Li, Yang Hu, Dave Li, Jingtao Ren, Howard Cheng, et al. 2026. LLaTTE: Scaling Laws for Multi-Stage Sequence Modeling in Large-Scale Ads Recommendation. *arXiv preprint arXiv:2601.20083* (2026).

[23] Ben Xue, Dan Liu, Lixiang Wang, Mingjie Sun, Peng Wang, Pengfei Zhang, Shaoyun Shi, Tianyu Xu, Yunhao Sha, Zhiqiang Liu, et al. 2026. Generative Recommendation for Large-Scale Advertising. *arXiv preprint arXiv:2602.22732* (2026).

[24] Chao Yi, Dian Chen, Gaoyang Guo, Jiakai Tang, Jian Wu, Jing Yu, Mao Zhang, Sunhao Dai, Wen Chen, Wenjun Yang, et al. 2025. Recgpt technical report. *arXiv preprint arXiv:2507.22879* (2025).

[25] Jiaqi Zhai, Lucy Liao, Xing Liu, Yueming Wang, Rui Li, Xuan Cao, Leon Gao, Zhaojie Gong, Fangda Gu, Jiayuan He, et al. 2024. Actions Speak Louder than Words: Trillion-Parameter Sequential Transducers for Generative Recommendations. In *International Conference on Machine Learning*. PMLR, 58484–58509.

[26] Jun Zhang, Yi Li, Yue Liu, Changping Wang, Yuan Wang, Yuling Xiong, Xun Liu, Haiyang Wu, Qian Li, Enming Zhang, et al. 2025. GPR: Towards a Generative Pre-trained One-Model Paradigm for Large-Scale Advertising Recommendation. *arXiv preprint arXiv:2511.10138* (2025).

[27] Chujie Zhao, Qun Hu, Shiping Song, Dagui Chen, Han Zhu, Jian Xu, and Bo Zheng. 2025. LLM-Auction: Generative Auction towards LLM-Native Advertising. *arXiv preprint arXiv:2512.10551* (2025).

[28] Zuowu Zheng, Ze Wang, Fan Yang, Jiangke Fan, Teng Zhang, Yongkang Wang, and Xingxing Wang. 2025. Ega-v2: An end-to-end generative framework for industrial advertising. *arXiv preprint arXiv:2505.17549* (2025).

[29] Guorui Zhou, Jiaxin Deng, Jinghao Zhang, Kuo Cai, Lejian Ren, Qiang Luo, Qianqian Wang, Qigen Hu, Rui Huang, Shiyao Wang, et al. 2025. OneRec Technical Report. *arXiv preprint arXiv:2506.13695* (2025).

[30] Guorui Zhou, Hengrui Hu, Hongtao Cheng, Huanjie Wang, Jiaxin Deng, Jinghao Zhang, Kuo Cai, Lejian Ren, Lu Ren, Liao Yu, et al. 2025. Onerec-v2 technical report. *arXiv preprint arXiv:2508.20900* (2025).

[31] Hao Zhou, Chengming Hu, Ye Yuan, Yufei Cui, Yili Jin, Can Chen, Haolun Wu, Dun Yuan, Li Jiang, Di Wu, et al. 2024. Large language model (llm) for telecommunications: A comprehensive survey on principles, key techniques, and opportunities. *IEEE Communications Surveys & Tutorials* 27, 3 (2024), 1955–2005.