Appendix

Anonymous Submission

Calculation of Cumulative Explained Variance

To quantify the uniformity of the representation space, we compute the cumulative explained variance ratio based on the standard PCA procedure. Let $\mathbf{E} \in \mathbb{R}^{N \times d}$ denote the item embedding matrix, where N is the number of items and d is the hidden dimension.

First, we perform Standardization (Zero-mean, Unit-variance) on the embedding matrix to ensure scale invariance. For each feature dimension j, the standardized element $\tilde{E}_{i,j}$ is calculated as:

$$\tilde{E}i, j = \frac{Ei, j - \mu_j}{\sigma_j} \tag{1}$$

where μ_j and σ_j are the mean and standard deviation of the j-th dimension across all items, respectively.

Next, we apply Principal Component Analysis (PCA) to the standardized matrix $\tilde{\mathbf{E}}$. This involves computing the covariance matrix $\mathbf{C} = \frac{1}{N-1}\tilde{\mathbf{E}}^\top \tilde{\mathbf{E}}$ and performing eigendecomposition to obtain the eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$, which represent the variance explained by each principal component.

The Cumulative Explained Variance Ratio for the top-K principal components is defined as the proportion of the total variance captured by the first K components:

$$\mathcal{V}_K = \frac{\sum i = 1^K \lambda_i}{\sum_{j=1}^{d} \lambda_j} \tag{2}$$

A lower value of $\mathcal{V}_K$ (for a small K) indicates that the information is distributed across more dimensions, reflecting a higher degree of isotropy in the representation space. In our experiments, we set the maximum $K = 16$ to observe the trend in the dominant spectral directions.

Theoretical Justification of Consensus Steering

In this section, we provide a rigorous theoretical grounding for the Collective Consensus Steering (CCS) module. We frame the sequential recommendation task as a high-dimensional signal recovery problem and demonstrate that our subspace rectification mechanism functions as a **Bayesian Shrinkage Estimator**, which is theoretically necessary to minimize estimation error under stochastic noise conditions.

The High-Variance Dilemma in Isolation Modeling

Let $\mathbf{h}_u^* \in \mathbb{R}^d$ denote the *true, latent intent* of user u. In real-world scenarios, the observed sequence representation $\mathbf{z}_u$ is often corrupted by stochastic perturbations (e.g., accidental clicks, context ambiguity). We model this observation process as:

$$\mathbf{z}_u = \mathbf{h}_u^* + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_d) \tag{3}$$

where $\boldsymbol{\epsilon}$ represents isotropic Gaussian noise. Traditional methods that rely on **instance-level isolation** effectively perform Maximum Likelihood Estimation (MLE) based solely on $\mathbf{z}_u$. The expected Mean Squared Error (MSE) of this estimator scales linearly with the embedding dimension d:

$$\mathbb{E}[\|\mathbf{z}_u - \mathbf{h}_u^*\|^2] = \text{Tr}(\text{Cov}(\boldsymbol{\epsilon})) = d \cdot \sigma^2 \tag{4}$$

In high-dimensional latent spaces (e.g., $d = 64$), this variance term ($d\sigma^2$) dominates the error, rendering the model highly sensitive to idiosyncratic noise.

Bayesian Shrinkage via Low-Rank Prior

To mitigate this high variance, PRISM introduces an **informative prior**: we postulate that valid user intents do not span the entire $\mathbb{R}^d$ space but reside within a compact, low-rank manifold defined by the Global Consensus Matrix $\mathbf{C} \in \mathbb{R}^{K \times d}$.

We specifically calibrate the cardinality of this basis set to $K = 16$, representing the **intrinsic dimensionality** of the collective intent space. This constraint ($K \ll d$) is crucial for constructing a robust shrinkage estimator. The rectification operation projects the noisy observation onto this prior subspace:

$$\hat{\mathbf{h}}_{CCS} = \mathbf{z}_u(\mathbf{C}^\top \mathbf{C}) = \mathbf{P}_\mathbf{C} \mathbf{z}_u \tag{5}$$

where $\mathbf{P}_\mathbf{C}$ is the projection matrix of rank $K = 16$. In the context of Stein's estimation theory Efron and Morris [1973], this operation acts as a **shrinkage operator** that contracts the high-variance individual estimate toward the stable population consensus.

Recall that in the main paper, we enforce the Consensus Factor Orthogonality loss ($\mathcal{L}_{orth}$) and normalize the factors. Therefore, in this theoretical derivation, we assume the consensus matrix C satisfies the orthonormality condition $CC^\top = I_K$ Chen *et al.* [2022]. This ensures that $P_C = C^\top C$ is a valid orthogonal projection matrix.

Derivation of Bias-Variance Decomposition

To rigorously justify the efficacy of this design, we perform a Bias-Variance decomposition on the estimation error. The MSE of our rectified estimator is derived as follows:

$$
\begin{aligned}
\text{MSE}(\hat{\mathbf{h}}_{CCS}) &= \mathbb{E}_{\boldsymbol{\epsilon}}\left[\|\hat{\mathbf{h}}_{CCS} - \mathbf{h}_u^*\|^2\right] \\
&= \mathbb{E}_{\boldsymbol{\epsilon}}\left[\|\mathbf{P}_{\mathbf{C}}(\mathbf{h}_u^* + \boldsymbol{\epsilon}) - \mathbf{h}_u^*\|^2\right] \\
&= \mathbb{E}_{\boldsymbol{\epsilon}}\left[\|(\mathbf{P}_{\mathbf{C}} - \mathbf{I})\mathbf{h}_u^* + \mathbf{P}_{\mathbf{C}}\boldsymbol{\epsilon}\|^2\right] \\
&= \underbrace{\|(\mathbf{I} - \mathbf{P}_{\mathbf{C}})\mathbf{h}_u^*\|^2}_{\text{Bias Term (Deterministic)}} + \underbrace{\mathbb{E}_{\boldsymbol{\epsilon}}[\|\mathbf{P}_{\mathbf{C}}\boldsymbol{\epsilon}\|^2]}_{\text{Variance Term (Stochastic)}} \\
&\quad + \underbrace{2(\mathbf{h}_u^*)^\top(\mathbf{P}_{\mathbf{C}} - \mathbf{I})^\top\mathbf{P}_{\mathbf{C}}\mathbb{E}[\boldsymbol{\epsilon}]}_{\text{Cross Term}=0}
\end{aligned}
\tag{6}
$$

Since the noise $\boldsymbol{\epsilon}$ is zero-mean ($\mathbb{E}[\boldsymbol{\epsilon}] = \mathbf{0}$), the cross term vanishes. The variance term can be further simplified using the trace operator properties:

$$
\begin{aligned}
\text{Variance} &= \mathbb{E}_{\boldsymbol{\epsilon}}\left[\text{Tr}(\boldsymbol{\epsilon}^\top\mathbf{P}_{\mathbf{C}}^\top\mathbf{P}_{\mathbf{C}}\boldsymbol{\epsilon})\right] \\
&= \text{Tr}\left(\mathbf{P}_{\mathbf{C}}\mathbb{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top]\mathbf{P}_{\mathbf{C}}^\top\right) \\
&= \text{Tr}\left(\mathbf{P}_{\mathbf{C}}(\sigma^2\mathbf{I})\mathbf{P}_{\mathbf{C}}^\top\right) \\
&= \sigma^2\text{Tr}(\mathbf{P}_{\mathbf{C}}) = K \cdot \sigma^2
\end{aligned}
\tag{7}
$$

This derivation leads to two critical insights:

- **Variance Reduction:** By projecting onto the subspace of rank $K = 16$, the estimation variance is drastically reduced from $d\sigma^2$ to $K\sigma^2$. Since $K \ll d$ (i.e., $16 \ll 64$), the stochastic noise is effectively filtered out in the orthogonal $d - K$ dimensions.
- **Bias Control via Alignment:** The shrinkage introduces a bias term $\|(\mathbf{I} - \mathbf{P}_{\mathbf{C}})\mathbf{h}_u^*\|^2$, representing the information loss due to projection. This theoretically justifies our **Global Alignment Loss** ($\mathcal{L}_{align}$): explicitly minimizing $\mathcal{L}_{align}$ forces the consensus subspace $\mathbf{C}$ to cover the high-density regions of true user intents $\mathbf{h}_u^*$, thereby minimizing this bias term.

The Consensus Steering module functions as an optimal trade-off mechanism. By constraining the latent structure to a low-rank basis ($K = 16$), PRISM sacrifices a negligible amount of expressivity (bias) to achieve a substantial reduction in variance, theoretically guaranteeing robust intent estimation in noisy sequential environments.

Experimental Setup

Dataset Statisitcs

Table 1 summarizes the statistics of the dataset used for the experiment, The descriptions of each dataset are below:

- **Amazon**[1] (`Beauty`, `Sports`, `Toys`) McAuley *et al.* [2015]: This datasets are all derived from a global e-commerce platform, Amazon, which contains data on a range of product reviews from 1996 to 2014. These datasets are widely used for SR task.

[1]http://jmcauley.ucsd.edu/data/amazon/links.html

Dataset	# Users	# Items	# Interactions	Avg. Length	Sparsity
`Beauty`	22,363	12,101	198,502	8.9	99.93%
`Sports`	35,598	18,357	296,337	8.3	99.95%
`Toys`	19,412	11,924	167,597	8.6	99.93%
`Yelp`	30,431	20,033	316,354	10.4	99.95%
`LastFM`	1,090	3,646	52,551	48.2	98.68%
`ML-1M`	6,041	3,417	999,611	165.5	95.16%

Table 1: Detailed descriptions of the three baseline datasets.

- `Yelp`[2]: This dataset is known as a popular business recommendation dataset. We only treat the transaction records after January 1st, 2019 since it is very large.
- `ML-1M`[3]Harper and Konstan [2015]: This is the popular movie recommendation dataset provided by MovieLens[4]. It has the longest average interaction length among our datasets.
- `LastFM`[5]: This dataset contains user interaction with music, such as artist listening records. It is used to recommend musicians to users in SR with long sequence lengths.

Details of Baselines

In our experiments, we compare fourteen SR baselines, divided into two categories. Here's the description of baselines in experiments:

- RNN or Transformer-based sequential models: GRU4Rec Hidasi *et al.* [2015] pioneers the adaptation of Gated Recurrent Units (GRUs) for sequential modeling in session-based recommendation. Caser Tang and Wang [2018] employs Convolutional Neural Networks (CNNs) to capture local patterns in short-term behavior sequences for recommendation enhancement. SASRec Kang and McAuley [2018] first uses self-attention to capture user interests. BERT4Rec Sun *et al.* [2019] introduces a deep bidirectional self-attention mechanism to model user behavior sequences, with a cloze task objective for masked prediction training. DuoRec Qiu *et al.* [2022] combines contrastive learning to boost self-attention-based recommendation. ICLRec Chen *et al.* [2022] learns latent intent distributions via intent-aware sequence clustering and optimizes representations through contrastive self-supervised learning. CL4SRec Xie *et al.* [2022] integrates contrastive learning with Transformer architecture to construct more accurate user behavior representations. MAERec Ye *et al.* [2023] proposes a graph masking autoencoder that adaptively captures global item-interaction topology for self-supervised augmentation. SASRec_{F} proposed by Lin *et al.* [2024], are advanced versions of SASRec Kang and McAuley [2018], which integrate item IDs and property features via concatenation. MSSR Lin *et al.* [2024] adopts a

[2]https://www.yelp.com/dataset

[3]https://grouplens.org/datasets/movielens/1m/

[4]https://grouplens.org/datasets/movielens/

[5]https://grouplens.org/datasets/hetrec-2011/

Hyper parameter	α	β	B	LR
Beauty	0.5	0.5	512	1e-3
Sports	0.7	0.7	256	5e-4
Toys	0.7	0.3	256	1e-3
Yelp	0.7	0.5	256	5e-4
LastFM	0.7	0.5	256	1e-3
ML-1M	0.3	0.5	256	1e-3

Table 2: Optimal hyper-parameter settings of PRISM on all datasets.

Datasets	CDs & Vinyl	Automotive	Grocery & Gourmet Food
# Sequences	112,379	193,651	127,496
# Items	15,520	18,703	11,778
# Interactions	457,589	806,939	623,940

Table 3: More detailed statements of the cross-domain datasets in Amazon.

multi-sequence integrated attention layer strategy to model user interests.

- Frequency-based or Transformer and Frequency-based sequential models: FMLPRec Zhou *et al.* [2022] proposes a filter-enhanced MLP designed to filter out possible noise effects in SR tasks. FEARec Du *et al.* [2023] proposes a frequency-enhanced hybrid attention network to alleviate the deficiency of self-attention in capturing high-frequency information. BSARec Shin *et al.* [2024] leverages the DFT to balance between the inductive bias and self-attention, DIFF Kim *et al.* [2025] proposed a noise filtering mechanism and a multi-sequence side-information fusion strategy.

Hyper-parameter Setting

The optimal parameters of PRISM on the three datasets are shown in Table 2, and the grid search ranges for each parameter are as follows: the learner weight α and β are in $\{0.1, 0.3, 0.5, 0.7, 0.9\}$, batch size B is in $\{32, 64, 128, 256, 512\}$ and learning rate LR is in $\{1e-3, 5e-4, 1e-4, 5e-5\}$

More Details on Cross-domain Scenario

Datasets Statistics.

Table 3 shows the statistical details of the datasets in the noisy scenario experiment. Following the prior works Li *et al.* [2023]; Yang *et al.* [2024]; Liu *et al.* [2025], we choose three public real-world benchmarks from Amazon [6] (i.e,. CDs & Vinyl, Automotive, Grocery & Gourmet Food). For our experiments, we utilized the common 5-core versions of these datasets, filtering out users and items with fewer than five interactions. The datasets were structured into sequential format, with the maximum historical interaction sequence length capped at ten.

[6]https://cseweb.ucsd.edu/ jmcauley/datasets/amazon_v2/

Model	Automotive		CDs		Grocery	
	H@10	N@10	H@10	N@10	H@10	N@10
FMLPRec	0.0320	0.0204	0.0367	0.0207	0.1043	0.0705
BSARec	0.0514	0.0430	0.0537	0.0438	0.1414	0.1215
Our Model$_{w/o\ Filter}$	0.0490	0.0453	0.0512	0.0478	0.1326	0.1211
Our Model$_{w/o\ Steer}$	0.0485	0.0454	0.0514	0.0481	0.1327	0.1230
Our Model	**0.0542**	**0.0458**	**0.0561**	**0.0489**	**0.1430**	**0.1246**

Table 4: Performance under noisy scenario in terms of HR@10 and NDCG@10 across the three target domains. The best results are marked in **bold**.

Methods		SASRec	DuoRec	FEARec	BSARec	Ours
Beauty	Train Time	13.588	34.423	248.335	15.657	16.745
	Inference	13.569	14.409	19.711	13.404	13.401
	Params (MiB)	3.349	3.349	3.349	3.350	3.372
	FLOPs (MB)	9.911	29.734	29.491	9.911	9.955
Sports	Train Time	18.998	50.918	372.549	25.722	26.753
	Inference	17.281	23.762	29.121	22.872	23.115
	Params (MiB)	4.876	4.876	4.876	4.877	4.899
	FLOPs (MB)	9.911	29.734	29.491	10.001	10.014

Table 5: Model efficiency comparison on Beauty and Sports datasets. Training/inference time is reported in seconds per epoch (s/epoch).

Performance of Handling Noisy Data.

Table 4 presents the recommendation performance comparison of PRISM against baselines under cross-domain settings, measured by HR@5 and NDCG@5. We aggregated training sets from three categories while maintaining independent self-category test and validation sets. Three key observations emerge: (i) PRISM establishes a new state-of-the-art across all target domains, outperforming BSARec, by a significant margin. (i) Notably, on the CDs segment, PRISM attains an NDCG@10 of 0.0489 versus 0.0438 for BSARec, yielding a relative gain of 11.64%. (iii) Crucially, removing either the adaptive spectral filter or the consensus steer causes a clear drop, confirming that the two modules jointly suppress domain-specific high-frequency noise and extract universal preference signals.

Efficiency Analysis

To evaluate PRISM's computational complexity and efficiency, we conduct comparative analyses of training time (s/epoch), inference time (s/epoch), parameters (MB), and FLOPs (MB) against baseline models across two benchmark datasets, with all experiments rigorously executed on identical NVIDIA RTX 4090 GPUs (24GB) to ensure fairness, as detailed in Table 5. Overall, PRISM exhibits a modest parameter increase while demonstrating faster training speeds than DuoRec and FEARec, though marginally slower ($\leq 1.06\times$) than BSARec. Crucially, PRISM achieves optimal inference

Dataset	Beauty				Sports & Outdoors				Toys & Games			
Metric	R@5	N@5	R@10	N@10	R@5	N@5	R@10	N@10	R@5	N@5	R@10	N@10
P5-SID [2023]	0.0465	0.0329	0.0638	0.0384	0.0295	0.0212	0.0403	0.0247	0.0216	0.0151	0.0325	0.0186
P5-CID [2023]	0.0465	0.0325	0.0668	0.0391	0.0295	0.0214	0.0420	0.0254	0.0223	0.0143	0.0357	0.0186
P5-SemID [2023]	0.0459	0.0327	0.0667	0.0394	0.0336	0.0243	0.0481	0.0290	0.0264	0.0178	0.0416	0.0227
TIGER [2023]	0.0352	0.0236	0.0533	0.0294	0.0176	0.0143	0.0311	0.0146	0.0274	0.0174	0.0438	0.0227
ID-GenRec [2024]	0.0463	0.0328	0.0665	0.0393	0.0273	0.0186	0.0403	0.0228	0.0462	0.0323	0.0651	0.0383
LETTER [2024a]	0.0364	0.0243	0.0560	0.0306	0.0209	0.0136	0.0331	0.0176	0.0309	0.0296	0.0493	0.0262
ELMRec [2024b]	0.0372	0.0267	0.0506	0.0310	0.0241	0.0181	0.0307	0.0203	0.0148	0.0119	0.0193	0.0131
LC-Rec [2024]	0.0503	0.0352	0.0715	0.0420	0.0259	0.0175	0.0384	0.0216	0.0543	0.0385	0.0753	0.0453
GRAM [2025]	<u>0.0641</u>	<u>0.0451</u>	<u>0.0890</u>	<u>0.0531</u>	<u>0.0375</u>	<u>0.0256</u>	<u>0.0554</u>	<u>0.0314</u>	<u>0.0718</u>	<u>0.0516</u>	<u>0.0987</u>	<u>0.0603</u>
PRISM	**0.0729**	**0.0518**	**0.0998**	**0.0605**	**0.0422**	**0.0295**	**0.0606**	**0.0354**	**0.0798**	**0.0571**	**0.1083**	**0.0662**
Improve.	13.73%	14.86%	12.13%	13.94%	12.53%	15.23%	9.39%	12.74%	11.14%	10.66%	9.73%	9.78%

Table 6: Overall performance comparison with generative recommendation models.

latency across all datasets. This empirically confirms that while the integration of frequency-domain MLPs incurs slight parameter and training overheads, PRISM significantly optimizes inference efficiency without compromising computational footprint.

Comparison with Generative Models

Baselines.

In this section, we conduct comparative experiments against state-of-the-art generative recommendation models. Specifically, nine advanced generative recommenders are selected for benchmarking:

- P5-SID, P5-CID, and P5-SemID Hua *et al.* [2023] assign numerical IDs to items using sequential indexing, spectral clustering, and item metadata, respectively.
- TIGER Rajput *et al.* [2023] employs a hierarchical RQ-VAE-based method to generate item IDs.
- ID-GenRec Tan *et al.* [2024] learns an ID generation model leveraging meta-information to produce item IDs.
- LETTER Wang *et al.* [2024a] integrates hierarchical semantics, collaborative signals, and code assignment diversity, and proposes a ranking-guided generation loss to enhance recommendation performance.
- ELMRec Wang *et al.* [2024b] incorporates user-item high-order interactions and adopts a numerical ID-based re-ranking strategy.
- LC-Rec Zheng *et al.* [2024] utilizes RQ-VAE IDs and further integrates linguistic and collaborative semantics through finetuning tasks.
- GRAM Lee *et al.* [2025] designs a semantics-to-lexicon translation mechanism and integrates it with multi-granular late fusion techniques.

Results.

Table 6 presents the performance comparison between PRISM and generative recommendation models across three datasets. For brevity, R denotes Recall and N denotes NDCG. The best results are highlighted in **bold** and suboptimal results are <u>underlined</u>. *Improve.* represents the percentage improvement of PRISM over the strongest baseline. It can be observed that PRISM achieves the best overall recommendation performance on all three datasets, with improvements ranging from 9.39% to 15.23%. Specifically, on the `Beauty` dataset, PRISM attains R@5/N@5 scores of 0.0729/0.0518. This represents an improvement of 13.73% in R@5 and 14.86% in N@5 relative to GRAM (0.0641/0.0451). These results provide strong empirical evidence for the effectiveness of dynamic spectral modulation and consensus learning in SR tasks.

References

Yongjun Chen, Zhiwei Liu, Jia Li, Julian McAuley, and Caiming Xiong. Intent contrastive learning for sequential recommendation. In *Proceedings of the ACM web conference 2022*, pages 2172–2182, 2022.

Xinyu Du, Huanhuan Yuan, Pengpeng Zhao, Jianfeng Qu, Fuzhen Zhuang, Guanfeng Liu, Yanchi Liu, and Victor S Sheng. Frequency enhanced hybrid attention network for sequential recommendation. In *Proceedings of the 46th International ACM SIGIR conference on research and development in information retrieval*, pages 78–88, 2023.

Bradley Efron and Carl Morris. Stein's estimation rule and its competitors—an empirical bayes approach. *Journal of the American Statistical Association*, 68(341):117–130, 1973.

F. Maxwell Harper and Joseph A. Konstan. The movielens datasets: History and context. *ACM Trans. Interact. Intell. Syst.*, 5(4), December 2015.

Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. Session-based recommendations with recurrent neural networks. *CoRR*, abs/1511.06939, 2015.

Wenyue Hua, Shuyuan Xu, Yingqiang Ge, and Yongfeng Zhang. How to index item ids for recommendation foundation models. In *Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region*, pages 195–204, 2023.

Wang-Cheng Kang and Julian McAuley. Self-attentive sequential recommendation. In *2018 IEEE international conference on data mining (ICDM)*, pages 197–206. IEEE, 2018.

Hyeyoung Kim, Minjin Choi, Sunkyung Lee, and Ilwoong Baek Jongwuk Lee. Diff: Dual side-information filtering and fusion for sequential recommendation. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2025.

Sunkyung Lee, Minjin Choi, Eunseong Choi, Hye-young Kim, and Jongwuk Lee. GRAM: Generative recommendation via semantic-aware multi-granular late fusion. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers*, pages 33294–33312. Association for Computational Linguistics, 2025.

Jiacheng Li, Ming Wang, Jin Li, Jinmiao Fu, Xin Shen, Jingbo Shang, and Julian McAuley. Text is all you need: Learning language representations for sequential recommendation. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1258–1267, 2023.

Xiaolin Lin, Jinwei Luo, Junwei Pan, Weike Pan, Zhong Ming, Xun Liu, Shudong Huang, and Jie Jiang. Multi-sequence attentive user representation learning for side-information integrated sequential recommendation. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, pages 414–423, 2024.

Shuo Liu, An Zhang, Guoqing Hu, Hong Qian, and Tat-Seng Chua. Preference diffusion for recommendation. In *The Thirteenth International Conference on Learning Representations*, 2025.

Julian McAuley, Christopher Targett, Qinfeng Shi, and Anton van den Hengel. Image-based recommendations on styles and substitutes. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '15, page 43–52, New York, NY, USA, 2015. Association for Computing Machinery.

Ruihong Qiu, Zi Huang, Hongzhi Yin, and Zijian Wang. Contrastive learning for representation degeneration problem in sequential recommendation. In *Proceedings of the fifteenth ACM international conference on web search and data mining*, pages 813–823, 2022.

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

Yehjin Shin, Jeongwhan Choi, Hyowon Wi, and Noseong Park. An attentive inductive bias for sequential recommendation beyond the self-attention. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 8984–8992, 2024.

Fei Sun, Jun Liu, Jian Wu, Changhua Pei, Xiao Lin, Wenwu Ou, and Peng Jiang. Bert4rec: Sequential recommendation with bidirectional encoder representations from transformer. In *Proceedings of the 28th ACM international conference on information and knowledge management*, pages 1441–1450, 2019.

Juntao Tan, Shuyuan Xu, Wenyue Hua, Yingqiang Ge, Zelong Li, and Yongfeng Zhang. Idgenrec: Llm-recsys alignment with textual id learning. In *Proceedings of the 47th international ACM SIGIR conference on research and development in information retrieval*, pages 355–364, 2024.

Jiaxi Tang and Ke Wang. Personalized top-n sequential recommendation via convolutional sequence embedding. In *Proceedings of the eleventh ACM international conference on web search and data mining*, pages 565–573, 2018.

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

Xinfeng Wang, Jin Cui, Fumiyo Fukumoto, and Yoshimi Suzuki. Enhancing high-order interaction awareness in llm-based recommender model. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 11696–11711, 2024.

Xu Xie, Fei Sun, Zhaoyang Liu, Shiwen Wu, Jinyang Gao, Jiandong Zhang, Bolin Ding, and Bin Cui. Contrastive learning for sequential recommendation. In *2022 IEEE 38th international conference on data engineering (ICDE)*, pages 1259–1273. IEEE, 2022.

Zhengyi Yang, Jiancan Wu, Zhicai Wang, Xiang Wang, Yancheng Yuan, and Xiangnan He. Generate what you prefer: Reshaping sequential recommendation via guided diffusion. *Advances in Neural Information Processing Systems*, 36, 2024.

Yaowen Ye, Lianghao Xia, and Chao Huang. Graph masked autoencoder for sequential recommendation. In *Proceedings of the 46th international ACM SIGIR conference on research and development in information retrieval*, pages 321–330, 2023.

Bowen Zheng, Yupeng Hou, Hongyu Lu, Yu Chen, Wayne Xin Zhao, Ming Chen, and Ji-Rong Wen. Adapting large language models by integrating collaborative semantics for recommendation. In *2024 IEEE 40th International Conference on Data Engineering (ICDE)*, pages 1435–1448. IEEE, 2024.

Kun Zhou, Hui Yu, Wayne Xin Zhao, and Ji-Rong Wen. Filter-enhanced mlp is all you need for sequential recommendation. In *Proceedings of the ACM web conference 2022*, pages 2388–2399, 2022.