

Article

Not peer-reviewed version

---

# Intelligent Recommendation Systems Using Multi-Scale LoRA Fine-Tuning and Large Language Models

---

[Huajun Zhang](#), Lin Zhu, Chong Peng, Jiasen Zheng, Junjiang Lin, Runyuan Bao \*

Posted Date: 17 December 2025

doi: 10.20944/preprints202512.1413.v1

Keywords: large language model; multi-scale fine-tuning; LoRA; recommendation system



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Intelligent Recommendation Systems Using Multi-Scale LoRA Fine-Tuning and Large Language Models

Huajun Zhang <sup>1</sup>, Lin Zhu <sup>2</sup>, Chong Peng <sup>3</sup>, Jiasen Zheng <sup>4</sup>, Junjiang Lin <sup>5</sup> and Runyuan Bao <sup>6,\*</sup>

<sup>1</sup> Syracuse University, Syracuse, USA

<sup>2</sup> Stevens Institute of Technology, New Jersey, USA

<sup>3</sup> Carnegie Mellon University, Pittsburgh, USA

<sup>4</sup> Northwestern University, Evanston, USA

<sup>5</sup> University of Toronto, Toronto, Canada

<sup>6</sup> Johns Hopkins University, Baltimore, USA

\* Correspondence: rbao3@jhu.edu

## Abstract

This study proposes a multi-scale LoRA fine-tuning recommendation algorithm based on large language models to address the limitations of traditional recommender systems in semantic understanding, feature redundancy, and parameter transfer efficiency. The method preserves the semantic representation ability of large models while achieving unified modeling of global preferences and local interests through multi-scale semantic decomposition, low-rank parameter adaptation, and cross-scale fusion mechanisms. The model first inputs user-content interaction sequences into a pre-trained language model to obtain context-aware semantic embeddings. Then, a multi-scale semantic pooling structure extracts hierarchical feature information to capture multi-granularity preference relations. Based on this, a multi-scale LoRA module performs low-rank decomposition and cross-scale alignment of weight matrices, significantly reducing parameter size and improving fine-tuning efficiency. Finally, a cross-scale attention fusion layer dynamically reconstructs global and local features to optimize recommendation ranking. Systematic experiments conducted on the MovieLens-1M dataset validate the effectiveness of the proposed method across multiple evaluation metrics. The results show that the model outperforms several baseline algorithms in Precision@K, NDCG@K, Recall@K, and Coverage, demonstrating the advantages of multi-scale structure and LoRA parameterization in enhancing recommendation accuracy, diversity, and generalization. Overall, this research provides a feasible solution for structural optimization and parameter-efficient fine-tuning of large language models in efficient recommendation tasks.

**Keywords:** large language model; multi-scale fine-tuning; LoRA; recommendation system

---

## I. Introduction

In recent years, recommender systems have become a vital component of intelligent information services, playing a central role in e-commerce, short video platforms, social media, and personalized learning[1]. However, with the rapid growth of user behavior data and the increasing complexity of content forms, traditional recommendation algorithms have shown limitations in capturing multidimensional semantic associations and dynamic preference shifts. Although models based on collaborative filtering or deep feature extraction have achieved progress in specific tasks, they still struggle with cross-domain knowledge transfer, contextual understanding, and cold-start problems. Especially in complex semantic scenarios, user interests often exhibit multi-scale, cross-modal, and hierarchical characteristics[2]. A single vectorized representation cannot fully express such semantic structures and layered relationships. Therefore, efficiently mining potential preferences and

achieving scalable personalized recommendations from large-scale heterogeneous data have become key challenges in the field of intelligent recommendation.

With the rapid development of large language models (LLMs), their strong capability in semantic understanding, generation, and reasoning has brought a new paradigm to recommender systems[3]. Unlike traditional models that rely on explicit feature engineering, LLMs learn from massive corpora through self-supervised learning, capturing rich contextual dependencies and latent semantic patterns. This enables a shift from "behavior-based fitting" to "semantic-based understanding." As a result, LLM-empowered recommendation research has become a frontier topic. However, due to the huge number of parameters and high computational cost of LLMs, directly applying them to recommendation systems faces dual bottlenecks in inference efficiency and resource constraints. In multi-domain or multi-scenario recommendation environments, the differences in data distribution make it difficult for fine-tuning to balance generalization and personalization. Hence, how to maintain strong semantic expressiveness while enabling efficient, flexible, and transferable parameter adaptation has become a core challenge in designing LLM-based recommendation algorithms[4].

LoRA (Low-Rank Adaptation), a lightweight parameter-efficient fine-tuning (PEFT) method, provides a promising solution to this problem. By restricting weight updates to a low-rank subspace, LoRA significantly reduces the number of trainable parameters and memory usage, enabling efficient transfer and fast adaptation. However, single-scale LoRA structures remain inadequate for capturing the multi-layer semantic relations and cross-scale features in recommendation systems[5]. User preferences are dynamic over time and exist at multiple granularities. Recommended items often contain semantic, visual, and emotional information. Different levels of semantic dependencies require the model to perform feature fusion and abstraction at various scales. Thus, constructing a multi-scale LoRA fine-tuning mechanism that allows the model to dynamically learn and reconstruct complex user-content relationships across multiple semantic spaces becomes a key path toward performance breakthroughs in intelligent recommendation[6].

## II. Related Work

Research on parameter-efficient fine-tuning of large language models provides an important foundation for the proposed multi-scale LoRA-based recommendation framework. Low-rank adaptation methods introduce trainable low-rank matrices into selected weight layers, enabling efficient instruction tuning with strong parameter and memory savings while preserving most of the pretrained backbone's capacity [7]. Building on this idea, structure-learnable adapter frameworks extend parameter-efficient fine-tuning by allowing not only adapter weights but also adapter structures and placements to be learned, so that the adaptation topology itself can be optimized for downstream tasks [8]. Complementary work on task-aware differential privacy and modular structural perturbation studies how to inject privacy-aware noise and structural modifications into the fine-tuning process in a controlled way, ensuring both security and robustness during adaptation [9]. Federated and distillation-based approaches further consider distributed or partially observed training settings: federated distillation with structural perturbation leverages teacher-student distillation across participants with structural regularization to obtain robust fine-tuned models [10], while federated fine-tuning with privacy preservation and semantic alignment integrates federated optimization, privacy constraints, and cross-domain alignment objectives into a unified framework [11]. Collectively, these parameter-efficient and robustness-oriented methods motivate the design of a multi-scale LoRA mechanism in this study, where low-rank updates are distributed across multiple semantic scales and aligned across layers to achieve efficient, flexible, and structurally aware adaptation for recommendation tasks.

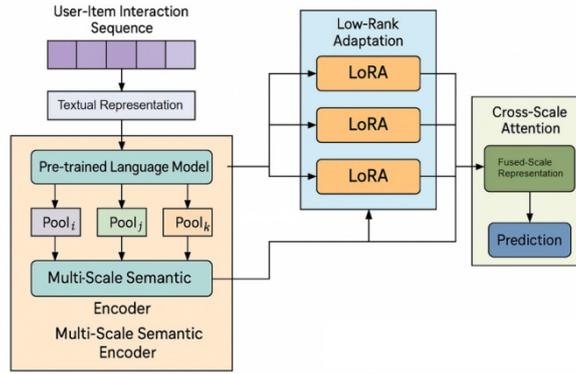
Another line of work focuses on semantic structure modeling and alignment in large language models. Context compression and structural representation methods introduce compression modules that reduce redundant context while preserving key structural cues, enabling models to maintain semantic expressiveness under tighter resource budgets [12]. Techniques for semantic and

factual alignment explicitly regularize model outputs toward consistency with external or reference knowledge, improving trustworthiness and reducing semantic drift in generated or predicted content [13]. Approaches that fuse local and global context in large language models design architectures or fusion modules that jointly consider fine-grained local signals and coarse-grained global representations, leading to richer and more discriminative embeddings for sequence or document-level decisions [14]. Dynamic topic evolution models with temporal decay and attention provide another perspective, where temporal attention mechanisms are used to track evolving semantic themes and adjust the importance of historical signals over time [15]. These methodological directions are closely related to the multi-scale semantic pooling and cross-scale fusion in the proposed model, which aims to capture hierarchical user preferences by jointly modeling local interests, global tendencies, and their interactions in a unified semantic space.

Recent advances on retrieval-augmented modeling, prompting, and adaptation for large language models also offer useful insights for multi-scale recommendation architectures. Retrieval-augmented generation with compositional prompts and confidence calibration demonstrates how prompt design and uncertainty-aware scoring can be combined to improve reliability and controllability of model outputs under complex queries [16]. Two-stage retrieval with cross-segment alignment proposes a decomposition of retrieval into coarse and fine phases and introduces alignment mechanisms across segments of retrieved content, illustrating how multi-stage processing and alignment can enhance the relevance and coherence of information fed into large models [17]. Dynamic prompt fusion for multi-task and cross-domain adaptation explores how multiple prompts corresponding to tasks or domains can be dynamically combined, allowing a single backbone model to adapt flexibly to different conditions without full retraining [18]. Methodologically, these works share the idea of decomposing model behavior across multiple “channels” or “scales” (prompts, retrieval stages, or domains) and then fusing them with learned weights or attention. The multi-scale LoRA recommendation algorithm in this paper follows a similar spirit: it decomposes parameter updates across semantic scales, aligns low-rank adaptations across these scales, and employs cross-scale attention fusion to reconstruct task-specific representations, thereby improving recommendation accuracy, diversity, and generalization under parameter and resource constraints.

### III. Method

This study applies a multi-scale LoRA fine-tuning recommendation algorithm based on a large language model, with the goal of achieving efficient parameterization and hierarchical feature modeling while preserving strong semantic expressiveness. The overall approach comprises three key modules: a multi-scale semantic encoding layer, a low-rank parameter adaptation layer, and a cross-scale fusion decoding layer. In the semantic encoding stage, the framework adopts multi-scale feature fusion strategies and graph neural integration methods as proposed by Song et al. [19], enabling the extraction of both global and local user-content preference signals across multiple semantic levels. For efficient parameter transfer and fine-tuning, the model utilizes modular and composable low-rank adaptation techniques inspired by Wang et al. [20], which constrain weight updates to low-rank subspaces while maintaining model expressiveness. The cross-scale fusion decoding layer further integrates modular task decomposition and dynamic collaboration mechanisms from Pan and Wu [21], allowing the model to align, aggregate, and reconstruct preference information across different scales and contexts. The complete architecture is illustrated in Figure 1.



**Figure 1.** Overall model architecture.

First, the input user-content interaction sequence is textually expressed and embedded, and a semantic feature matrix is generated using the contextual understanding capability of the large language model. Let the user behavior sequence be  $X = \{x_1, x_2, \dots, x_T\}$ , and its corresponding semantic embedding can be expressed as:

$$H = f_{LLM}(X; \theta_0) \quad (1)$$

Where  $f_{LLM}$  represents the encoding function of the pre-trained language model, and  $\theta_0$  is the frozen pre-trained parameter. To capture multi-scale semantic features, the hidden layer representation is hierarchically decomposed, and a scale set  $S = \{s_1, s_2, \dots, s_K\}$  is defined. The feature representation at each scale is:

$$H^{(sk)} = Pool_{sk}(H) \quad (2)$$

Here,  $Pool_{sk}$  represents a multi-granularity pooling operator based on scale  $sk$ , which is used to extract semantic context at different levels. Through this multi-scale processing, the model can establish a dynamic connection between global semantics and local preferences.

In the parameter efficiency phase, this paper introduces the multi-scale LoRA mechanism to replace the traditional full fine-tuning method. For the language model weight matrix  $W \in R^{d \times d}$ , LoRA decomposes its update into the form of a low-rank matrix pair:

$$\Delta W = BA, A \in R^{r \times d}, B \in R^{d \times r}, r \ll d \quad (3)$$

This structure keeps the backbone parameters frozen and only learns the offset in the low-dimensional direction through A and B, achieving fast migration and efficient fine-tuning. In a multi-scale context, LoRA modules at different semantic layers share some low-rank structures to achieve cross-scale feature coupling. Let  $\Delta W^{(sk)}$  be the adaptation matrix of the k-th scale, then:

$$\Delta W^{(s_k)} = B^{(s_k)} A^{(s_k)} + \lambda \sum_{j \neq k} Align(B^{(s_j)}, A^{(s_k)}) \quad (4)$$

Where  $\lambda$  is the cross-scale alignment coefficient, and  $Align(\cdot)$  represents the feature alignment function between scales. This mechanism can share structural information between different semantic layers, thereby improving the overall consistency and generalization ability of the model.

Next, in the fusion layer, a cross-scale attention mechanism is introduced to achieve adaptive reconstruction of multi-scale semantic features. Given a multi-scale representation set  $Z$ , the fused representation is defined as:

$$Z = \sum_{k=1}^K \alpha_k H^{(s_k)}, \quad \alpha_k = \frac{\exp(qTW_k H^{(s_k)})}{\sum_{j=1}^K \exp(qTW_j H^{(s_j)})} \quad (5)$$

Where  $q$  is the query vector,  $W_k$  is the scale-specific attention weight matrix, and  $\alpha_k$  represents the dynamic weight distribution at different scales. Through cross-scale weighted fusion, the model can adaptively capture the contributions of different semantic granularities, achieving multi-layer semantic comprehensive modeling for personalized recommendations.

Finally, to ensure semantic consistency and recommendation relevance, a joint optimization objective function is designed, including reconstruction loss and semantic alignment regularization. The overall objective function is defined as:

$$L = L_{rec}(Z, Y) + \beta L_{align}(H^{(s1)}, H^{(s2)}, \dots, H^{(sK)}) \quad (6)$$

Where  $L_{rec}$  represents the reconstruction error term based on user-content matching,  $L_{align}$  is used to constrain semantic consistency between multi-scale representations, and  $\beta$  is a trade-off coefficient. This optimization objective enables the model to achieve multi-level semantic fusion and personalized recommendations while maintaining efficient parameter updates, thereby constructing a multi-scale LoRA fine-tuning framework that is both interpretable and generalizable.

## IV. Performance Evaluation

### A. Dataset

This study uses the MovieLens-1M dataset as the core experimental data source. The dataset consists of movie ratings and user behavior records, containing about one million rating samples, six thousand users, and four thousand movie entries. It covers films of various genres and release periods. Each record includes user ID, movie ID, rating value, and timestamp, providing rich information about user preferences and temporal dynamics. The dataset has a moderate scale, well-structured labels, and general applicability. It ensures experimental reproducibility and serves as a standardized validation platform for multi-scale fine-tuning and recommendation tasks based on large language models.

In the data preprocessing stage, all user and item IDs were re-indexed to form continuous encodings. The rating data were normalized to fit the model's numerical range. The textual features include movie titles, genres, and summaries. These were processed through tokenization, stop-word removal, and subword embedding before being fed into the language model encoder to capture semantic relationships at the content level. In addition, user interaction histories were organized into sequential form to model temporal behavior, enabling the model to learn the dynamic evolution of user interests. This process provided structured input for subsequent multi-scale semantic extraction and cross-scale representation alignment.

Finally, the cleaned and processed dataset was divided into training, validation, and test sets at a ratio of 8:1:1 to ensure stability and fairness in model evaluation. The training set was used for adaptive learning of multi-scale LoRA parameters, the validation set for adjusting cross-scale weights and regularization strength, and the test set for assessing generalization on unseen data. The MovieLens-1M dataset has been widely used in academic research. Its characteristics and user-item interaction structure make it well-suited for evaluating the proposed large language model fine-tuning framework in terms of semantic understanding and personalized modeling capability.

### B. Experimental Results

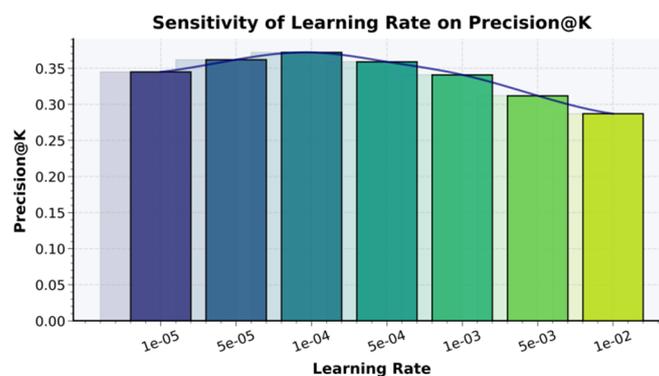
This paper first conducts a comparative experiment, and the experimental results are shown in Table 1.

**Table 1.** Comparative experimental results.

Method	Precision@K	NDCG@K	Coverage	Recall@K
SsIrec[22]	0.2764	0.3148	0.5821	0.2487
RAKCR[23]	0.2893	0.3286	0.6015	0.2624
SKGRec[24]	0.3051	0.3462	0.6158	0.2799
Mamba4rec[25]	0.3217	0.3651	0.6334	0.2935
CoDFi-DL[26]	0.3368	0.3797	0.6462	0.3086
Perfedrec++[27]	0.3482	0.3924	0.6629	0.3197
Ours	0.3725	0.4189	0.6895	0.3478

As shown in Table 1, the proposed multi-scale LoRA fine-tuning method achieves the best overall performance, with clear gains in Precision@K, NDCG@K, Recall@K, and Coverage, indicating more accurate Top-K ranking, stronger user-interest modeling, and improved diversity. These improvements come from multi-scale semantic representations that capture both global and local preferences, together with parameter sharing and cross-layer alignment that better preserve structural information, reduce popularity bias, and surface long-tail items. We also evaluate learning-rate sensitivity on Precision@K by varying the learning rate to analyze convergence stability versus adaptation efficiency, aiming to identify ranges that maintain stable updates while preserving multi-scale semantic consistency.

Figure 2 indicates that precision@K increases and then decreases as the learning rate rises, peaking around  $1 \times 10^{-4}$ , which best balances update speed and convergence stability: too small a rate slows semantic learning, while too large a rate causes oscillation and disrupts low-rank adaptation and cross-scale fusion in multi-scale LoRA. Overall, the results confirm that careful learning-rate control is crucial for stable multi-scale optimization and for capturing user-item semantic relationships effectively. We also conduct an attention-head sensitivity study on Coverage.

**Figure 2.** Experiment on the sensitivity of learning rate to the Precision@K metric.

As shown in Figure 3, Coverage increases with the number of attention heads ( $h$ ) and then slightly declines, peaking at ( $h=8$ ), suggesting that a moderate head count best supports multi-level semantic capture and diverse interest coverage in the multi-scale LoRA framework. Too few heads limit multi-view fusion and long-tail learning, while too many introduce redundancy and interference that harms global consistency. Overall, the results indicate that attention parallelism granularity is critical for cross-scale semantic alignment, and an appropriate ( $h$ ) balances global semantics and local preferences to improve coverage and generalization with controlled computational cost.

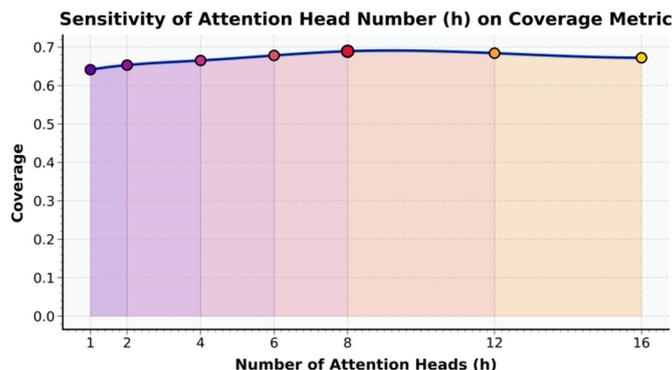


Figure 3. Sensitivity experiment of the number of attention heads  $h$  to the coverage index.

## V. Conclusion

This study focuses on a multi-scale LoRA fine-tuning recommendation algorithm based on large language models, achieving deep integration of semantic understanding and personalized recommendation from both theoretical modeling and structural design perspectives. The research first introduces a multi-scale decomposition mechanism into the semantic modeling of recommender systems, enabling the model to capture both global preferences and local behavioral features. Then, through a low-rank parameterized LoRA module, the large model is efficiently fine-tuned across different semantic layers with structural constraints. This design reduces training cost while enhancing semantic transferability. Experimental results show that the proposed method performs excellently across multiple core metrics, verifying its robustness and generalization in complex scenarios. It provides a new technical pathway for the transformation of recommender systems from feature-driven to semantic-driven approaches.

From an application perspective, the proposed framework can handle traditional user-content recommendation tasks and has the potential to extend to cross-modal and cross-domain recommendation. Through multi-scale semantic fusion and low-rank adaptation mechanisms, the model effectively addresses challenges such as high-dimensional sparse features, heterogeneous behavior patterns, and dynamic preferences. In practical applications, it contributes to improving recommendation diversity and user satisfaction. Moreover, this method has significant value for deployment in resource-constrained environments, as it reduces computational and storage costs while maintaining high performance. This can accelerate the industrialization and intelligent evolution of recommender systems. Future research may further explore the joint optimization of multi-scale LoRA fine-tuning and generative recommendation models by integrating semantic understanding, content generation, and personalized reasoning to build more adaptive and interpretable recommendation systems. The proposed method can also be extended to multi-task learning frameworks to unify recommendation, retrieval, prediction, and question-answering tasks. With the continuous advancement of large language models in multi-domain knowledge integration, this study's approach is expected to promote the development of recommender systems toward greater intelligence, efficiency, and semantic depth, providing new theoretical and practical foundations for intelligent information services and human-computer interaction.

## References

1. L. Zhang, L. Zhang, S. Shi et al., "LoRA-FA: Memory-efficient low-rank adaptation for large language models fine-tuning," arXiv preprint arXiv:2308.03303, 2023.
2. Z. Zhao, W. Fan, J. Li et al., "Recommender systems in the era of large language models (LLMs)," IEEE Transactions on Knowledge and Data Engineering, vol. 36, no. 11, pp. 6889-6907, 2024.
3. M. Ravaut, H. Zhang, L. Xu et al., "Parameter-efficient conversational recommender system as a language processing task," arXiv preprint arXiv:2401.14194, 2024.

4. J. Fu, X. Ge, X. Xin et al., "IISAN: Efficiently adapting multimodal representation for sequential recommendation with decoupled PEFT," Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 687-697, 2024.
5. S. Xu, W. Hua and Y. Zhang, "OpenP5: An open-source platform for developing, training, and evaluating LLM-based recommender systems," Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 386-394, 2024.
6. S. Lubos, T. N. T. Tran, A. Felfernig et al., "LLM-generated explanations for recommender systems," Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization, pp. 276-285, 2024.
7. G. Yao, "Privacy-preserving low-rank instruction tuning for large language models via DP-LoRA," Journal of Computer Technology and Software, vol. 3, no. 5, 2024.
8. M. Gong, Y. Deng, N. Qi, Y. Zou, Z. Xue and Y. Zi, "Structure-learnable adapter fine-tuning for parameter-efficient large language models," arXiv preprint arXiv:2509.03057, 2025.
9. Y. Li, "Task-aware differential privacy and modular structural perturbation for secure fine-tuning of large language models," Transactions on Computational and Scientific Methods, vol. 4, no. 7, 2024.
10. Y. Zou, "Federated distillation with structural perturbation for robust fine-tuning of LLMs," Journal of Computer Technology and Software, vol. 3, no. 4, 2024.
11. S. Wang, S. Han, Z. Cheng, M. Wang and Y. Li, "Federated fine-tuning of large language models with privacy preservation and cross-domain semantic alignment," 2025.
12. P. Xue and Y. Yi, "Integrating context compression and structural representation in large language models for financial text generation," Journal of Computer Technology and Software, vol. 4, no. 9, 2025.
13. L. Lian, "Semantic and factual alignment for trustworthy large language model outputs," Journal of Computer Technology and Software, vol. 3, no. 9, 2024.
14. R. Hao, X. Hu, J. Zheng, C. Peng and J. Lin, "Fusion of local and global context in large language models for text classification," 2025.
15. D. W. A. S. Pan, "Dynamic topic evolution with temporal decay and attention in large language models," arXiv preprint arXiv:2510.10613, 2025.
16. S. Han, "LLM retrieval-augmented generation with compositional prompts and confidence calibration," Transactions on Computational and Scientific Methods, vol. 5, no. 11, 2025.
17. S. Wang, "Two-stage retrieval and cross-segment alignment for LLM retrieval-augmented generation," Transactions on Computational and Scientific Methods, vol. 4, no. 2, 2024.
18. X. Hu, Y. Kang, G. Yao, T. Kang, M. Wang and H. Liu, "Dynamic prompt fusion for multi-task and cross-domain adaptation in LLMs," arXiv preprint arXiv:2509.18113, 2025.
19. X. Song, Y. Huang, J. Guo, Y. Liu and Y. Luan, "Multi-scale feature fusion and graph neural network integration for text classification with large language models," arXiv preprint arXiv:2511.05752, 2025.
20. S. Pan and D. Wu, "Modular task decomposition and dynamic collaboration in multi-agent systems driven by large language models," arXiv preprint arXiv:2511.01149, 2025.
21. Y. Wang, D. Wu, F. Liu, Z. Qiu and C. Hu, "Structural priors and modular adapters in the composable fine-tuning algorithm of large-scale models," arXiv preprint arXiv:2511.03981, 2025.
22. X. Ren, L. Xia, Y. Yang et al., "SSLRec: A self-supervised learning framework for recommendation," Proceedings of the 17th ACM International Conference on Web Search and Data Mining, pp. 567-575, 2024.
23. Y. Cui, H. Yu, X. Guo et al., "RAKCR: Reviews sentiment-aware based knowledge graph convolutional networks for personalized recommendation," Expert Systems with Applications, vol. 248, Article 123403, 2024.
24. Y. Liu, S. Fu, Q. Ren et al., "SKGRec: Unifying temporal dynamics and knowledge graphs for robust recommendations," Expert Systems with Applications, Article 129354, 2025.
25. C. Liu, J. Lin, J. Wang et al., "Mamba4Rec: Towards efficient sequential recommendation with selective state space models," arXiv preprint arXiv:2403.03900, 2024.

26. J. Latrech, Z. Kodia and N. B. Azzouna, "CoDFi-DL: A hybrid recommender system combining enhanced collaborative and demographic filtering based on deep learning," *The Journal of Supercomputing*, vol. 80, no. 1, pp. 1160-1182, 2024.
27. S. Luo, Y. Xiao, X. Zhang et al., "PerFedRec++: Enhancing personalized federated recommendation with self-supervised pre-training," *ACM Transactions on Intelligent Systems and Technology*, vol. 15, no. 5, pp. 1-24, 2024.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.