



LLM-Assisted Multimodal Hybrid Recommendation System for Amazon Electronics

Harsh Singh Tiwari¹, Achanta Srinivas Varshith^{2*}, Muthaiah Uchimuthu³

^{1,2,3}School of Computing Amrita Vishwa Vidyapeetham, Chennai, India

*Correspondence to: Vedant Ghodake, Department of Information Technology, Vishwakarma Institute of Technology, Pune, India; E-mail: vedant.ghodake241@vit.edu

Received: July 13, 2026; Manuscript No: JADS-26-1788; Editor Assigned: July 29, 2026; PreQc No: JADS-26-1788 (PQ); Reviewed: August 03, 2026; Revised: August 20, 2026; Manuscript No: JADS-26-1788 (R); Published: October 06, 2026

Citation: Tiwari HS, Varshith AS, Uchimuthu M (2026) LLM-Assisted Multimodal Hybrid Recommendation System for Amazon Electronics. Adv. Data Sci. Comput. Intell. Vol.1 Iss.1, October (2026), pp:21-28.

Copyright: © 2026 Harsh Singh Tiwari, Achanta Srinivas Varshith, Muthaiah Uchimuthu. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

This theoretical article proposes an LLM-assisted multimodal hybrid recommendation framework for the Amazon Electronics domain. Traditional recommendation methods that rely primarily on user-item interactions often degrade under sparse-data and cold-start conditions. The proposed architecture combines three complementary sources of information: Alternating Least Squares (ALS) collaborative factors, lightweight MiniLM-based semantic embeddings derived from product meta-data and reviews, and visual product features projected into the collaborative embedding space. These signals are integrated through a weighted hybrid scoring mechanism intended to preserve behavioural personalisation while improving semantic and multimodal coverage. The article synthesises 21 recent studies on collaborative, LLM-based, multimodal, graph-based, and hybrid recommendation, identifies unresolved integration and deployment gaps, and specifies a methodology for future implementation and empirical evaluation using Precision@K, Recall@K, NDCG@K, and cold-start cohort analysis. The contribution is therefore a research hypothesis and system design rather than a report of completed experimental results.

Keywords: Recommender Systems; Collaborative Filtering; Large Language Models; Multimodal; Semantic Embeddings; MiniLM; ALS; Cold-Start; Amazon Electronics

INTRODUCTION

Recommendation systems have become integral to digital platform operations, including product discovery engines for e-commerce websites such as Amazon and Flipkart, and content recommendation systems for streaming platforms such as Netflix and Spotify. With the increasing volume of products and users, there is a growing need for smart, scalable, and context-aware recommender systems.

Traditional Collaborative Filtering (CF)-based recommendation methods, although highly effective in data-rich situations, are subject to several well-known shortcomings. Such approaches heavily depend on user-item interaction matrices and therefore struggle under two major conditions: (i) the cold-start problem, which refers to

the absence of sufficient interaction history for new users or items, and (ii) data sparsity, where the interaction matrix is so sparse that meaningful signals cannot be obtained for a large number of user-item pairs. Therefore, additional information sources are needed to overcome these problems.

With the advent of Large Language Models (LLMs), recommendation systems have gained an additional capability to understand the complex semantics of user intent, item descriptions, and reviews. However, LLM-based recommenders may lack collaborative components and therefore provide limited behavioural personalisation when sufficient interaction history is available.

To address these limitations, multimodal recommendation systems have emerged that use both textual and visual product data. Although many current multimodal systems exhibit excellent performance, they often rely on heavyweight end-to-end vision and text transformers. Such infrastructure introduces substantial computational and memory overhead, making these systems less suitable for low-latency, real-time applications.

In contrast to recent hybrid models that either depend on heavyweight multimodal encoders or maintain separate text and collaborative pipelines, this paper introduces an LLM-Assisted Multimodal Hybrid Recommendation System that addresses the above-mentioned challenges through a lightweight approach. Our approach uses:

Collaborative Filtering

Through ALS to leverage user item interaction patterns.

Semantic Embeddings

Through MiniLM, a lightweight transformer model, to capture contextual semantics from product reviews and metadata.

Multimodal Fusion

Through direct projection of visual product representations into the collaborative embedding space without the overhead of image-specific transformers.

Weighted Hybrid Scoring Mechanism

To combine the above-mentioned signals and generate personalised Top-N recommendations on the Amazon Electronics dataset.

MOTIVATION

The motivation for this work is based on five major observations in the evolving landscape of recommender systems:

Failure of CF under sparsity and cold-start:

Collaborative filtering methods are fundamentally dependent on the density of user-item interaction data. In real-world deployments, interaction matrices are highly sparse often exceeding 99% sparsity and new users or newly launched products may have no historical interactions. Under these conditions, purely CF-based systems may fail to generate recommendations or produce highly inaccurate ones.

LLMs lack behavioural grounding:

While LLM-based recommenders are strong at comprehending natural language, reviews, and product semantics, they operate without direct knowledge of collective user behaviour patterns. This can result in

recommendations that are semantically coherent but poorly personalised to individual users.

Multimodal systems are computationally prohibitive:

Systems that process raw image and text streams through separate large-scale encoders impose significant inference latency and memory costs. This makes them unsuitable for real-time recommendation in production environments.

Need for unified but efficient fusion:

These limitations motivate a lightweight framework that combines collaborative, semantic, and multimodal signals without the prohibitive computational cost of a large, end-to-end multimodal transformer.

Broad real-world applications in e-commerce:

A strong and reliable hybrid system is important for high-volume e-commerce platforms such as Amazon Electronics, where massive product catalogues, frequent cold-start items such as newly launched gadgets, and rich textual reviews require a scalable, low-latency solution.

LITERATURE SURVEY

This section reviews 21 studies published or made publicly available from 2023 to 2025 in four interconnected areas: collaborative filtering and LLM integration, multimodal recommendation, graph-based recommendation, and hybrid or specialised systems. The review examines how behavioural, semantic, and visual information are represented and fused, as well as the computational and evaluation constraints associated with the proposed framework.

Collaborative Filtering and LLM Integration

Zhang, introduced CoLLM, which injects collaborative embeddings into an LLM so that language-based recommendation is grounded in user-item interaction patterns [1]. The method improves recommendation quality and cold-start behaviour, but it does not explicitly model visual product information.

Kim, proposed A-LLMRec, an efficient architecture that enables an LLM to reuse collaborative knowledge from a pretrained CF recommender [2]. The model performs strongly across warm, cold-start, few-shot, cold-user, and cross-domain settings while avoiding extensive LLM fine-tuning. However, product-image fusion is not a central component of the design. Lin, combined collaborative filtering with LLM-based textual understanding to improve accuracy, diversity, and user satisfaction [3]. Their results support the value of combining behavioural and semantic information, although the framework gives limited attention to visual and other product modalities. Liu, proposed CCF-LLM, which represents interaction history through hybrid prompts and applies attentive cross-modal fusion to combine collaborative and semantic embeddings

[4]. The work demonstrates stronger synergy between CF and LLM signals than simple instruction tuning or direct embedding injection, but it does not explicitly focus on product-image modelling.

Li, investigated LLM-based recommendation systems, focusing on the semantic representation of users and items [5]. The study highlights scalability and multimodal integration as significant challenges for production deployment. Kumar, compared LLM-based recommendation approaches and reported advantages over classical NLP-based recommendation methods. However, multimodal fusion remains insufficiently explored [6].

Liu, surveyed LLM-based recommender systems [7]. The study covers text prompts, instruction tuning, representation learning, and generation-based approaches. The survey highlights the potential of LLMs and explains the lack of consistent large-scale and multimodal evaluation.

Munson, explored the use of LLMs for recommendation and suggested that their world knowledge and language understanding can outperform traditional text models, while domain-specific multimodal integration remains a challenge [8].

Multimodal Recommendation

Wang, investigated large-scale multimodal benchmarks and architectures, demonstrating that multimodal recommendation can outperform unimodal recommendation across product categories [9]. Nevertheless, the evaluated systems are mostly based on traditional text encoders and image features rather than LLM-based semantic modeling.

Liu, provided a comprehensive survey of multimodal recommendation systems involving text, images, audio, and video [10]. The survey explains the advantages of cross modal fusion but gives limited treatment to newer LLM-driven recommendation pipelines.

Luo, proposed a sequential recommender called Molar [11]. A multimodal LLM is utilised as the basis of a unified framework that represents items and aligns content based user representations with collaborative ID-based signals. The method improves recommendation accuracy, but its additional encoding and post-alignment stages may increase computational and architectural complexity.

Lopez-Avila and Du, surveyed the use of LLMs in multimodal recommender systems [12]. The work summarises existing research on modality handling, disentanglement, alignment, fusion, datasets, and evaluation.

Xu, reviewed recent multimodal recommender system advances through four technical components: feature extraction, encoders, multimodal fusion, and loss functions [13]. The survey identifies major research

directions, but its scope is broad and does not provide an Amazon Electronics-specific system design.

Zhou, proposed a multimodal LLM framework for sequential recommendation that models temporal user behaviour and multimodal item content [14]. Its scalability on very large item catalogues was not fully evaluated. Smith, used deep neural models for text-image recommendation and reported gains over unimodal baselines, but did not use LLM-derived semantic embeddings [15].

Graph Neural Network Based Approaches

Li, proposed a GNN-based multimodal representation-learning framework with deep mutual learning for sparse recommendation settings [16]. The method produces robust multimodal representations, but it does not incorporate the semantic reasoning available from an LLM backbone.

Chen, discussed graph-based recommendation modelling, where complex and higher-order user-item relationships are captured [17].

Guo, proposed a GNN-based multimodal fusion system intended to improve robustness under noisy interaction data [18]. The method benefits from graph and multimodal signals, but does not incorporate LLM-assisted semantic representations.

Hybrid and Specialised Systems

Zhao, surveyed hybrid recommender systems and concluded that combining several recommendation paradigms improves performance over isolated methods [19]. However, LLM-assisted hybrid architectures receive limited treatment.

Raj, investigated latent-feature integration and optimisation for matrix-factorisation-based recommendation [20]. The approach improves recommendation quality, but does not use LLM-generated embeddings to enrich the latent space.

Chen, demonstrated that LLMs can perform zero shot point-of-interest recommendation without target-domain training [21]. The result is promising for cold-start recommendation, but the method remains focused on the POI domain rather than general e-commerce recommendation involving product text and images.

Summary and Comparison

Table I summarises the 21 reviewed studies, including their core methods, principal findings, and remaining research gaps.

Table II provides a higher-level comparison of the four major recommendation paradigms.

RESEARCH GAPS

Although there have been substantial advances in all four paradigms examined, several unanswered questions remain and motivate the proposed system.

- Existing literature has often considered collaborative filtering and LLM-based semantics as distinct, non-communicating pipelines, resulting in weak CF-LLM integration. The two information sources are not tightly coupled, which can result in inadequate personalisation: LLM outputs are not grounded in interaction behaviour, while CF models lack sufficient semantic information.
- For LLM and multimodal systems, a large-scale transformer backbone is often required for both text and image encoding. This introduces substantial inference latency and memory overhead, making real-time recommendations at scale difficult.
- Current systems often provide limited explanations for their recommendations. This opacity can reduce user confidence in the system and make debugging and auditing more complex.
- Textual, visual, and behavioural information is often fused in a shallow or ad hoc manner. There is a lack of a principled and uniform fusion mechanism that embeds multimodal features in the same space as collaborative embeddings.
- Academic benchmarks show that complex hybrid models can outperform simpler ones, but only a limited number of studies discuss engineering considerations for real-world deployment, such as inference speed, memory requirements, and API compatibility.

The proposed system is intended to address these gaps as follows.

The proposed framework is inspired by CoLLM [1], A-LLMRec [16], Molar [17], and CCF-LLM [21], but is adapted toward a lightweight e-commerce architecture that explicitly combines collaborative, textual, and visual product signals.

The Collaborative-Semantic Alignment layer incorporates embeddings generated by the collaborative filtering method (ALS factors) alongside semantic embeddings generated by the language model. This aligns semantic information with user behaviour and is intended to reduce the risk of semantically plausible but behaviourally unsupported recommendations.

Efficient Lightweight Architecture

MiniLM is chosen over large-scale language models. It provides strong semantic representation capabilities while requiring substantially fewer computational resources and can be deployed on commodity hardware.

The proposed architecture encodes multimodal visual information without relying on a heavyweight image encoder within the collaborative embedding space. This enables the integration of visual information while reducing computational cost.

To address cold-start conditions, the architecture combines collaborative personalisation when interaction history exists with semantic and visual item representations when behavioural evidence is limited.

OBJECTIVES

This research has the following objectives:

- Perform an extensive literature review of collaborative, LLM-based, and multimodal recommender systems published between 2023 and 2025.
- Identify the main challenges, including the cold-start problem, data sparsity, and the lack of unified integration of semantic, behavioural, and multimodal information.
- Identify and select relevant features from the Amazon Electronics dataset, including user interactions, product reviews, ratings, product metadata, and product images, and preprocess them.

Ref.	Authors (Year)	Core Method	Key Findings	Research Gap
[1]	Zhang et al. (2023)	CF embeddings integrated with LLMs	Improves accuracy and cold-start performance	Multimodal features not considered
[2]	Li et al. (2024)	GNN with deep mutual learning	Produces robust multimodal representations	LLM-based semantic understanding is absent
[3]	Wang et al. (2024)	Large-scale multimodal bench-marking	Multimodal fusion outperforms uni-modal approaches	Does not incorporate LLM-based semantic modelling
[4]	Liu et al. (2024)	Survey of multimodal recom-mender systems	Highlights the benefits of multi-modal fusion	Limited focus on LLM-driven rec-ommendation approaches

[5]	Chen et al. (2024)	Graph-based recommendation models	Captures complex user-item relationships	Integration of multimodal information and LLMs remains unexplored
[6]	Li et al. (2024)	LLM-assisted recommender systems	Enhances semantic understanding of users and items	Faces scalability and multimodal-fusion challenges
[7]	Munson et al. (2025)	Review of LLMs in recommendation	LLMs outperform traditional text-based models	Lacks domain-specific multimodal recommendation systems
[8]	Raj et al. (2025)	Latent-feature integration and optimisation	Improves overall recommendation quality	Does not incorporate LLM-generated embeddings
[9]	Liu et al. (2024)	Survey of LLM-powered recommender systems	Demonstrates the strong potential of LLM-based recommendation	Limited evaluation on large-scale multimodal datasets
[10]	Kumar et al. (2025)	Comparative analysis of LLM-based methods	LLMs outperform classical NLP-based recommendation methods	Multimodal fusion remains insufficient
[11]	Zhou et al. (2025)	Multimodal LLMs for sequential recommendation	Improves sequence-aware recommendation performance	Scalability on large item catalogues is not evaluated
[12]	Guo et al. (2024)	GNN with multimodal fusion	Improves robustness under noisy interaction data	Does not incorporate LLM-assisted semantic representations
[13]	Chen et al. (2025)	Zero-shot recommendation using LLMs	Enables effective zero-shot point-of-interest recommendation	Limited primarily to the point-of-interest domain
[14]	Zhao et al. (2024)	Survey of hybrid recommender systems	Hybrid models generally outperform single-method systems	Does not sufficiently examine LLM-assisted hybrid architectures
[15]	Smith et al. (2024)	Deep learning for text-image fusion	Improves recommendation accuracy through multimodal information	Does not use LLMs for semantic representation
[16]	Kim et al. (2024)	A-LLMRec: collaborative knowledge integrated with LLM reasoning	Performs effectively across warm, cold-start, few-shot and cross-domain scenarios	Product-image fusion is not a central component
[17]	Luo et al. (2024)	Molar: multimodal LLM with collaborative-filtering alignment	Aligns multimodal content representations with collaborative and ID-based signals	The additional alignment pipeline may increase architectural and computational complexity
[18]	Lopez-Avila and Du (2025)	Survey of LLMs in multimodal recommender systems	Maps the use of LLMs with textual, visual, structured and interaction data	Provides a survey rather than an implemented recommendation pipeline

[19]	Xu et al. (2025)	Survey of recent multimodal recommender-system advances	Reviews feature extraction, en-coders, fusion methods, loss functions and research trends	Broad survey without an Amazon Electronics-specific implementation
[20]	Lin et al. (2024)	Enhanced recommendation combining collaborative filtering and LLMs	Combines behavioural preference modelling with LLM-based textual understanding	Provides limited treatment of visual and other product modalities
[21]	Liu et al. (2024)	CCF-LLM: collaborative cross-modal fusion using hybrid prompts and attentive fusion	Effectively combines semantic knowledge with collaborative user-behaviour signals	Focuses on semantic-collaborative fusion without explicit product-image modelling

TABLE I: SUMMARY OF RELATED WORKS

Approach	Strength	Limitation	Key Idea
Collaborative Filtering	Good personalisation	Fails in cold-start	User-item interactions
LLM-based Models	Captures semantics	No behavioural grounding	Text understanding
Multimodal Systems	Rich features	High computational cost	Text + image
GNN-based Models	Captures complex relations	Complex & heavy	Graph relationships

TABLE II: COMPARISON OF RECOMMENDATION PARADIGMS

- Perform feature extraction from reviews, product meta data, and ratings for the collaborative and semantic pipelines.
- Generate dense semantic embeddings using the MiniLM transformer-based sentence encoder.
- Apply the Alternating Least Squares (ALS) collaborative filtering model and train it on the user-item interaction matrix.
- Design a multimodal hybrid recommendation model that combines collaborative, semantic, and visual signals using an α -weighted scoring model.
- Define an empirical evaluation protocol based on Precision@K, Recall@K, and NDCG@K.
- Evaluate cold-start performance using interaction-frequency cohorts and compare the hybrid model against collaborative-only, semantic-only, and multimodal ablation baselines.

PROPOSED METHODOLOGY

System Architecture Overview

The proposed system consists of three stages: preprocessing of user-item interaction data, preprocessing of product text metadata, and preprocessing of product visual features, followed by feature fusion to generate a product recommendation score. The high-level architecture, shown in Fig. 1, consists of the following modules: (i) a Data Preprocessing Module,

(ii) A Dual Embedding Generation Module (collaborative + semantic), and (iii) a Hybrid Scoring and Ranking Module.

Data Preprocessing

The planned evaluation benchmark is the Amazon Electronics dataset. The proposed preprocessing pipeline involves:

- Deduplication and exclusion of items and users with fewer than a specified minimum number of interactions.
- Construction of a sparse user-item rating matrix $R \in \mathbb{R}^{|U| \times |I|}$, where $|U|$ and $|I|$ denote the numbers of users and items, respectively.
- Concatenation of the product title, category, and combined product-review text into a single textual representation for each product.
- Visual feature encoding by extracting and normalising product thumbnail images for visual feature representation.

Collaborative Filtering via ALS

Collaborative filtering is performed using the Alternating Least Squares algorithm, which decomposes the rating matrix R into two low-dimensional matrices:

$$R \approx UV^T \quad (1)$$

where $U \in \mathbb{R}^{|U| \times d}$, $V \in \mathbb{R}^{|I| \times d}$, and d is the embedding dimension. The following objective function is minimised:

$$\mathcal{L}_{CF} = \sum_{(u,i) \in \Omega} (r_{ui} - \mathbf{u}_u^\top \mathbf{v}_i)^2 + \lambda (\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2) \quad (2)$$

Training and validation are conducted on disjoint subsets of observed interactions Ω , with λ controlling regularisation.

Multimodal Visual Feature Integration

The proposed design uses a lightweight convolutional encoder to map product thumbnail images into a compact visual feature vector $\mathbf{f}_i^{\text{img}}$. Instead of supplying raw vectors to an independent large-scale Vision Transformer, visual features are projected into the d -dimensional collaborative space via a learned linear projection matrix $\mathbf{W}_{\text{img}}$ (optimized via secondary alignment against item latent factors):

$$\hat{\mathbf{v}}_i = \mathbf{v}_i + \mathbf{W}_{\text{img}} \mathbf{f}_i^{\text{img}} \quad (3)$$

The augmented item embedding $\hat{\mathbf{v}}_i$ replaces $\mathbf{v}_i$ in subsequent steps. Thus, the collaborative score for a user-item pair becomes:

$$s_{CF}(u, i) = \mathbf{u}_u^\top \hat{\mathbf{v}}_i \quad (4)$$

Hybrid Scoring and Ranking

The collaborative score $s_{CF}(u, i)$ and semantic score $s_{\text{sem}}(u, i)$ are linearly combined using a weighting factor α to obtain the final recommendation score:

$$s_{\text{hybrid}}(u, i) = \alpha \cdot s_{CF}(u, i) + (1 - \alpha) \cdot s_{\text{sem}}(u, i) \quad (5)$$

where $\alpha \in [0, 1]$ regulates the contribution of behavioural and semantic signals.

Evaluation Metrics

Three standard ranking metrics are proposed for evaluating the implemented model at cut-off K :

Precision@K

The proportion of recommended items in the Top- K list that are relevant.

Recall@K

The proportion of all relevant items that are retrieved in the Top- K list.

NDCG@K

A position-sensitive metric that assigns greater credit when relevant items appear earlier in the ranking.

DISCUSSION

Limitations, Expected Challenges and Practical Applicability

Although the proposed approach provides a lightweight solution to multimodal recommendation within the Amazon Electronics domain, several limitations, challenges, and aspects of practical applicability should be considered.

Limitations

Although the approach is computationally efficient through the use of a lightweight sentence encoder (MiniLM) and a linear projection of image features (Wimg), there may be cases where the model does not capture fine-grained nuances that could be captured by more advanced multimodal models. Furthermore, model performance depends on the balancing hyperparameter α , which may need to vary across different sparse-user cohorts.

Expected Challenges

Updating the user-item interaction matrix (R) and recomputing collaborative factors in near-real-time for an extremely large and continuously changing e-commerce catalogue such as Amazon Electronics presents several technical challenges. Moreover, processing cold-start items without sufficient text information, reviews, or thumbnail images presents additional difficulties.

Practical Applicability

Despite these limitations, the proposed approach is well suited to applications where inference time, memory footprint, and infrastructure costs are important considerations.

CONCLUSIONS AND FUTURE WORK

This theoretical article presented a review of 21 recent studies and proposed an architecture for an LLM-assisted multimodal hybrid recommendation system in the Amazon Electronics domain. The reviewed literature shows substantial progress in collaborative-LLM integration, multimodal representation learning, graph-based modelling, and hybrid recommendation. Nevertheless, a deployment-oriented framework

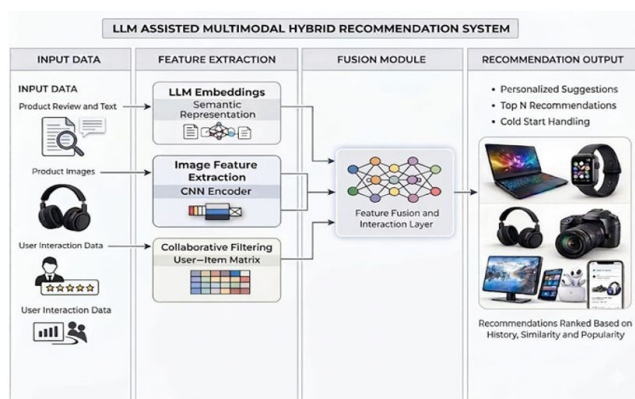


Fig. 1: The proposed LLM-Assisted Multimodal Hybrid Recommendation System. Three parallel feature extraction pipelines are used for the input data: LLM Embeddings (semantic representation via MiniLM), Image Feature Extraction (CNN Encoder), and Collaborative Filtering (User-Item Matrix). The extracted features are combined in a Feature Fusion and Interaction Layer, which generates personalised Top-N recommendations based on user history, similarity, and popularity.

That jointly balances behavioural grounding, semantic understanding, visual product context, cold-start robustness, and computational efficiency remains insufficiently established.

The proposed architecture addresses this gap by combining ALS-based collaborative factors, lightweight MiniLM semantic embeddings, and projected image features within a shared hybrid-scoring pipeline. Its expected advantage is not yet an empirical finding; rather, it is the central hypothesis to be tested through controlled implementation, ablation studies, ranking metrics, cold-start cohort analysis, and efficiency measurements.

Future work will focus on:

- Full implementation and training on the complete Amazon Electronics dataset.
- Empirical evaluation using Precision@K, Recall@K, and NDCG@K with different values of α and embedding dimension d .
- Cold-start analysis based on interaction frequency and user cohorts.
- Experiments with graph-based extensions, such as user item interaction graphs, to further enhance collaborative representations.
- Investigation of explainability mechanisms based on user contentment, usability, utility, and engagement traces to build user trust and improve system transparency.

REFERENCES

1. Zhang Y, Feng F, Zhang J, Bao K, Wang Q, He X. CoLLM: Integrating Collaborative Embeddings into Large Language Models for Recommendation. arXiv. 2023.

2. Kim S, Kang H, Choi S, Kim D, Yang M, Park C. Large Language Models meet Collaborative Filtering: An Efficient All-round LLM-based Recommender System. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '24). 2024. p. 1395-1406.
3. Lin X, Cheng Z, Yun L, Lu Q, Luo Y. Enhanced Recommendation Combining Collaborative Filtering and Large Language Models. arXiv. 2024.
4. Liu Z, Zhang H, Dong K, Fang Y. Collaborative Cross-modal Fusion with Large Language Model for Recommendation. arXiv. 2024.
5. Zhao Z, Fan W, Li J, Liu Y, Mei X, Wang Y, et al. Recommender Systems in the Era of Large Language Models (LLMs). IEEE Trans Knowl Data Eng. 2024;36(11): 6889-6907.
6. Elmoghazy SS, Shouman MA, Elminir HK, Selim GEI. Comparative analysis of methodologies and approaches in recommender systems utilizing large language models. Artif Intell Rev. 2025;58:207.
7. Liu Y, et al. A Comprehensive Survey on LLM-Powered Recommender Systems. arXiv. 2024.
8. Munson J, Cuzzo T, Nesar S, Zosso D. A review of large language models and the recommendation task. Discov Artif Intell. 2025;5:203.
9. Wang Y, et al. Large-Scale Benchmarks and Architectures for Multimodal Recommendation. Expert Syst Appl. 2024.
10. Liu Q, Hu J, Xiao Y, Zhao X, Gao J, Wang W, et al. Multimodal Recommender Systems: A Survey. ACM Comput Surv. 2025;57(2):26.
11. Luo Y, Qin Q, Zhang H, Cheng M, Yan R, Wang K, et al. Molar: Multimodal LLMs with Collaborative Filtering Alignment for Enhanced Sequential Recommendation. arXiv. 2024.
12. Lopez-Avila A, Du J. A Survey on Large Language Models in Multimodal Recommender Systems. arXiv. 2025.
13. Xu J, Chen Z, Yang S, Li J, Wang W, Hu X, et al. A Survey on Multimodal Recommender Systems: Recent Advances and Future Directions. arXiv. 2025.
14. Wang Z, Liu B, Huang W, Hao T, Zhou H, Guo Y. Leveraging multimodal large language model for multimodal sequential recommendation. Sci Rep. 2025;15:28960.
15. Smith A, et al. Deep Learning Models for Recommendation with Text and Visual Content. Inf Syst. 2024.
16. Li J, Yang C, Ye G, Nguyen QVH. Graph neural networks with deep mutual learning for designing multi-modal recommendation systems. Inf Sci. 2024;654:119815.
17. Chen Z, et al. Advances in Graph-Based Recommendation Models. ACM Trans Recomm Syst. 2024.
18. Guo L, et al. Graph Neural Networks with Multimodal Fusion for Recommendation Systems. Inf Fusion. 2024.
19. Zhao M, et al. Hybrid Recommendation Systems: Recent Trends and Challenges. SN Comput Sci. 2024.
20. Raj S, et al. Latent Feature Integration and Optimization in Recommender Systems. Sci Rep. 2025.
21. Kim J, Seo Y, Kim Y, Kang J, Shin J, Trirat P, et al. Large language models are zero-shot point-of-interest recommenders. Data Min Knowl Disc. 2025;39(6):80.