A Hybrid Framework for E-Commerce Recommendations in Emerging Markets: Leveraging Sentiment Analysis and LLM Embeddings.

Ashutosh Ahuja
Independent Researcher
West Hartford, Connecticut, United States
ashutosh.ahuja11@gmail.com

Mandakini Gupta
Independent Researcher
West Hartford, Connecticut, United States
gupta.mandakini@gmail.com

ABSTRACT

Product search and recommendation systems are critical for enhancing customer experiences on e-commerce platforms, driving engagement, and increasing sales. Traditional approaches, such as collaborative filtering (CF) and content-based filtering (CBF), often struggle to capture the semantic richness of customer reviews and product descriptions, limiting their ability to provide personalized suggestions. In this study, we propose a novel hybrid recommendation framework that leverages Large Language Models (LLMs) to integrate semantic understanding and sentiment analysis into the recommendation process. Using a simulated e-commerce dataset of ~27000 customer reviews, we demonstrate that combining LLM-generated embeddings with collaborative filtering significantly improves recommendation accuracy and relevance. Our framework addresses key challenges like data sparsity and noisy feedback, showcasing the transformative potential of LLMs in advancing e-commerce personalization. This work highlights the benefits of combining semantic insights with traditional recommendation methods to create scalable and context-aware systems, with potential applications in emerging markets where diverse consumer behavior and data sparsity pose significant challenges.

CCS CONCEPTS

• Computing Methodologies →Natural Language Processing

KEYWORDS

Customer Behavior Analysis, Semantic Understanding, Collaborating efforts, Product Recommendation, and Machine Learning.

WWW Companion '25, April 28-May 2, 2025, Sydney, NSW, Australia

ACM 979-8-4007-1331-6/25/04
https://doi.org/10.1145/3701716.3717856

ACM Reference format:

Ashutosh Ahuja and Mandakini Gupta. 2025. A Hybrid Framework for E-Commerce Recommendations in Emerging Markets: Leveraging Sentiment Analysis and LLM Embeddings. In *Proceedings of WWW Companion '25, April 28-May 2, 2025, Sydney, NSW, Australia, 6 pages.* https://doi.org/10.1145/3701716.3717856

1 Introduction

In e-commerce, product recommendation systems are crucial for enhancing user engagement and driving sales. E-commerce platforms in emerging markets, such as India, face unique challenges like diverse customer preferences, data sparsity, and rapid digital adoption. However, traditional algorithms, such as collaborative filtering (CF) and content-based filtering (CBF), often fail to capture the nuanced preferences of customers. These methods rely on sparse or noisy data and typically ignore unstructured feedback, such as customer reviews and sentiments, which play a critical role in influencing purchasing decisions. Recent advancements in Large Language Models (LLMs), such as RoBERTa and DistilBERT, have shown their ability to generate semantic representations of text, opening new possibilities for improving recommendation systems. In this paper, we propose a hybrid recommendation framework that integrates LLM-generated embeddings with collaborative filtering to enhance product recommendations on e-commerce platforms. Our framework incorporates sentiment analysis of customer reviews, enabling a deeper understanding of customer preferences and contextual relevance. By leveraging the semantic capabilities of LLMs, our system addresses common challenges such as data sparsity and cold-start problems, which often hinder the performance of traditional models. Using a simulated e-commerce dataset comprising ~27,000 reviews, we evaluate the hybrid framework through metrics such as Precision@K and Recall@K. The simulated dataset was designed to mimic real-world e-commerce platforms, capturing key challenges such as data sparsity and noisy feedback, while maintaining manageable complexity for experimentation. Specifically, our framework achieves a 7.69% improvement in Precision@5 and a 13.3% improvement in Recall@5 over traditional CF methods. This study highlights the potential of combining LLMs and collaborative filtering to create scalable, personalized, and context-aware recommendation systems for real-world e-commerce applications, with particular relevance to emerging markets.

2 Methodology

Our proposed framework is a hybrid recommendation system that combines Large Language Models (LLMs) with collaborative filtering (CF) to enhance product recommendations. The system integrates semantic embeddings generated by LLMs and sentiment analysis of customer reviews to provide personalized and context-aware recommendations. The framework operates in three main stages:

Data Processing: Customer reviews and user-item interactions are processed to extract sentiment and generate semantic embeddings.
Recommendation Generation: The outputs of collaborative filtering and LLM-based embeddings are combined using a weighted average.
Evaluation: The system is evaluated using metrics such as Precision@K, Recall@K, and Mean Reciprocal Rank (MRR).

2.1 Key Components

2.1.1 Sentiment Analysis: Customer reviews are analyzed to extract sentiment (positive, negative, or neutral) using a fine-tuned sentiment analysis model. This allows the system to prioritize products with positive feedback and deprioritize those with negative reviews.

2.1.2 Sentiment Embeddings: Semantic embeddings for product descriptions and reviews are generated using pre-trained LLMs (e.g., RoBERTa, DistilBERT). These embeddings capture the semantic richness of textual data, enabling the system to understand nuanced relationships between products.

2.1.3 Collaborative Filtering: A matrix factorization-based CF algorithm is used to identify product similarities based on user-item interactions. This component ensures that the system leverages historical user behavior to make recommendations.

2.1.4 Hybrid Recommendation Generation: The outputs of CF and LLM-based recommendations are combined using a weighted average. The weights are optimized to balance the contributions of each component, ensuring that the system leverages both user-item interaction data and semantic insights from customer reviews. A dynamic weighting mechanism could further optimize the balance between CF and LLM embeddings. Future work may explore adaptive learning-based approaches to adjust these weights in real-time based on user interaction patterns.

2.2 Novelty of the Approach

2.2.1 Integration of Sentiment Analysis: By incorporating sentiment analysis, the system captures nuanced customer preferences that traditional CF methods often miss.

2.2.2 LLM-Generated Embeddings: The use of LLMs to generate semantic embeddings allows the system to understand contextual relationships between products, addressing challenges such as data sparsity and cold-start problems.

2.3 Recommendation Algorithm

The recommendations are generated by combining the outputs of collaborative filtering (CF) and LLM-based semantic embeddings using a weighted average approach. These models were chosen for their proven performance in text representation tasks and their suitability for understanding nuanced e-commerce data. The CF component identifies product similarities based on user-item interactions, while the LLM embeddings provide semantic and contextual relevance derived from customer feedback. The weights for combining these components are optimized through grid search on the validation set, with the best-performing weights being 0.6 for LLM embeddings and 0.4 for CF scores. This balanced approach allows the system to address key challenges such as data sparsity, noisy reviews, and cold-start problems, ensuring that recommendations are both personalized and context-aware.

3 Dataset and Experimental Set up

3.1 Dataset and Statistics

We utilized a custom dataset designed to simulate a real-world online retail environment. The dataset incorporates user reviews, product descriptions, and user-item interaction data to support the development and evaluation of our hybrid recommendation framework. It comprises 27,000 simulated customer reviews covering approximately 1,800 unique products and 900 users. The dataset was generated to reflect the structure and diversity of real-world e-commerce environments, ensuring a balance across product categories and sentiment distributions. Given the resource constraints of this study and the potential privacy concerns associated with publicly available datasets containing user-generated content, we opted for a simulated dataset. This approach allowed us to focus on specific research objectives while ensuring cost-effectiveness and alignment with ethical considerations. Each review is labeled with sentiment (positive, neutral, or negative) derived through LLM-based sentiment analysis. The dataset contains approximately ~360,000 words, with an average review length of 13.3 words. The sentiment distribution is maintained at 45% positive, 30% neutral, and 25% negative, corresponding to approximately 12,150 positive, 8100 neutral and 6750 negative reviews. The dataset was constructed using a combination of manually curated seed reviews and synthetic augmentation techniques. To ensure diversity in sentiment and product coverage, we applied paraphrasing and controlled augmentation methods to generate additional reviews while maintaining linguistic variability. Sentiment labels were assigned using a fine-tuned sentiment analysis model to ensure consistency across the dataset.

Dataset Source and Relevance: The dataset was designed to simulate a real-world online retail environment, capturing key challenges such as data sparsity, noisy feedback, and cold-start problems. This makes it an ideal testbed for evaluating hybrid recommendation systems. While the dataset of 27,000 reviews provides initial meaningful insights, future work could explore larger, publicly available datasets, such as the

Amazon Product Data, to further validate the framework's scalability and applicability to diverse real-world scenarios.

To better illustrate the structure of the dataset, below table presents a sample of the data used in this study. Each entry includes a user identifier, product id, customer review text, numerical rating, sentimental label, and the corresponding embedding vectors. This format provides the dataset captures both structured and unstructured information, providing a comprehensive analysis of user preference.

User	**Prod.**	**Review Text**	**Rating (Sentiment)**	**Embeddings**
U001	P001	"Terrible quality, broke within..".	1(Negative)	[0.32, 0.45,]
U002	P002	"Average product, decent for.."	3 (Neutral)	[-0.22, 0.14]
U003	P003	"This coffee maker is fantastic."	5 (Positive)	[0.32, 0.45]

Table 1: Example of the custom e-commerce dataset, including user reviews, product ratings, and derived sentiment labels.

3.2 Data Preprocessing

To prepare the dataset for modeling, the following preprocessing steps were applied:

Sentiment Analysis: Each customer review was analyzed using a sentiment model, labeling reviews as positive, neutral, or negative. The model was fine-tuned with a learning rate of 0.00005, a batch size of 16, and a sequence length of 128 tokens. The sentiment scores were used to weight product recommendations, prioritizing items with positive feedback and deprioritizing those with negative reviews.

Embedding Generation: Semantic embeddings for product descriptions and reviews were generated using pre-trained RoBERTa and DistilBERT models, fine-tuned on our dataset. The embeddings capture 512-dimensional representations of product descriptions and reviews, enabling the system to understand nuanced relationships between products.

Dataset Splits: The dataset was divided into training (80%), validation (10%), and testing (10%) sets to ensure robust evaluation.

Text Cleaning: Review text and product descriptions were cleaned by removing special characters, stop words, and irrelevant tokens.

Normalization: User ratings were normalized to a common scale (0 to 1) for compatibility with collaborative filtering algorithms.

3.3 Baselines

The hybrid recommendation framework was evaluated through a two-stage experimental setup. The baseline model relied on collaborative filtering, using only user-item interaction data without incorporating contextual inputs such as customer reviews or product descriptions. The CF model was trained on 80% of the dataset, with the remaining 20% used for testing, and its performance was assessed using standard evaluation metrics, including Precision@5, Recall@5, and MRR. In contrast, the LLM-enhanced model integrated sentiment analysis and semantic embeddings derived from customer reviews with the CF output. The embeddings were generated using pre-trained RoBERTa and DistilBERT models, fine-tuned on the e-commerce dataset with a learning rate of 0.00005, a batch size of 16, and a sequence length of 128 tokens. Fine-tuning was performed over three epochs, with early stopping applied to prevent overfitting. The hybrid framework combined the outputs of collaborative filtering and LLM embeddings using a weighted average approach, with weights optimized based on validation performance. The final configuration assigned a weight of 0.6 to LLM embeddings and 0.4 to collaborative filtering scores, allowing the system to effectively address challenges such as data sparsity and the cold-start problem. The collaborative filtering baseline in this study is implemented using matrix factorization (ALS-based CF), a widely used approach in recommendation systems. This ensures a robust baseline for comparison against the hybrid model.

3.4 Evaluation Metrics

3.4.1 Precision@K: Measures the proportion of relevant recommendations in the top K results. The metric reflects the system's ability to provide accurate suggestions.

3.4.2 Recall@K: Measures the system's ability to retrieve relevant products from the entire dataset. This metric is particularly important for assessing the framework's coverage of user preferences.

3.4.3 Mean Reciprocal Rank (MRR): Evaluates the ranking quality of the first relevant recommendation. This metric is useful for understanding how well the system prioritizes the most relevant items.

These metrics are computed using 5-fold cross-validation to ensure the robustness of our results. Additionally, we compare the performance of our hybrid framework to baseline models (e.g., traditional CF and content-based filtering) to demonstrate its effectiveness.

4 Results and Evaluation

4.1 Baseline and LLM Enhanced Results

The baseline collaborative filtering (CF) model, which relies solely on user-item interaction data, was evaluated using Precision@5, Recall@5, and Mean Reciprocal Rank (MRR). These metrics assess the relevance, ranking quality, and retrieval capability of the recommendations, respectively. The baseline model achieves a Precision@5 of 0.65, a Recall@5 of 0.60, and an MRR of 0.55, as shown in Table 2. These results reflect the limitations of traditional CF systems, such as data sparsity and noisy feedback, which hinder their ability to provide accurate and context-aware recommendations. For instance, the Recall@5 value of 0.60 highlights the model's difficulty in retrieving relevant recommendations consistently, while the MRR of 0.55 indicates suboptimal ranking quality. To address these limitations, we integrated pre-trained LLM embeddings and applied sentiment analysis to customer reviews. These enhancements allowed the hybrid system to incorporate semantic understanding and contextual relevance into the recommendations. The LLM-enhanced models achieved notable improvements across all evaluation metrics, as shown in Table 2 and Figure 1. For instance, Precision@5 increased from 0.65 in the baseline to 0.70 for RoBERTa (fine-tuned), while Recall@5 improved from 0.60 to 0.68. The integration of LLM embeddings notably enriched the recommendation process by addressing challenges such as data sparsity and noisy feedback. Among the LLM variants evaluated, RoBERTa (ft) achieved the highest scores across all metrics, demonstrating the effectiveness of combining semantic insights with traditional recommendation methods.

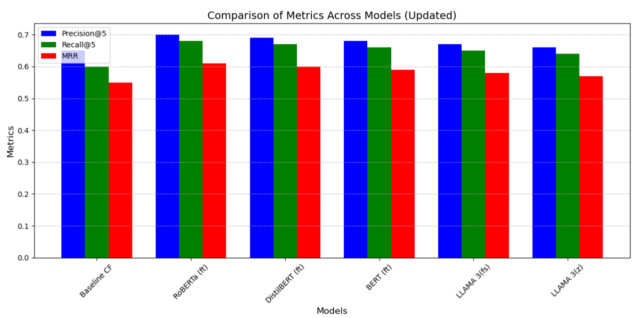

Fig 1: Comparison of Baseline CF and Enhanced LLM models across three evaluations metrics.

Model	Precision@5	Recall@5	MRR
Baseline CF	0.65	0.60	0.55
RoBERTa (ft)	0.70	0.68	0.61
DistilBERT (ft)	0.69	0.67	0.60
BERT (ft)	0.68	0.66	0.59
LLAMA 3(fs)	0.67	0.65	0.58
LLAMA 3(z)	0.66	0.64	0.57

Table 2: Performance comparison of the baseline CF model and LLM enhanced Models.

4.3 Comparison

The performance comparison between the baseline CF model and the LLM-enhanced system is illustrated in Table 3 and Figure 2. The hybrid model achieved notable gains, with percentage improvements of 7.69% in Precision@5, 13.3% in Recall@5, and 10.9% in MRR. These results demonstrate the transformative potential of LLM-based embeddings and sentiment analysis in enhancing recommendation quality.

Metric	Baseline CF	RoBERTa (ft)	Improvement %
Precision@5	0.65	**0.70**	+7.69%
Recall@5	0.60	**0.68**	+13.3%
MRR	0.55	**0.61**	+10.9%

Table 3: Comparison of Baseline CF and Best performing RoBERTa (ft) model across three evaluation metrics.

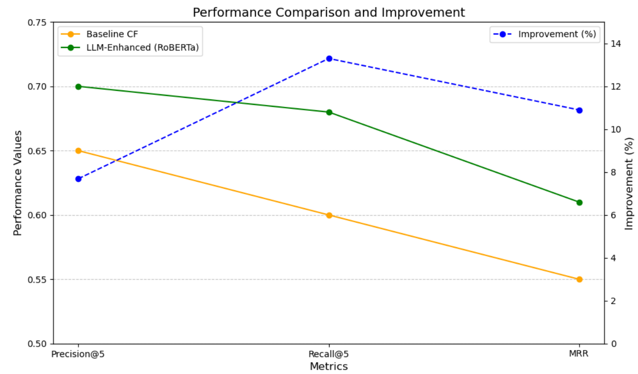

Fig 2: Visualizes these improvements across metrics, highlighting the hybrid framework's ability to better capture customer preferences and provide more context-aware recommendations.

5 Analysis and Discussion

The integration of Large Language Models (LLMs) into collaborative filtering systems resulted in substantial improvements in recommendation performance. The LLM-enhanced system, particularly the fine-tuned RoBERTa model, consistently outperformed the baseline collaborative filtering model across all evaluation metrics. For instance, Precision@5 improved from 0.65 in the baseline model to 0.70 with the LLM-enhanced system, representing a 7.69% increase. Similarly, Recall@5 and MRR demonstrated gains of 13.3% and 10.9%, respectively, highlighting the hybrid framework's ability to address challenges such as data sparsity and noisy feedback. These metrics underscore the value of combining LLM embeddings with collaborative filtering to create more effective and personalized recommendation systems. One of the most significant contributions of the hybrid system was its ability to incorporate sentiment analysis into the recommendation process. By analyzing customer reviews, the system captured nuanced feedback, enabling it to better align recommendations with customer preferences. While not directly measured, the integration of sentiment analysis and LLM embeddings is expected to mitigate cold-start issues by leveraging textual data. For example, products with negative sentiments were deprioritized in favor of alternatives with more favorable reviews, leading to more context-aware suggestions. This approach was particularly effective in improving Recall@5, as it allowed the system to identify a broader range of relevant recommendations that traditional collaborative filtering methods often missed. Among the LLMs evaluated, RoBERTa (ft) achieved the highest scores across all metrics, outperforming both DistilBERT and BERT. RoBERTa's ability to generate detailed semantic embeddings was instrumental in capturing the complex relationships between customer reviews and product descriptions, contributing to its superior performance. DistilBERT, while slightly less accurate, offered a balance between computational efficiency and effectiveness, making it a viable option for real-time systems with limited resources. Despite these advancements, the study also identified several limitations. The computational cost associated with generating embeddings from large pre-trained models, such as RoBERTa, poses scalability challenges for smaller e-commerce platforms. Furthermore, the fine-tuning process requires substantial labeled data and computational resources, which may not be readily available in all scenarios. Addressing these limitations in future work could involve exploring lightweight LLM architectures or efficient fine-tuning methods.To balance recommendation accuracy and efficiency, this study already includes DistilBERT as a lightweight alternative to larger LLMs. However, further optimizations, such as quantization techniques or distillation-based training, could enhance computational efficiency without significant loss in performance.

In practical terms, the hybrid framework offers significant advantages for e-commerce platforms. By integrating semantic understanding and sentiment analysis, the system caters to diverse customer preferences, mitigates the cold-start problem, and enhances overall user satisfaction. These findings highlight the transformative potential of LLM-based systems in advancing personalized recommendations and provide a strong foundation for future innovations in the field.

6 Conclusion

This paper demonstrates the effectiveness of integrating LLMs with collaborative filtering to enhance product recommendations in e-commerce. By incorporating semantic embeddings and sentiment analysis, the proposed framework achieves notable improvements and addressing key challenges outlined earlier. These findings highlight the ability of hybrid recommendation systems to deliver more personalized and context-aware user experiences, paving the way for scalable advancements in recommendation technologies. The integration of semantic understanding and sentiment analysis not only improves recommendation accuracy but also enables e-commerce platforms to better align with customer preferences, ultimately driving user engagement and satisfaction. Looking ahead, future work could explore the application of this framework in emerging markets, such as India, where diverse consumer behavior and rapid digital adoption present unique challenges for e-commerce recommendation systems. Additionally, extending the framework to other domains, such as content recommendation or social media, could further demonstrate its versatility and impact. To further validate the generalizability of this approach, future work will explore evaluations on large-scale public datasets such as Amazon Reviews or Yelp. By continuing to innovate in this space, we can unlock new possibilities for intelligent, user-centric recommendation systems that adapt to the evolving needs of consumers and businesses alike.

REFERENCES

[1] Bellar O, Baina A, Ballafkih M. Sentiment Analysis: Predicting Product Reviews for E-Commerce Recommendations Using Deep Learning and Transformers. Mathematics. 2024; 12(15):2403. https://doi.org/10.3390/math12152403

[2] Kim, S., Kang, H., Choi, S., Kim, D., Yang, M., & Park, C. (2024, August). 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 (pp. 1395-1406).

[3] Zhou, X., Zhao, X., & Li, G. (2024). LLM-Enhanced Data Management. arXiv preprint arXiv:2402.02643.

[4] Liao, W., Zeng, B., Yin, X., & Wei, P. (2021). An improved aspect-category sentiment analysis model for text sentiment analysis based on RoBERTa. Applied Intelligence, 51, 3522-3533.

[5] K. L. Tan, C. P. Lee, K. S. M. Anbananthen and K. M. Lim, "RoBERTa-LSTM: A Hybrid Model for Sentiment Analysis With Transformer and Recurrent Neural Network," in IEEE Access, vol. 10, pp. 21517-21525, 2022, doi: 10.1109/ACCESS.2022.3152828.

[6] Kulvinder Singh, Dhawan, S. & Bali, N. An Ensemble Learning Hybrid Recommendation System Using Content-Based, Collaborative Filtering, Supervised Learning and Boosting Algorithms. Aut. Control Comp. Sci. 58, 491–505 (2024). https://doi.org/10.3103/S0146411624700615

[7] S. S. Roy, A. Kumar and R. Suresh Kumar, "Metadata and Review-Based Hybrid Apparel Recommendation System Using Cascaded Large Language Models," in IEEE Access, vol. 12, pp. 140053-140071, 2024, doi: 10.1109/ACCESS.2024.3462793.

[8] Ahuja, A., & Gupta, M. (2024). Optimizing Predictive Maintenance with Machine Learning And Iot: A Business Strategy For Reducing Downtime And Operational Costs.

[9] Mandalapu, S.R., Narayanan, B. & Putheti, S. A hybrid collaborative filtering mechanism for product recommendation system. Multimed Tools Appl 83, 12775–12798 (2024). https://doi.org/10.1007/s11042-023-16056-8

[10] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert:Pre-training of deep bidirectional transformers for language understanding.

[11] Abdalla, H.B., Gheisari, M. & Awlla, A.H. Hybrid self-attention BiLSTM and incentive learning-based collaborative filtering for e-commerce recommendation systems. Electron Commer Res (2024). https://doi.org/10.1007/s10660-024-09888-5Conference Name:ACM Woodstock conference

[12] Liu, Q., Wu, X., Wang, Y., Zhang, Z., Tian, F., Zheng, Y., & Zhao, X. (2024, May). LLM-ESR: Large Language Models Enhancement for Long-tailed Sequential Recommendation. In The Thirty-eighth Annual Conference on Neural Information Processing Systems.

[13] K. Bao, J. Zhang, Y. Zhang, W. Wang, F. Feng, and X. He. Tallrec: An effective and efficient tuning framework to align large language model with recommendation. In Proceedings of the 17th ACM Conference on Recommender Systems.

[14] Zhang, W., Deng, Y., Liu, B., Pan, S. J., & Bing, L. (2023). Sentiment analysis in the era of large language models: A reality check. arXiv preprint arXiv:2305.15005.

[15] Mughal, N., Mujtaba, G., Kumar, A., & Daudpota, S. M. (2024). Comparative Analysis of Deep Natural Networks and Large Language Models for Aspect-Based Sentiment Analysis. IEEE Access.

[16] Zhou, C., Song, D., Tian, Y., Wu, Z., Wang, H., Zhang, X., ... & Zhang, S. (2024). A Comprehensive Evaluation of Large Language Models on Aspect-Based Sentiment Analysis. arXiv preprint arXiv:2412.02279.

[17] Conference Location:El Paso, Texas USAYang, H., Zhao, Y., Wu, Y., Wang, S., Zheng, T., Zhang, H., ... & Qin, B. (2024). Large language models meet text-centric multimodal sentiment analysis: A survey. arXiv preprint arXiv:2406.08068.

[18] Kumawat, S., Yadav, I., Pahal, N., & Goel, D. (2021, January). Sentiment analysis using language models: A study. In 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 984-988). IEEE.

[19] Törnberg, P. (2023). How to use Large Language Models for Text Analysis. arXiv preprint arXiv:2307.13106.

[20] Shaik Vadla, M. K., Suresh, M. A., & Viswanathan, V. K. (2024). Enhancing product design through AI-driven sentiment analysis of Amazon reviews using BERT. Algorithms, 17(2), 59.

[21] Wu, L., Zheng, Z., Qiu, Z., Wang, H., Gu, H., Shen, T., ... & Chen, E. (2024). A survey on large language models for recommendation. World Wide Web, 27(5), 60.

[22] Ahuja, A., & Analytics, P. (2022). REVOLUTIONIZING CLAIM ADJUDICATION: DESIGNING INTELLIGENT, PANDEMIC-RESILIENT CONTACT CENTER SYSTEMS IN HEALTHCARE TECHNOLOGY. Well Testing Journal, 31(1), 64-83.

[23] Zhao, X., & Sun, Y. (2022). Amazon fine food reviews with BERT model. Procedia Computer Science, 208, 401-406.

[24] Praveen, S. V., Gajjar, P., Ray, R. K., & Dutt, A. (2024). Crafting clarity: Leveraging large language models to decode consumer reviews. Journal of Retailing and Consumer Services, 81, 103975.

[25] Xu, X., Xu, Z., Ling, Z., Jin, Z., & Du, S. (2024). Emerging Synergies Between Large Language Models and Machine Learning in Ecommerce Recommendations. arXiv preprint arXiv:2403.02760.

[26] Ghatora, P. S., Hosseini, S. E., Pervez, S., Iqbal, M. J., & Shaukat, N. (2024). Sentiment Analysis of Product Reviews Using Machine Learning and Pre-Trained LLM. Big Data and Cognitive Computing, 8(12), 199.

[27] Assiri, A., Gumaei, A., Mehmood, F., Abbas, T., & Ullah, S. (2024). DeBERTa-GRU: Sentiment Analysis for Large Language Model. Computers, Materials & Continua, 79(3).

[28] Carneros-Prado, D., Villa, L., Johnson, E., Dobrescu, C. C., Barragán, A., & García-Martínez, B. (2023, November). Comparative study of large language models as emotion and sentiment analysis systems: A case-specific analysis of GPT vs. IBM Watson. In International Conference on Ubiquitous Computing and Ambient Intelligence (pp. 229-239). Cham: Springer Nature Switzerland.

[29] Chumakov, S., Kovantsev, A., & Surikov, A. (2023). Generative approach to aspect based sentiment analysis with GPT language models. Procedia Computer Science, 229, 284-293.

[30] Ahuja, A. (2022). Revolutionizing Claim Adjudication Designing Intelligent. Pandemic-Resilient Contact Center systems in Healthcare Technology, 10.

[31] Baert, G., Gahbiche, S., Gadek, G., & Pauchet, A. (2020, December). Arabizi language models for sentiment analysis. In Proceedings of the 28th international conference on computational linguistics (pp. 592-603).

[32] Shah, F. A., Sabir, A., & Sharma, R. (2024). A Fine-grained Sentiment Analysis of App Reviews using Large Language Models: An Evaluation Study. arXiv preprint arXiv:2409.07162.

[33] Roumeliotis, K. I., Tselikas, N. D., & Nasiopoulos, D. K. (2024). Leveraging Large Language Models in Tourism: A Comparative Study of the Latest GPT Omni Models and BERT NLP for Customer Review Classification and Sentiment Analysis. Information, 15(12), 792.

[34] Liu, J., Liu, C., Zhou, P., Ye, Q., Chong, D., Zhou, K., ... & Yu, P. S. (2023). Llmrec: Benchmarking large language models on recommendation task. arXiv preprint arXiv:2308.12241.

[35] Rong, D., Yao, L., Zheng, Y., Yu, S., Xu, X., Liu, M., & Wang, Z. (2024, December). LLM Enhanced Representation for Cold Start Service Recommendation. In International Conference on Service-Oriented Computing (pp. 153-167). Singapore: Springer Nature Singapore.