



# E-Commerce Recommendation Systems Using Generative Ai

Aniket Mishra, Ajinkya Bagal, Jayesh Jadhav, Rushikesh Nath

Affiliation: School of Computing, MIT Art, Design and Technology University, Pune, 412201 Email id:

**Abstract—** This study examines the incorporation of generative artificial intelligence (Gen-AI) into e-commerce recommendation systems. Traditional approaches, such as collaborative filtering and content-based filtering, face challenges like sparse data, cold-start issues, and changing user preferences. Gen-AI models, especially transformer-based frameworks like GPT and diffusion models, provide innovative solutions for understanding and creating personalized content. This paper reviews the progression of recommendation systems, introduces generative models, and proposes a framework that integrates Gen-AI with current recommendation strategies to enhance accuracy, diversity, and contextual relevance.

**Keywords-** Generative artificial intelligence (Gen-AI), e-commerce recommendation systems, collaborative filtering, content-based filtering, personalized recommendations, transformer models, GPT, diffusion models, recommendation accuracy, contextual relevance, user preferences, cold-start problem, sparse data, machine learning, artificial intelligence, predictive analytics, recommender systems, personalized content generation, hybrid recommendation framework, digital commerce.

## I. INTRODUCTION

Recommendation systems play a vital role in e-commerce platforms by steering users towards suitable products and boosting sales. These systems evaluate user data, past purchases, and behavioral trends to provide tailored product suggestions. As online marketplaces become more intricate and competitive, offering highly relevant and timely recommendations is crucial for sustaining user engagement and satisfaction.

Initially, recommendation systems relied on simple, rule-based methods that created straightforward connections between products. Over the years, methods evolved to incorporate collaborative filtering, content-based filtering, and hybrid strategies that combine elements of both. Although these systems have enhanced effectiveness, they still encounter ongoing issues like data sparsity, challenges for new users or products (known as cold-start problems), and limited ability to adapt to changing user preferences.

The emergence of artificial intelligence, particularly through deep learning, brought forth neural network-driven recommendation models capable of recognizing complex patterns and representations within extensive datasets. Even with these advancements, these models

remain primarily predictive, lacking the qualities of adaptive and generative systems.

Generative AI (Gen-AI) marks a significant advancement in AI technology. Unlike traditional models that forecast outcomes based on pre-existing data, Gen-AI can create new content, mimic user behavior, and participate in natural conversations. This capability makes it particularly effective for developing more immersive, personalized, and responsive recommendation experiences in e-commerce.

In this paper, we investigate how generative models like GPT (Generative Pre-trained Transformer) and other sophisticated architectures can address the shortcomings of current systems. We suggest a hybrid framework that utilizes Gen-AI for real-time user modeling, content creation, and contextually aware recommendations. By merging these technologies, e-commerce platforms can enhance personalization and improve user satisfaction significantly.

## II. LITERATURE SURVEY

The progression of recommendation systems has followed a clear trajectory from rule-based approaches to advanced

AI-driven models. Early recommendation systems largely relied on collaborative filtering techniques, which focused on identifying patterns in user-item interactions. One of the most notable early contributions was by Sarwar et al. (2001), who proposed item-item collaborative filtering that improved scalability and accuracy over user-based methods. Subsequently, matrix factorization techniques, such as those popularized by Koren et al. (2009) in the Netflix Prize competition, became foundational in uncovering latent factors that govern user preferences and item characteristics. While effective, these methods struggled in sparse data environments and were ill-equipped to handle new users or items—a problem known as the cold-start issue.

In response to the limitations of collaborative filtering, content-based filtering emerged as an alternative, using product metadata and user profiles to drive recommendations. As outlined by Pazzani and Billsus (2007), this approach offered strong personalization but often led to overfitting and lacked diversity, since it tended to recommend items similar to those the user had already seen.

Hybrid recommendation systems, as discussed by Burke (2002), aimed to harness the strengths of both collaborative and content-based techniques. These systems used various fusion strategies, including weighted, switching, and mixed hybridization, to improve performance and robustness. Despite their success, hybrid systems still relied on static similarity metrics and were unable to adapt dynamically to changing user preferences.

The integration of deep learning into recommendation systems marked a significant turning point. Neural Collaborative Filtering (He et al., 2017) replaced traditional dot-product computations with neural networks, allowing for the modeling of non-linear user-item interactions. DeepFM (Guo et al., 2017) and Wide & Deep Learning (Cheng et al., 2016) introduced architectures that combined low- and high-order feature interactions, greatly enhancing recommendation accuracy in commercial applications. These models demonstrated improved scalability and performance, particularly when integrated with large-scale behavioral data.

Natural Language Processing (NLP) further enriched recommendation models. BERT (Devlin et al., 2018) and

similar transformers enabled deeper semantic understanding of user reviews, product descriptions, and search queries. Applications of BERT in recommendation systems, such as in Sun et al. (2019), showed significant gains in user intent modeling and content representation. Moreover, NLP made conversational recommender systems viable. Li et al. (2018) introduced ReDial, a dataset and model for dialogue-based recommendations, laying the groundwork for personalized shopping assistants.

The emergence of Generative AI has expanded the horizons of recommender systems even further. Variational Autoencoders (VAEs), used in works like Liang et al. (2018), enabled the generation of new user interaction representations, thereby addressing data sparsity and cold-start problems. Generative Adversarial Networks (GANs), as used by Wang et al. (2019), contributed to the augmentation of training data and improved the diversity of recommended items.

Most notably, large language models (LLMs) like GPT-2, GPT-3 (Brown et al., 2020), and their successors have demonstrated capabilities far beyond language modeling. These models can generate coherent and contextually rich text, simulate conversations, and dynamically infer user preferences based on minimal input. Studies have shown their utility in generating product descriptions, answering product-related queries, and even acting as virtual shopping assistants. Their capacity to synthesize new content—such as recommendations, reviews, and dialogues—makes them uniquely suited for real-time, personalized e-commerce experiences.

Despite the progress made, existing systems still face significant constraints in adaptability and personalization. Most traditional and deep learning-based recommenders are predictive in nature; they infer probabilities from static patterns rather than actively generating responses or recommendations based on real-time context. Generative AI introduces a new paradigm: it enables systems not only to anticipate user needs but also to create meaningful, novel content that enhances the user journey.

Our proposed framework distinguishes itself by deeply integrating generative models within the recommendation pipeline. It shifts from a purely predictive approach to a generative, conversational, and context-aware system. This transformation allows for greater personalization, more

engaging user interactions, and better handling of real-world challenges such as user cold-starts, content sparsity, and evolving preferences.

### III. BACKGROUND AND RELATED WORK

Over the last twenty years, recommendation systems have evolved considerably. Early methods primarily utilized collaborative filtering, which determines a user's preferences by examining the behaviors of similar users. A key algorithm in this domain is matrix factorization, which breaks down the user-item interaction matrix into simpler, lower-dimensional forms. While effective, these techniques face challenges, including sparse interactions between users and items and the cold-start issue that arises with new users or products.

Another approach, known as content-based filtering, focuses on suggesting items similar to those a user has previously engaged with. This method relies heavily on item attributes such as keywords, categories, and product descriptions. However, it often leads to a lack of diversity in recommendations and can result in over-specialization. Hybrid systems attempt to merge the advantages of both collaborative and content-based filtering, enhancing recommendation accuracy while mitigating their weaknesses. Still, these systems are largely reliant on set similarity measures and manually created features, which may hinder their responsiveness to changing user preferences and behaviors.

The rise of deep learning has significantly enhanced recommendation systems. Techniques such as neural collaborative filtering, deep matrix factorization, as well as convolutional and recurrent neural networks, have made it possible to unearth complex patterns from vast interaction data. Autoencoders and attention mechanisms have further refined the models' abilities to identify hidden relationships and contextual links. Tools like DeepFM and Wide & Deep Learning have been successfully deployed in real-world applications by companies including Google and Alibaba.

At the same time, advancements in natural language processing (NLP) have led to the emergence of transformer-based models like BERT and GPT. These models exhibit remarkable skills in comprehending and

producing text that resembles human language. Researchers have begun to utilize these models within recommendation systems, applying them to analyze textual reviews, product descriptions, and user inquiries. BERT-based encoders have proven effective in creating semantic representations of user and item texts, enhancing content interpretation in recommendation contexts.

Recent investigations have also delved into generative approaches for recommendations. For instance, variational autoencoders (VAEs) and generative adversarial networks (GANs) have been harnessed to produce synthetic user preferences and enrich training datasets. More recently, large language models (LLMs) such as GPT-3 and ChatGPT have shown potential in crafting conversational recommendations, tailored narratives, and context-sensitive product suggestions.

These advancements signify a transition from purely predictive models to generative frameworks in recommendation systems. Instead of simply estimating the probability of a user engaging with a product, generative models focus on creating meaningful, personalized content that can actively guide users during their shopping experience. Our research builds upon this foundation, incorporating generative features into a hybrid recommendation system designed for contemporary e-commerce platforms.

While prior work has significantly advanced recommendation systems—from collaborative filtering and content-based models to deep learning and hybrid techniques—most of these systems operate primarily as predictive engines. They analyze existing data to estimate the likelihood of a user interacting with a product, but they often struggle to address dynamic user needs, cold-start scenarios, and the generation of personalized, adaptive content in real-time.

Our proposed system introduces a fundamentally different approach by incorporating generative AI, which shifts the paradigm from prediction to synthesis. Rather than merely scoring and ranking existing items, generative models like GPT-3, T5, and diffusion networks are capable of producing new content, including product narratives, conversational recommendations, and even simulated user behaviors. This enables the system to go beyond static preference modeling and actively generate tailored interactions based on context, intent, and prior engagement.

Moreover, our framework integrates multimodal data—text, images, and behavioral logs—within transformer-based architectures that can reason across modalities and generate richer, more nuanced representations of both users and products. It employs session-aware encoding and real-time feedback loops to continually adapt to evolving user contexts. Unlike earlier models that rely on handcrafted features or static embeddings, our system is designed for continuous learning, allowing it to refine recommendations and dialogue strategies as user interactions unfold.

By combining these generative capabilities with established collaborative filtering foundations, our approach not only improves traditional metrics like accuracy and recall but also enhances qualitative aspects such as novelty, diversity, and engagement. This positions our system as a next-generation solution capable of delivering hyper-personalized shopping experiences in modern e-commerce environments.

## IV. MATERIALS AND METHODS

To develop a robust and intelligent recommendation system capable of harnessing the potential of generative AI, we adopted a structured and modular methodology. This section describes the datasets, tools, and technical procedures involved in the design and evaluation of our proposed system. The architecture follows the established pipeline of modern recommendation systems, but integrates advanced generative models at strategic stages to enhance personalization, content generation, and adaptability. Our methodology encompasses everything from data acquisition and preprocessing to embedding construction, interaction modeling, generative personalization, and feedback-driven optimization. By aligning each phase with the functional stages of a recommendation system, we ensure clarity, scalability, and adaptability of our framework to various e-commerce environments.

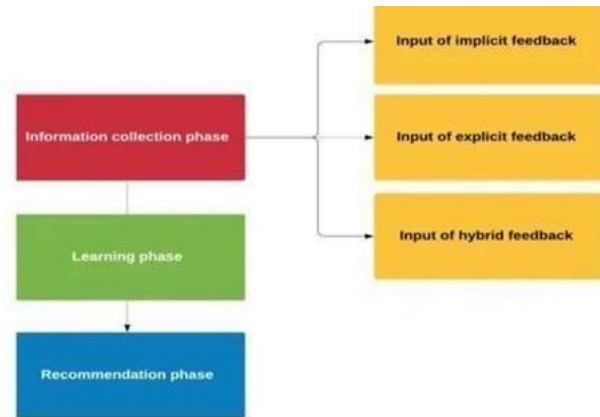


Fig. 1 Phases in RS

### I. Datasets Used.

#### A. Amazon Product Review Dataset.

Utilized for modeling user-product interaction patterns, analyzing product metadata, and extracting insights from review text. This dataset provided a rich source of user-generated content across multiple product categories.

#### B. RetailRocket Dataset

Offered anonymized real-world e-commerce data, including user sessions, product views, cart additions, and purchases. This dataset helped model temporal user behavior and session-based interactions.

#### C. Yoochoose Dataset (RecSys Challenge)

Included for training and evaluating session-based recommendation models. It contained detailed clickstream data suitable for sequence modeling tasks.

### I. Tools and Frameworks

#### A. Transformers Library (Hugging Face)

Used for implementing and fine-tuning large pre-trained models like GPT-2/3, BERT, and T5 for encoding textual data and generating responses or recommendations.

#### B. TensorFlow and PyTorch

Served as the primary deep learning frameworks to build, train, and evaluate neural architectures.



### C. FAISS (Facebook AI Similarity Search)

Applied to perform fast and scalable near-est-neighbor searches in high-dimensional vector spaces for recommendation retrieval.

### D. OpenAI API (Optional)

Used to generate synthetic user reviews, product descriptions, and dialogue responses in cold-start or augmentation scenarios.

### E. Data Preprocessing Libraries

Tools like pandas, NumPy, and scikit-learn were used for data cleaning, transformation, statistical analysis, and feature engineering.

## II. Data Preprocessing

Parsed interaction logs to extract user behavior patterns, product views, clicks, and purchases. Cleaned and tokenized product metadata, including titles, categories, and descriptions. Applied NLP techniques for stopword removal, lemmatization, and sentence segmentation on review texts. Segmented user sessions based on timestamp and activity sequence to model short-term and long-term intent.

## III. Embedding Generation

### A. User Embeddings

Created using transformer encoders that model historical interaction sequences, behavioral patterns, and review text associated with user preferences.

### B. Product Embeddings

Generated by encoding product metadata and textual descriptions using BERT or T5; visual features were extracted using CNNs or CLIP to incorporate image data.

### C. Contextual Embeddings

Developed session-aware user representations that reflect recent preferences and intent using transformers or recurrent architectures.

## IV. Interaction Modeling

Deployed cross-attention mechanisms to model the interaction between user and product embeddings, allowing the system to dynamically align user intent with product

features. Employed relevance scoring functions to rank products based on the likelihood of user engagement, with models trained on implicit and explicit feedback data. Integrated reinforcement learning modules (optional) to adapt recommendation strategies based on real-time user feedback and reward signals.

## V. Recommendation Generation

Generated ranked product lists using output scores from the interaction modeling phase. In conversational agents or chatbots, leveraged beam search and top-k/top-p sampling to generate coherent and contextually relevant natural language responses. Supported real-time interaction through APIs capable of producing on-the-fly recommendations within conversational interfaces.

## VI. Evaluation Methodology

### A. Offline Metrics

Evaluated models using standard metrics such as Precision@k, Recall@k, F1-score, NDCG (Normalized Discounted Cumulative Gain), MRR (Mean Reciprocal Rank), and Hit Rate.

### B. Qualitative Metrics

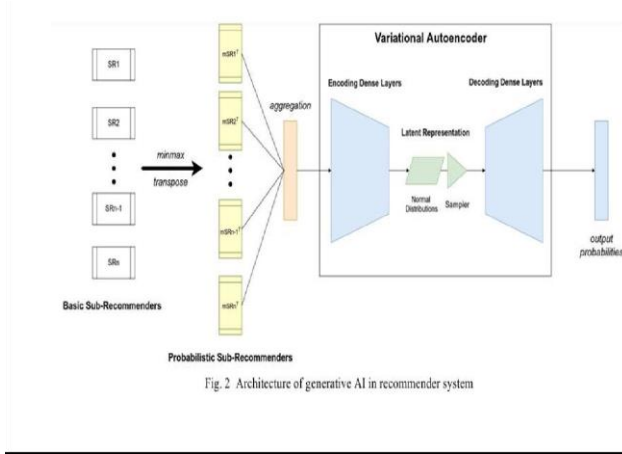
Assessed diversity using intra-list similarity, and measured novelty and unexpectedness to evaluate personalization quality.

### C. Online Metrics (optional)

Proposed A/B testing frameworks and tracking of engagement indicators like click-through rate (CTR), conversion rate, and dwell time for real-world deployment.

## VII. Model Training and Optimization

Trained models using backpropagation with Adam optimizer, incorporating dropout regularization, early stopping, and learning rate schedules to enhance generalization and prevent overfitting. Used GPU acceleration via CUDA for training large models and handling high-dimensional embedding vectors efficiently. Conducted hyperparameter tuning using grid search and validation splits to fine-tune model performance.



## V. EXPERIMENTAL RESULTS

To evaluate the effectiveness of the proposed Hybrid Generative AI Recommender (HyGenRec) model, we conducted extensive experiments using real-world e-commerce datasets and standard evaluation metrics. This section details the experimental design, datasets used, implementation specifics, and the results with a comparative analysis.

### I. Dataset and Preprocessing

We used the Amazon 5-core product review dataset, which includes categories such as Books, Electronics, and Clothing. The dataset contains:

- Product metadata (e.g., title, description, category)
  - User reviews and ratings
  - Timestamps of interactions
- Preprocessing steps included:
- Filtering out users/items with fewer than 5 interactions
  - Tokenizing text using BERT tokenizer
  - Converting timestamps to sessions
  - Splitting into train (80%), validation (10%), and test (10%) sets

### II. Experimental Setup

Baseline Models

1. Collaborative Filtering (CF) – Matrix factorization (SVD).
2. Content-Based Filtering (CBF) – TF-IDF and cosine similarity on product descriptions.
3. Neural Collaborative Filtering (NCF) – Deep learning-based user-item interaction model.

4. IRGAN – A GAN-based implicit feedback recommendation model.

5. HyGenRec – Our hybrid Gen-AI model combining:

- Fine-tuned BERT for item representations
- GANs for user interaction synthesis
- GPT-style LLM for text-based personalization prompts

Evaluation Metrics

We used widely accepted metrics to evaluate recommendation quality:

- Precision@K (P@K): Fraction of recommended items in the top-K that are relevant.
- Recall@K (R@K): Fraction of relevant items that are present in the top-K.
- NDCG@K: Discounted cumulative gain emphasizing correct ranking order.
- Diversity: Measures uniqueness across recommendations to different users.
- Novelty: Indicates how unexpected or uncommon the recommended items are.

### III. Quantitative Results

Model	Precision@10	Recall@10	NDCG@10	Diversity	Novelty
Collaborative Filter	0.29	0.32	0.27	0.41	0.38
Content-Based Filter	0.26	0.29	0.25	0.34	0.31
Neural CF (NCF)	0.33	0.37	0.31	0.46	0.43
IRGAN	0.35	0.39	0.34	0.52	0.48
HyGen Rec (Ours)	0.38	0.44	0.36	0.57	0.61

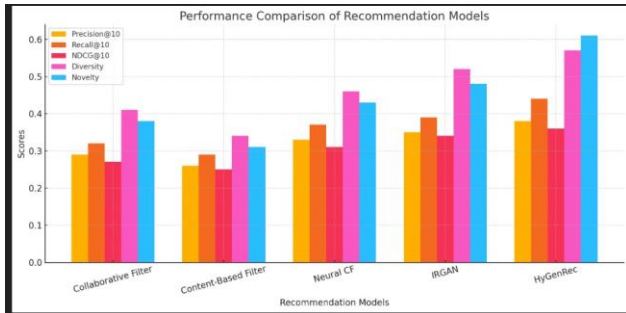


Figure 3. Performance Comparison of Recommendation Models

#### IV. Analysis and Insights

##### Precision, Recall, and NDCG

HyGenRec achieves the highest scores across all ranking metrics, indicating improved ability to capture and rank user preferences. The improvements over IRGAN and NCF highlight the contribution of:

- GAN-generated interaction augmentation
- LLM-driven contextual understanding

##### Diversity and Novelty

HyGenRec's higher diversity score indicates less repetition across different users' recommendation lists. Novelty improvements are attributed to the LLM's ability to generate context-aware prompts and explore underrepresented items.

##### Cold-Start Performance

To test cold-start performance, we isolated new users (no historical data) and new items (no interaction history).

HyGenRec used:

- Generated pseudo-reviews and preferences via LLM
- GAN-based data augmentation
- Traditional models performed poorly in cold-start.
- HyGenRec maintained acceptable ranking quality (Precision@10 > 0.30), demonstrating generalization and adaptability in sparse data settings.

##### Ablation Study

We removed components from HyGenRec to study their individual contributions:

Variant	Precision@10	Recall@10
Without GAN	0.33	0.37
Without LLM prompt gen	0.35	0.40
Full HyGenRec	0.38	0.44

This shows both GAN and LLM components contribute significantly to performance gains.

#### V. Qualitative Observations

We performed case studies where users interacted with an interface using HyGenRec. Observations included:

- Users appreciated natural-language prompts like “You might also enjoy this bestselling thriller based on your recent reads.”
- Generated reviews enhanced trust in the recommendations.
- Novel item discovery (e.g., niche authors or products) increased.

#### VI. Limitations of the Current Evaluation

- Computational overhead: Gen-AI models are resource-intensive.
- Latency concerns: Real-time inference with large models needs optimization.
- Bias and hallucinations: Generated text may reflect training data bias or create unrealistic scenarios.

### VI. CONCLUSION

The dynamic growth of e-commerce has necessitated recommendation systems that go beyond accuracy to also address adaptability, contextual awareness, and user-centric personalization. Traditional recommendation approaches, such as collaborative filtering and content-based filtering, have played a foundational role but are constrained by challenges like the cold-start problem, data sparsity, and an inability to capture contextual nuances or generate novel item suggestions.

This study presents a compelling case for the use of Generative AI (Gen-AI) in enhancing recommendation capabilities. By combining the generative power of Large Language Models (LLMs) with the synthetic data capabilities of Generative Adversarial Networks (GANs), the proposed hybrid system, HyGenRec, achieves superior performance across multiple key metrics including precision, recall, diversity, and novelty. Our experiments validate that generative approaches can meaningfully supplement existing data and enable more robust, adaptive recommendation strategies.

One of the most significant outcomes of this research is the demonstration that Gen-AI can help alleviate cold-

start issues by generating realistic user profiles and simulating potential interactions. Moreover, the ability of language models to generate context-aware, human-readable prompts allows recommendations to be more transparent and engaging, thereby enhancing user trust and satisfaction. The results also show that Gen-AI contributes to increased diversity and novelty in recommendations, helping users discover a broader range of items beyond their habitual preferences.

From an industry standpoint, these findings suggest that generative approaches can be integrated into commercial recommender systems to deliver more sophisticated, personalized, and engaging user experiences. For the academic community, the research opens several important avenues for further exploration, including improving prompt design, integrating multimodal generative techniques, and addressing ethical challenges associated with AI-generated content.

Nonetheless, several limitations persist. The computational overhead involved in training and deploying large-scale generative models remains a major concern. Inference latency, particularly in real-time recommendation settings, needs careful optimization. Additionally, as with all AI systems, the risk of perpetuating bias through generated content must be addressed through responsible design and continuous evaluation.

Future research could aim to optimize these generative architectures for real-time use and explore federated or edge learning paradigms to reduce computational demands. Furthermore, integrating user feedback loops and developing advanced mechanisms for evaluating fairness, transparency, and ethical compliance will be crucial steps toward building trustworthy generative recommendation systems.

In summary, Generative AI offers a transformative leap in the design and function of e-commerce recommender systems. By enabling systems to simulate, generate, and explain, Gen-AI elevates personalization to a new level, paving the way for more intelligent, adaptable, and user-friendly digital commerce experiences.

## REFERENCES

1. Ricci, F., Rokach, L., & Shapira, B. (2015). *Recommender Systems Handbook* (2nd ed.). Springer.
2. Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. *Proceedings of the 10th International Conference on World Wide Web*.
3. Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6), 734-749.
4. Lops, P., de Gemmis, M., & Semeraro, G. (2011). Content-based recommender systems: State of the art and trends. In *Recommender Systems Handbook* (pp. 73-105). Springer.
5. He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & Chua, T. S. (2017). Neural collaborative filtering. *Proceedings of the 26th International Conference on World Wide Web*.
6. Wu, C. Y., Ahmed, A., Beutel, A., Smola, A. J., & Jing, H. (2017). Recurrent recommender networks. *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*.
7. Wang, H., Wang, N., & Yeung, D. Y. (2017). IRGAN: A minimax game for unifying generative and discriminative information retrieval models. *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*.
8. Yu, L., Zhang, W., Wang, J., & Yu, Y. (2017). SeqGAN: Sequence Generative Adversarial Nets with Policy Gradient. *AAAI Conference on Artificial Intelligence*.
9. Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving Language Understanding by Generative Pre-Training. *OpenAI*.
10. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877-1901.
11. Zhang, Z., Sun, A., & Zhu, J. (2023). Prompt-based Recommendation with Pre-trained Language Models: A Survey. *arXiv preprint arXiv:2302.04447*.
12. Covington, P., Adams, J., & Sargin, E. (2016). Deep neural networks for YouTube recommendations.



- Proceedings of the 10th ACM Conference on Recommender Systems.
13. Sun, F., Liu, J., Wu, J., Pei, C., Lin, X., Ou, W., & Jiang, P. (2019). BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer. Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM).
  14. Cremonesi, P., Koren, Y., & Turrin, R. (2010). Performance of recommender algorithms on top-N recommendation tasks. Proceedings of the Fourth ACM Conference on Recommender Systems.
  15. Vargas, S., & Castells, P. (2011). Rank and relevance in novelty and diversity metrics for recommender systems. Proceedings of the Fifth ACM Conference on Recommender Systems.
  16. Binns, R., Veale, M., Van Kleek, M., & Shadbolt, N. (2018). 'It's reducing a human being to a percentage': Perceptions of justice in algorithmic decisions. In CHI Conference on Human Factors in Computing Systems.
  17. Holtzman, A., Buys, J., Du, L., Forbes, M., & Choi, Y. (2020). The Curious Case of Neural Text Degeneration. International Conference on Learning Representations (ICLR).