# Advancing Content-Based Recommendations: Integrating Large Language Models in Chatbots

A PROJECT REPORT

Submitted by

**Sri Karthik Avala**
**20MIA1032**

**Nikitha A R**
**20MIA1025**

**Akshitha Bachu**
**19MIA1096**

in partial fulfilment for the award of the degree of

Master of Technology

in

Business Analytics (5 Year Integrated Programme)

# CERTIFICATE

This is to certify that the report entitled Advancing Content-Based Recommendations: Integrating Large Language Models in Chatbots is prepared and submitted by Sri Karthik Avala (Reg. No. 20MIA1032), Nikitha A R (Reg. No. 20MIA1025) and Akshitha Bachu (Reg. No. 19MIA1096) to Vellore Institute of Technology, Chennai, in partial fulfilment of the requirement for the award of the degree of Master of Technology in Business Analytics (5 year Integrated Programme) and as part of CSE4077 –Recommender Systems Project is a bona-fide record carried out under my guidance. The project fulfils the requirements as per the regulations of this University and in my opinion meets the necessary standards for submission.

Guide/Supervisor

Name: Dr. Pradeep K

Date:

HoD

Name: Dr. Sivabalakrishnan

Date:

(Seal of SCOPE)

# ACKNOWLEDGEMENT

We would like to express our special thanks and gratitude to our teacher Prof Pradeep K , faculty of school of Computer Science and Engineering who gave us the golden opportunity to do this wonderful project titled **Advancing Content-Based Recommendations: Integrating Large Language Models in Chatbots**, and also for providing us with proper guidance and suggestions to develop this project. This helped us in doing a lot of research and we came to know about so many new things. Secondly, we would also like to thank our friends and group mates who helped us a lot in finalizing this project within the limited time frame. We would also like to thank our family for their consistent support throughout this time.

## Arrangement of contents

## Abstract

In the realm of conversational AI and recommendation systems, leveraging Large Language Models (LLMs) has become pivotal for enhancing user engagement and satisfaction. This study explores the integration of the llama3 model, powered by ollama, within a chatbot framework to improve content-based recommendations.

The primary focus is on utilizing LLMs, specifically llama3 with Retrieve-and-Generate (RAG) capabilities, to convert structured JSON data into meaningful embeddings. These embeddings serve as foundational elements for both answering user queries and generating personalized recommendations.

Key contributions of this research include:

- **Embedding Generation:** Utilizing llama3's capabilities to transform JSON data into embeddings that capture semantic relationships and context.
- **Recommendation System:** Enhancing the chatbot's ability to provide relevant and personalized recommendations based on user interactions and preferences.
- **User Engagement:** Improving the overall user experience by delivering more accurate and context-aware responses.

Through empirical evaluation and implementation, this work demonstrates the effectiveness of integrating LLMs in enhancing content-based recommendations within chatbot environments. Future directions may involve optimizing retrieval and generation processes to further refine recommendation accuracy and responsiveness.

This research underscores the transformative potential of integrating advanced language models in enhancing AI-driven conversational systems, paving the way for more sophisticated and user-centric applications in recommendation technology.

# Introduction

Large language models (LLMs) are a big leap in recommendation systems, now being integrated into chatbots to improve content-based recommendations. These models use their understanding of language to suggest things that match what users are interested in. By analyzing how users talk, chatbots with LLMs can offer more relevant suggestions. This makes recommendations more accurate and makes chatting with these bots more useful and enjoyable. This study explores how LLMs in chatbots enhance recommendations, showing how this technology can make personalized suggestions better and change how recommendations are done.

# Literature Review

Paper-1

**Large Language Models as Data Augmenters for Cold-Start Item Recommendation**

"Large Language Models as Data Augmenters for Cold-Start Item Recommendation" by Jianling Wang, Haokai Lu, James Caverlee, Ed Chi, and Minmin Chen, the authors explore the significant enhancement of cold-start item recommendations through the integration of Large Language Models (LLMs). LLMs are shown to improve reasoning and inference capabilities for user preferences, providing augmented training signals that boost recommendation performance. The study evaluates models such as NeuMF and SASRec for their generalizability, demonstrating that LLMs can generate new features and improve encoding within recommendation systems. This integration not only enhances efficiency and accuracy but also addresses the common issue of data sparsity in cold-start scenarios, where items lack sufficient interaction data for accurate embeddings. The authors propose using pairwise BPR loss, ID embedding, and meta features for cold-start items, along with pairwise comparison prompts to better infer user preferences. Despite challenges such as slow API calls and data sparsity, the method shows improved recommendation effectiveness without added computational overhead. Experiments on real-world datasets validate these performance enhancements, highlighting the potential of LLM-generated data to fill the data gap for cold-start items and improve user preference understanding. This approach underscores the transformative impact of LLMs in enhancing cold-start item recommendations by generating and integrating augmented training signals effectively.

Paper-2

## LLM-Enhanced User-Item Interactions: Leveraging Edge Information for Optimized Recommendations

"LLM-Enhanced User-Item Interactions: Leveraging Edge Information for Optimized Recommendations" explores the integration of large language models (LLMs) with graph neural networks (GNNs) in recommendation systems to enhance accuracy and personalization. Traditional systems relying on collaborative filtering and content-based methods often struggle with utilizing edge information in graphs effectively. LLMs like GPT-2 and ChatGPT excel in natural language processing, enhancing recommendations by capturing intricate user preferences. However, LLMs face challenges in leveraging graph edges. Conversely, GNNs effectively model graph structures, offering insights into user-item relationships. By combining LLMs with GNNs, researchers aim to enhance recommendation systems by utilizing graph structures for more personalized recommendations. The proposed framework focuses on improving LLMs' understanding of graph relationships through innovative prompt construction mechanisms, resulting in better recommendation relevance and personalization. This integration presents a novel approach to optimizing recommendations and suggests new research directions.

Paper-3

## Rethinking Large Language Model Architectures for Sequential Recommendations

"Rethinking Large Language Model Architectures for Sequential Recommendations" provides a comprehensive overview of the challenges and advancements in sequential recommendation systems. Traditional models often struggle with the computational complexity of decoding algorithms like beam search, which can hinder their efficiency and scalability. The introduction of Lite-LLM4Rec addresses these issues by proposing a more streamlined architecture that reduces computational overhead and improves inference efficiency. By leveraging a hierarchical LLM structure and item projection head, Lite-LLM4Rec achieves significant performance improvements without sacrificing efficiency. This model represents a paradigm shift in the design of large language models for sequential recommendations, offering a more sustainable and effective solution for handling extensive context information in recommendation tasks. The contributions of Lite-LLM4Rec extend beyond its technical innovations, offering practical implications for improving the scalability and efficiency of real-world recommendation systems. Overall, Lite-LLM4Rec sets a new benchmark for LLM architectures in sequential recommendation, showcasing the potential for future advancements in this field.

**Paper-4**

### Rec-GPT4V: Multimodal Recommendation with Large Vision-Language Models

"Rec-GPT4V: Multimodal Recommendation with Large Vision-Language Models" explores the landscape of multimodal recommendation systems (MMRSs), focusing on the integration of large vision-language models (LVLMs) to address the challenges in understanding both images and text for personalized recommendations. Traditional recommendation systems have mainly relied on textual data, but with the increasing availability of visual content, there is a growing need for systems that can effectively process and understand multimodal information. LVLMs, such as Rec-GPT4V, offer a promising solution by leveraging their temporal understanding and static image interpretation capabilities to provide more comprehensive recommendations. LVLMs, while powerful, face challenges in understanding user preferences and handling multiple image dynamics. The VST reasoning scheme proposed by Rec-GPT4V aims to overcome these limitations by distilling information from images and utilizing user history and image summaries for personalized recommendations. This approach not only enhances the performance of LVLMs in multimodal recommendation scenarios but also improves the overall user experience by providing more relevant and engaging recommendations. The paper contributes to the field by introducing the VST strategy and the Rec-GPT4V model, which leverage LVLMs for multimodal recommendations effectively. By proposing innovative reasoning strategies and addressing the limitations of LVLMs in handling multimodal data, the paper opens up new possibilities for enhancing recommendation systems. The empirical evaluation of Rec-GPT4V on real-world datasets validates its effectiveness in improving recommendation quality and user engagement, highlighting the practical implications of integrating LVLMs into multimodal recommendation systems.

Paper-5

### Chat-REC: Towards Interactive and Explainable LLMs-Augmented Recommender System

The paper "Chat-REC: Towards Interactive and Explainable LLMs-Augmented Recommender System" introduces the Chat-Rec framework, which aims to improve top-k recommendations with normalized discounted cumulative gain (NDCG) scores. The framework utilizes large language models (LLMs) to enhance user preference learning and facilitate cross-domain recommendations. It addresses the limitations of existing recommender systems, such as poor interactivity, explainability, and feedback mechanisms. Chat-Rec optimizes the candidate set for movie recommendations and evaluates its performance on real-world datasets for recommendation and rating tasks. The framework uses prompts to convert user interactions into recommendations, offering solutions for natural, explainable recommendations and addressing challenges in making recommendations across multiple domains. The Chat-Rec framework is particularly innovative in its approach to enhancing recommender systems. Traditional recommender systems often lack interactivity and struggle with providing explanations for their recommendations, leading to a lack of user trust and engagement. Chat-Rec addresses these challenges by leveraging LLMs to provide natural and explainable recommendations. By optimizing the candidate set for movie recommendations and utilizing prompts to convert user interactions into recommendations, Chat-Rec enhances user preference learning and facilitates cross-domain recommendations. This approach not only improves the quality of recommendations but also enhances user engagement by providing explanations for why certain recommendations are made. Additionally, the framework's focus on addressing the cold start problem for new items and users further demonstrates its practical implications for improving the overall user experience in recommender systems.

Paper-6

### LLM-Enhanced User-Item Interactions: Leveraging Edge Information for Optimized Recommendations

The paper "LLM-Enhanced User-Item Interactions: Leveraging Edge Information for Optimized Recommendations" presents a method that enhances recommendation systems by leveraging large language models (LLMs) and graph neural networks (GNNs) to improve recommendation accuracy and personalization. The method focuses on utilizing LLMs to re-rank items, providing insights into diversity without requiring fine-tuning. By integrating LLMs with GNNs, the method enhances the understanding of graph relationships, leading to improved recommendation quality. Various metrics are used to evaluate the effectiveness of the method, including Jaccard similarity, cosine similarity, and distance in a taxonomy. Comparisons with existing methods demonstrate the superiority of the proposed approach in re-ranking candidate documents. The study also investigates the reasoning ability and structured answers of LLMs, correlating them with task difficulty. The paper highlights the significance of recommendation systems in providing diverse and meaningful recommendations to users, emphasizing the need for diverse recommendations to facilitate user choice. It discusses the challenges faced by traditional recommendation systems in generating diverse recommendations, particularly in cold-start scenarios where there is limited data available for new items. LLMs show promise in enhancing recommendation diversity through re-ranking, but their performance is not yet superior to traditional methods. The paper introduces the LLM-InS model, which addresses these challenges by improving cold-start item recommendation using large language models. The model leverages a Hierarchical Interaction Simulator to mimic user interactions, enabling it to train on both cold and warm items effectively. This simulation-based approach is crucial for generating accurate behavioral embeddings for cold items, allowing the model to make relevant and personalized recommendations even for items with limited historical data. Overall, the paper provides valuable insights into enhancing recommendation diversity using LLMs and sets a foundation for further research in this area.

Paper-7

### Enhancing Recommendation Diversity by Re-ranking with Large Language Models

The paper "Enhancing Recommendation Diversity by Re-ranking with Large Language Models" explores the use of large language models (LLMs) for diversity re-ranking in recommender systems. LLMs are used to re-rank items, providing insights into diversity without the need for fine-tuning. The study investigates the capabilities of LLMs, such as reasoning ability and structured answers, correlating them with task difficulty. Various similarity metrics, including Jaccard similarity, cosine similarity, and distance in a taxonomy, are employed to evaluate the effectiveness of LLMs in re-ranking candidate documents. The paper compares the proposed approach with existing methods and explores the capabilities of models like ChatGPT and GPT-4 in re-ranking.The importance of diverse recommendations in recommender systems for meaningful user choice is highlighted, emphasizing the role of LLMs in enhancing diversity re-ranking. However, the paper notes that while LLMs show promise, they are not yet superior to traditional methods in re-ranking. The study also finds that traditional re-ranking methods are faster and less resource-demanding than LLM-based approaches. Overall, the paper contributes to the understanding of how LLMs can be leveraged to improve recommendation diversity, providing insights into their capabilities and limitations in comparison to traditional methods.

Paper-8

### Integrating Large Language Models into Recommendation via Mutual Augmentation and Adaptive Aggregation

The paper "Integrating Large Language Models into Recommendation via Mutual Augmentation and Adaptive Aggregation" introduces the Llama4Rec model, which enhances recommendation performance by integrating large language models (LLMs) with conventional recommendation models. Llama4Rec outperforms baseline methods in recommendation performance on real-world datasets, with an ablation study showing that all components contribute significantly to the overall performance improvement. The model leverages strengths of both LLMs and conventional models, addressing the lack of generalizability in current methods and the computational inefficiency of LLMs. Llama4Rec achieves this by combining collaborative and sequential information, using data and prompt augmentation strategies tailored for recommendation models, and employing an adaptive aggregation module for refined recommendation results. Empirical studies validate Llama4Rec's superiority over baseline methods consistently, demonstrating its potential for enhancing recommendation performance. The integration of large language models (LLMs) into recommendation systems represents a significant advancement in the field, offering new possibilities for improving recommendation accuracy and relevance. Traditional recommendation approaches often struggle with cold-start scenarios and lack the ability to effectively leverage textual data for understanding user preferences. LLMs, on the other hand, excel in natural language understanding tasks and can capture complex user-item interactions from textual data. By integrating LLMs with conventional recommendation models, such as matrix factorization and collaborative filtering, Llama4Rec combines the strengths of both approaches, mitigating the weaknesses of each. This integration allows Llama4Rec to refine recommendation results through data and prompt augmentation, enhancing model performance by leveraging collaborative and sequential information effectively.One of the key innovations of Llama4Rec is its adaptive aggregation module, which refines recommendations by merging predictions from different models. This module adapts to the characteristics of each recommendation scenario, dynamically adjusting the aggregation strategy to achieve optimal results. Additionally, Llama4Rec's data and prompt augmentation strategies are specifically designed for recommendation models, ensuring that the model can effectively leverage the rich information contained in textual data for improving recommendation accuracy.

Paper-9

**Large Language Model Interaction Simulator for Cold-Start Item Recommendation**

The paper "Large Language Model Interaction Simulator for Cold-Start Item Recommendation" introduces the LLM-InS model, which aims to improve cold-start item recommendation using large language models (LLMs). The model outperforms existing methods in this area, showing significant improvements in performance metrics. LLM-InS utilizes a Hierarchical Interaction Simulator to mimic user interactions, allowing for the training of both cold and warm items. By simulating user behavior patterns, the model enhances recommendation capabilities for both types of items, addressing challenges in generating accurate behavioral embeddings for cold items. Additionally, the paper reviews existing cold-start recommendation models and their limitations, highlighting the impact of embedding logic differences and simple NLP models on cold item recommendation. LLM-InS represents a significant contribution to the field by offering a new approach to cold-start item recommendation and demonstrating the effectiveness of using simulated interactions for training recommendation models. The LLM-InS model is particularly innovative in its approach to cold-start item recommendation. Traditional recommendation systems often struggle with cold-start scenarios, where there is limited or no historical data available for new items. LLM-InS addresses this challenge by simulating user interactions, allowing it to train on both cold and warm items effectively. This simulation-based approach is crucial for generating accurate behavioral embeddings for cold items, enabling the model to make relevant and personalized recommendations even for items with limited historical data. By integrating natural language understanding perspectives into the recommendation process, LLM-InS offers a novel and effective solution to the cold-start problem, showcasing the potential of large language models in enhancing recommendation systems. The model's ability to outperform existing methods underscores its practical implications for improving recommendation accuracy and relevance, especially in scenarios with limited data availability for new items.

Paper-10

**LLM-Enhanced User-Item Interactions: Leveraging Edge Information for Optimized Recommendations**

In the paper "LLM-Enhanced User-Item Interactions: Leveraging Edge Information for Optimized Recommendations" by Xinyuan Wang, Liang Wu, Liangjie Hong, Hao Liu, and Yanjie Fu, the authors investigate the integration of Large Language Models (LLMs) with graph neural networks to enhance recommendation systems. This approach aims to improve the relevance and quality of recommendation results by leveraging the relationship mining capabilities of graph neural networks and the deep representation and generative logic of LLMs. The paper introduces a novel prompt construction framework that enables LLMs to understand graph data, addressing the challenge of effectively exploiting edge information in graphs, which existing methods often overlook. By focusing on connectivity information in graph data, the proposed graph attentive LLM system enhances both recommendation accuracy and personalization. Evaluations on real-world datasets demonstrate significant improvements in recommendation relevance, highlighting the practical implications of this integration. The framework also incorporates new prompting methods and attention mechanisms, significantly boosting model performance. This innovative approach opens new paths for intelligent recommendation systems by providing a new technological pathway for applying LLMs across various fields. Overall, the research showcases how combining LLMs with graph neural networks can offer a more nuanced understanding of user-item relationships, leading to optimized recommendations and personalized user experiences.

Paper -11

**On Generative Agents in Recommendation**

In the paper "On Generative Agents in Recommendation" by An Zhang, Yuxin Chen, Leheng Sheng, Xiang Wang, and Tat-Seng Chua, the authors explore the use of Large Language Models (LLMs) to empower generative agents within recommendation systems. These agents, equipped with profile, memory, and action modules, consistently and impressively identify items aligned with user preferences, maintaining high accuracy and recall regardless of the number of items. The memory operations, including retrieval, writing, and reflection, enable dynamic interactions with the environment. The paper introduces Agent4Rec, a simulator designed to evaluate the effectiveness of these generative agents in recommendation scenarios. However, Agent4Rec primarily utilizes offline datasets lacking detailed item descriptions and operates within a limited action space, excluding factors like social networks and advertising. These limitations, along with potential hallucinations in LLMs, can lead to inaccurate simulation outcomes. The review enhances the understanding of LLM-empowered generative agents in recommendation systems, identifying critical limitations and challenges for future exploration. The authors acknowledge the need for a wider action space to better simulate user decisions and recognize the constraints of using offline datasets. Despite these challenges, the research on Agent4Rec bridges the gap between academic research and real-world recommendation deployments, exploring the filter bubble effect and causal relationships in recommendation tasks. The simulator replicates the filter bubble effect and delves into causal discovery, highlighting the potential of generative agents to align with user preferences while recognizing their limitations. Overall, the paper contributes to the innovative application of generative agents in recommendation systems, focusing on simulating user behavior and preferences for personalized recommendations, and paves the way for future advancements in the field.

Paper-12

## LLM-based Federated Recommendation

In the paper "LLM-based Federated Recommendation" by Jujia Zhao, Wenjie Wang, Chen Xu, Zhaochun Ren, See-Kiong Ng, and Tat-Seng Chua, the authors explore the significant potential of Large Language Models (LLMs) in enhancing recommendation systems, particularly for cold-start problems. The research demonstrates that LLMs improve performance by enhancing contextual understanding and leveraging global knowledge, with larger models excelling in reasoning and inferring user preferences. The augmentation of training signals using LLMs effectively boosts cold-start recommendation performance by providing additional context and filling data gaps. Federated learning frameworks such as FedAvg and FedProx have been employed to deploy LLMs in a distributed manner, though these methods often result in performance imbalances and high communication costs due to the large number of parameters. To address these challenges, the authors propose the Privacy-Preserving LLM-based Recommendation (PPLR) framework, which achieves balanced client performance and resource efficiency through dynamic parameter aggregation and flexible storage strategies, outperforming traditional federated learning methods. The paper highlights a shift in focus towards trustworthiness elements in LLM-based recommendation systems, including fairness, robustness, and explainability, while noting that privacy preservation remains underexplored. The PPLR framework introduces privacy-preserving tasks that balance client performance and enhance data privacy, addressing significant challenges such as client performance imbalances and high communication costs. The framework implements efficient parameter aggregation methods, providing an equitable and resource-efficient approach for LLM-based recommendations. Future enhancements of the PPLR framework include fine-grained aggregation and broader adaptation to various recommendation scenarios. Overall, the research demonstrates that LLMs significantly improve the performance and efficiency of recommendation systems, particularly in federated learning environments, with the PPLR framework addressing critical challenges and paving the way for continued optimization and expanded applicability in real-world scenarios.

Paper-13

## SPAR: Personalized Content-Based Recommendation via Long Engagement Attention

In the paper "SPAR: Personalized Content-Based Recommendation via Long Engagement Attention" explores the current landscape of content-based recommendation systems, highlighting the limitations of existing methods and the need for more advanced techniques. Content-based recommendation systems rely on text-based features for matching user preferences with item characteristics. While these systems have shown effectiveness in certain scenarios, they often struggle with processing long user text and capturing complex user-item interactions. Additionally, existing methods lack fine-grained token-level signals, which are crucial for understanding the nuances of user preferences and interactions. Recent advancements in large language models (LLMs) have shown promise in enhancing user profiling and engagement prediction in recommendation systems. LLMs are capable of extracting global interests from user histories, providing a more holistic understanding of user preferences. However, integrating LLMs into recommendation systems poses challenges, such as the need for efficient processing of long user engagement histories and the optimization of user-content interaction dynamics. The SPAR framework addresses these challenges by leveraging LLMs and poly-attention mechanisms for personalized content-based recommendations. SPAR enhances user profiling by extracting global interests from user histories and improves user-item interaction dynamics. By integrating session-based PLM encoding and lightweight attention layers, SPAR achieves state-of-the-art performance in content-based recommendation, outperforming existing methods in accuracy and engagement prediction. Overall, the literature review emphasizes the importance of advanced techniques like LLMs and attention mechanisms in overcoming the limitations of traditional content-based recommendation systems. SPAR's approach offers insights into potential trade-offs in designing recommendation systems and sets a new benchmark in content-based recommendation accuracy and user engagement prediction.

Paper-14

### Amazon-M2: A Multilingual Multi-locale Shopping Session Dataset for Recommendation and Text Generation

In the paper "Amazon-M2: A Multilingual Multi-locale Shopping Session Dataset for Recommendation and Text Generation" delves into the current landscape of session-based recommendation systems and critiques existing datasets. It underlines the necessity for tailored language models in recommendation tasks, particularly in addressing the challenges of data sparsity and diversity. Session-based recommendation systems are vital in e-commerce, offering personalized suggestions based on users' browsing behaviors within a session. While collaborative filtering (CF) and content-based recommendation methods are common, they struggle with issues like the cold start problem and data scarcity. The rise of large language models (LLMs) presents a promising avenue for improvement, given their capabilities in natural language processing (NLP) and reasoning. However, existing session datasets are often limited in scope, focusing on narrow domains and lacking diversity in linguistic and cultural aspects, which hinders the development of more effective recommendation algorithms. The introduction of the Amazon-M2 dataset aims to bridge these gaps by providing a rich, multilingual, and multi-locale dataset for session-based recommendation and text generation tasks. This dataset enables the development of personalized recommendation and text generation strategies that can accommodate language and location variations, thereby enhancing the overall performance of recommendation systems and advancing research in session-based recommendations.

Paper-15

**GenRec: Large Language Model for Generative Recommendation**

In this paper "GenRec: Large Language Model for Generative Recommendation" by Jianchao Ji, Zelong Li, Shuyuan Xu, Wenyue Hua, Yingqiang Ge, Juntao Tan, and Yongfeng Zhang from Rutgers University introduces GenRec, a model that leverages Large Language Models (LLMs) for generative recommendation using text data. The paper demonstrates significant improvement in recommendation tasks on large datasets, using metrics like Hit Ratio and NDCG for evaluation. GenRec utilizes specialized prompts for recommendation tasks and shows adaptability across diverse applications in recommendation systems. It offers personalized and contextually relevant recommendations, enhancing user experience. The paper encourages further research on LLMs for enhancing recommendation systems and proposes future work to refine prompts, incorporate user interaction data, and test the model further. GenRec represents a paradigm shift to generative recommendation from traditional discriminative methods, showcasing the potential of LLMs in recommendation systems.

Paper-16

**Leveraging Large Language Models in Conversational Recommender Systems**

"Leveraging Large Language Models in Conversational Recommender Systems" by Luke Friedman, Sameer Ahuja, David Allen, Zhenning Tan, Hakim Sidahmed, Changbo Long, Jun Xie, Gabriel Schubiner, Ajay Patel, Harsh Lara, Brian Chu, Zexi Chen, and Manoj Tiwari proposes a roadmap for large-scale Conversational Recommender Systems (CRS) using Large Language Models (LLMs) with new implementations. The paper introduces RecLLM, a large-scale CRS for YouTube videos, leveraging LLMs for user preference understanding, dialogue management, and explainable recommendations. It discusses the evolution of dialogue management from rule-based to model-based language generation and the integration of LLMs with external resources for better CRS performance. Challenges include limited conversational data for training large-scale CRS and ethical problems in recommender systems, which the paper aims to address. The proposed framework focuses on building a controllable and explainable large-scale conversational recommender system, with future work including human evaluations, public dataset release, and UI enhancements.

Paper-17

**LLM-Rec: Personalized Recommendation via Prompting Large Language Models**

"LLM-Rec: Personalized Recommendation via Prompting Large Language Models" by Hanjia Lyu, Song Jiang, Hanqing Zeng, Yinglong Xia, Qifan Wang, Si Zhang, Ren Chen, Christopher Leung, Jiajie Tang, and Jiebo Luo introduces LLM-REC, a framework that enhances recommendation quality through diverse prompting strategies for text augmentation. The framework outperforms complex methods, empowering simple MLP models and boosting performance significantly compared to other content-based methods. LLM-REC incorporates text enrichment strategies for personalized recommendations, including engagement-guided prompting and recommendation-driven prompting for text augmentation. The study compares LLM-REC with TagGPT and Knowledge Augmented Recommendation (KAR) methods. Challenges include the extra computational cost associated with the LLM-REC framework and promptly incorporating the latest knowledge for LLMs. Overall, LLM-REC enhances text-based recommendations using diverse prompting strategies, improving recommendation accuracy, and relevance for users.

Paper-18

### Exploring the Upper Limits of Text-Based Collaborative Filtering Using Large Language Models: Discoveries and Insights

In the paper"Exploring the Upper Limits of Text-Based Collaborative Filtering Using Large Language Models: Discoveries and Insights" explores the landscape of text-based collaborative filtering (TCF) using large language models (LMs). It acknowledges the significant advancements in language models within natural language processing (NLP), such as BERT, GPT series, and ChatGPT, and their widespread application in item recommendation tasks. The review identifies two primary research directions: item representation and user encoders, highlighting the complexity and importance of these components in TCF. Furthermore, it recognizes the limitations of the user-item feedback assumption in the recommendation task, emphasizing the need for novel approaches. The review also points out specific challenges faced by TCF, including the high computational cost for retraining LMs and issues related to recall and ranking with numerous candidate items in models like ChatGPT. Additionally, it notes the evolving nature of TCF models, indicating that while they challenge the traditional ID-based collaborative filtering paradigm, there is room for improvement and further advancements, especially in utilizing larger text encoders and addressing transferability challenges.

Paper-19

### PALR: Personalization Aware LLMs for Recommendation

Yang et al. introduce PALR, a framework that leverages Large Language Models (LLMs) for personalized recommendations, specifically excelling in sequential recommendation tasks. PALR surpasses state-of-the-art models by effectively re-ranking top recommendations from various retrieval methods, integrating user history with LLMs. The framework fine-tunes an LLM with 7 billion parameters for ranking purposes, incorporating natural language user profile generation and item recommendation. While LLMs enhance NLP research and recommender systems with vast knowledge, they may lack complete knowledge of newly released shopping items and could generate incomplete or hallucinatory results. PALR addresses these challenges by enhancing personalized recommendations using LLMs' reasoning abilities and effectively incorporating user behaviors to generate user-preferred items. Experimentation with MovieLens-1M and Amazon

Beauty datasets demonstrates PALR's superiority over existing models in sequential recommendation tasks, showcasing its potential for enhancing recommendation systems with LLMs. PALR's integration of user behaviors with LLMs allows for more personalized recommendations, enhancing the user experience. By breaking down tasks into user profile, retrieval, and ranking sub-tasks, PALR effectively manages the complexity of recommendation systems. The framework's ability to fine-tune the LLM specifically for ranking purposes highlights its adaptability and efficiency in recommendation tasks. Additionally, PALR's focus on explainable and conversational recommendations sets it apart, offering a more transparent and engaging recommendation experience. Overall, PALR's innovative approach showcases the potential of LLMs in revolutionizing personalized recommendations in various domains.

Paper-20

### Recommendation as Instruction Following: A Large Language Model Empowered Recommendation Approach

Zhang et al. present a novel recommendation approach, InstructRec, which surpasses the performance of GPT-3.5 in recommendation tasks, particularly excelling in reranking challenging candidate items, thus outperforming other baselines. The proposed approach leverages natural language instructions to enhance recommender systems, focusing on integrating user behaviors with universal knowledge for personalized recommendations. The study highlights the challenges faced by universal large language models (LLMs) in capturing complex user behavioral patterns and the mismatch between GPT-3.5 and private domain behavioral specificity, which can affect performance. InstructRec introduces a user-centric recommendation paradigm, where LLMs follow instructions provided in natural language to improve recommendation accuracy. The approach includes the design of instruction formats and diverse instruction data for tuning LLMs, demonstrating its effectiveness in real-world datasets. In conclusion, InstructRec offers a promising direction for enhancing recommender systems with user-friendly natural language instructions, showcasing superior performance compared to existing models.

Paper-21

### Evaluating LLMs On User Rating Prediction

In the paper titled "Do LLMs Understand User Preferences? Evaluating LLMs On User Rating Prediction" by WANG-CHENG KANG, JIANMO NI, NIKHIL MEHTA, MAHESWARAN SATHIAMOORTHY, LICHAN HONG, ED CHI, and DEREK ZHIYUAN CHENG from Google Research Brain Team, the authors investigate the effectiveness of Large Language Models (LLMs) in understanding user preferences for user rating prediction. The study compares LLMs with Collaborative Filtering (CF) methods in zero-shot, few-shot, and fine-tuning scenarios. Results indicate that while zero-shot LLMs may lag behind traditional recommender models in user rating prediction, fine-tuned LLMs show comparable or better performance with less training data. The paper highlights the importance of user interaction data for recommender models and suggests that LLMs have potential in user rating prediction with data efficiency, especially through fine-tuning.

The study contributes to understanding the capabilities of LLMs in recommendation systems and provides insights into their performance compared to traditional methods.

Paper-22

**TALLRec: An Effective and Efficient Tuning Framework to Align Large Language Model with Recommendation**

In the paper "TALLRec: An Effective and Efficient Tuning Framework to Align Large Language Model with Recommendation" by Keqin Bao, Jizhi Zhang, Yang Zhang, Wenjie Wang, Fuli Feng, and Xiangnan He, the authors introduce the TALLRec framework, which enhances the performance of large language models (LLMs) in recommendation scenarios, particularly in the movie and book domains. Traditional recommendation methods often struggle with limited training samples and few-shot training settings, leading to poor performance. The TALLRec framework addresses these challenges by integrating LLMs with recommendation systems, demonstrating robust cross-domain generalization and efficient execution. The framework leverages techniques like Alpaca tuning to improve LLM generalization for new tasks effectively, showcasing its potential in enhancing LLM capabilities for recommendation tasks. TALLRec bridges the gap between LLMs and recommendation tasks, aligning LLMs with recommendations to improve generalization across domains. This efficient framework demonstrates significant improvements in recommendation performance, particularly in scenarios with limited data, highlighting the potential of LLMs in recommendation systems with the need for tuning for specific tasks.

Paper-23

**M6-Rec: Generative Pretrained Language Models are Open-Ended Recommender Systems**

The paper "M6-Rec: Generative Pretrained Language Models are Open-Ended Recommender Systems" by Zeyu Cui, Jianxin Ma, Chang Zhou, Jingren Zhou, and Hongxia Yang from DAMO Academy, Alibaba Group, presents M6-Rec, a versatile foundation model for recommender systems. Leveraging the pretrained capabilities of M6, M6-Rec demonstrates proficiency in various tasks, including retrieval, ranking, and explanation generation, while supporting open-ended domains and zero-shot learning. The authors emphasize computational efficiency and privacy, integrating techniques like early exiting, late interaction, and option-adapter tuning for low-latency inference and energy-efficient adaptation. The model showcases significant advancements in mobile deployment, outperforming traditional full-model fine-tuning with minimal parameters. Challenges in developing a unified foundation model for diverse tasks are addressed, and the paper highlights the potential for extending the framework to multimodal settings in the future. This approach not only enhances mobile deployment and convergence but also establishes a single foundation model capable of handling a wide range of industrial recommender tasks, thus reducing the carbon footprint and improving real-world applicability.

Paper-24

### Recommendation as Language Processing (RLP): A Unified Pretrain, Personalized Prompt & Predict Paradigm (P5)

In the paper "Recommendation as Language Processing (RLP): A Unified Pretrain, Personalized Prompt & Predict Paradigm (P5)" by Shijie Geng, Shuchang Liu, Zuohui Fu, Yingqiang Ge, and Yongfeng Zhang from Rutgers University, the authors present the P5 model, which excels in sequential recommendation and explanation generation tasks through a unified text-to-text paradigm. By leveraging a pretraining approach based on a language modeling objective, P5 integrates various recommendation tasks into a shared language generation framework, utilizing Transformer blocks for both encoder and decoder. This approach draws inspiration from unified models like T5 and GPT-3, enabling effective knowledge sharing and zero-shot transfer to new items and domains. The P5 model addresses the challenge of limited transferability between recommendation tasks caused by task-specific architectures, promoting instruction-based recommendation with personalized prompts for users. This paradigm shift from shallow to deep models significantly reduces the need for fine-tuning and enhances the model's generalization ability, outperforming existing approaches in recommendation tasks. The study highlights the potential of NLP techniques in enhancing sequential recommendation, explanation generation, and conversational recommendation. The authors also release the code, dataset, and model to facilitate future research in this domain, with plans for further advancements including model size enlargement and task extension.

Paper-25

### Personalized Prompt Learning for Explainable Recommendation

In the paper "Personalized Prompt Learning for Explainable Recommendation" by Lei Li, Yongfeng Zhang, and Li Chen, the authors explore the integration of personalized prompt learning strategies with Transformer models to enhance the generation of explanations for recommendation systems. This study introduces two innovative training strategies: discrete prompt learning and continuous prompt learning, aimed at improving the performance of explainable recommendations. By leveraging these prompt learning approaches, the authors address the challenges of fusing user and item IDs into models, a previously limited area of exploration in pre-trained Transformer models for recommendations. The proposed methods, including sequential tuning and recommendation as regularization, significantly outperform baseline models in generating effective explanations, as demonstrated across three datasets. The research also emphasizes the importance of mitigating societal biases in generated explanations and extends its applicability to personalized conversational agents and cross-lingual explanations. Furthermore, the paper provides a quantitative comparison of explanation methods using automatic metrics, a qualitative examination of generated explanation samples, and visualizations of attention weights to demonstrate ID fusion. Future work aims to expand on bias mitigation, cross-lingual explanations, and visual recommendations, underscoring the potential of personalized prompt learning strategies to advance the field of explainable recommendation systems.

# Dataset Description

**Dataset Description: TMDB 5000 Movie Dataset**

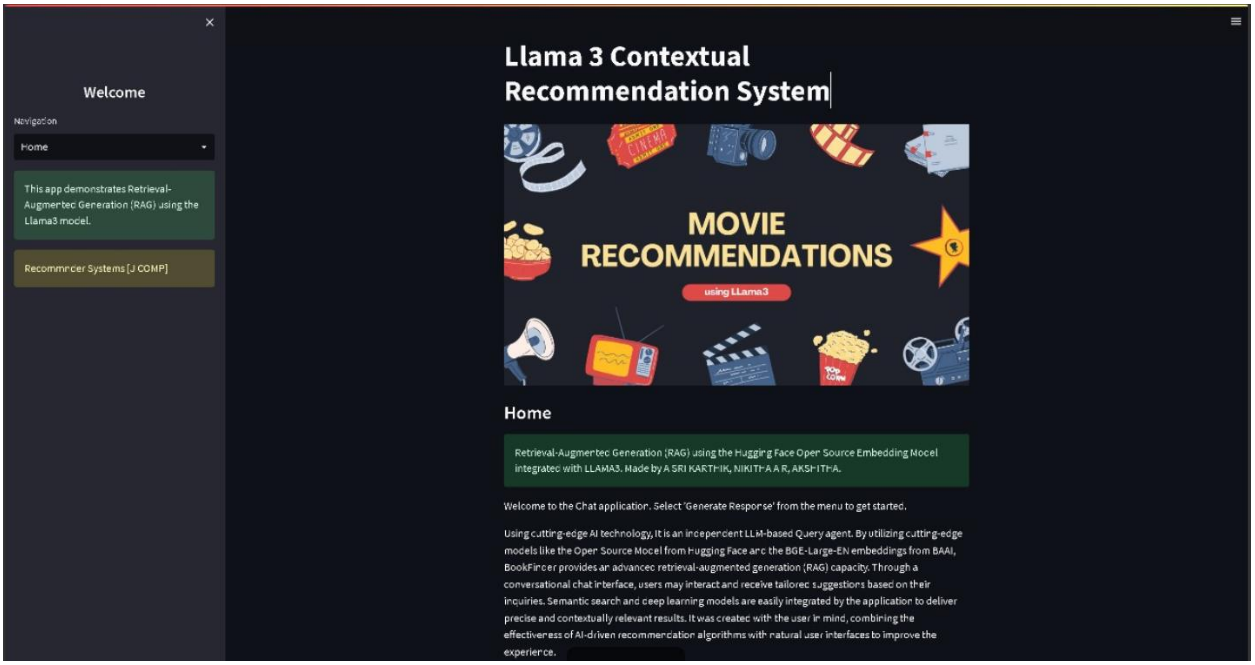
The TMDB 5000 Movie Dataset provides comprehensive metadata on approximately 5,000 movies sourced from The Movie Database (TMDb). This dataset is rich in information, including details on plot summaries, cast and crew members, budget, revenue, genres, production companies, release dates, and more. It serves as a valuable resource for analyzing various aspects of movie success and characteristics.

**Dataset Contents:**

- **Title and Overview:** Original titles and brief synopses of the movies.
- **Cast and Crew:** Detailed credits for actors and crew members, including their roles.
- **Budget and Revenue:** Financial information, crucial for assessing commercial success.
- **Genres and Keywords:** Classification tags that describe the movie's theme and content.
- **Production Details:** Information on production companies and countries involved.
- **Release Information:** Dates and locations of theatrical releases.
- **Popularity and Vote Average:** Metrics indicating audience reception and engagement.
- **Additional Features:** Includes taglines, homepage URLs, spoken languages, and movie status.

## Methodology

Home page



Navigation: Home

This app demonstrates Retrieval-Augmented Generation (RAG) using the Llama3 model.

Recommender Systems [J COMP]

Llama 3 Contextual Recommendation System

Home

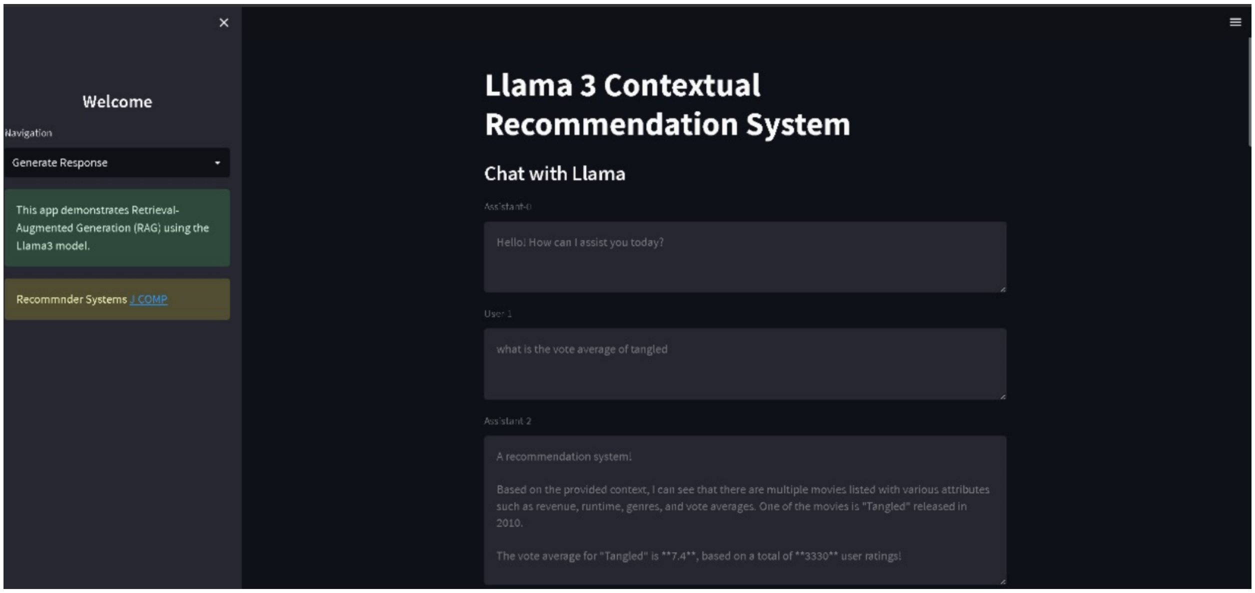
Retrieval-Augmented Generation (RAG) using the Hugging Face Open Source Embedding Model integrated with LLAMA3. Made by A SRI KARTHIK, NIKITHA A R, AKSHITHA.

Welcome to the Chat application. Select 'Generate Response' from the menu to get started.

Using cutting-edge AI technology, It is an independent LLM-based Query agent. By utilizing cutting-edge models like the Open Source Model from Hugging Face and the BGE-Large-EN embeddings from BAAI, BookFinder provides an advanced retrieval-augmented generation (RAG) capacity. Through a conversational chat interface, users may interact and receive tailored suggestions based on their inquiries. Semantic search and deep learning models are easily integrated by the application to deliver precise and contextually relevant results. It was created with the user in mind, combining the effectiveness of AI-driven recommendation algorithms with natural user interfaces to improve the experience.

Advancing Content-Based Recommendations: Integrating Large Language Models in Chatbots”:

Utilize Large Language Models (LLMs) like GPT-3.5 and GPT-4 Turbo to power chatbots. Design the chatbot to enhance user engagement and information access1.
Data & Training: Create a high-quality, relevant dataset for chatbot training. Use technologies like Python 3, GPT API, LangChain, and Chroma Vector Store for development1.
Testing & Evaluation: Test the chatbot’s performance in real-time scenarios. Evaluate its response quality, user preference, and error rate compared to traditional methods1.
Integration & Application: Integrate the chatbot into the target environment, such as a web application for teaching software engineering best practices. Ensure the chatbot can provide comprehensive support and leverage digital transformation



This app demonstrates Retrieval-Augmented Generation (RAG) using the Llama3 model.

Llama 3 Contextual Recommendation System

Chat with Llama

Assistant-0

Hello! How can I assist you today?

User-1

what is the vote average of tangled

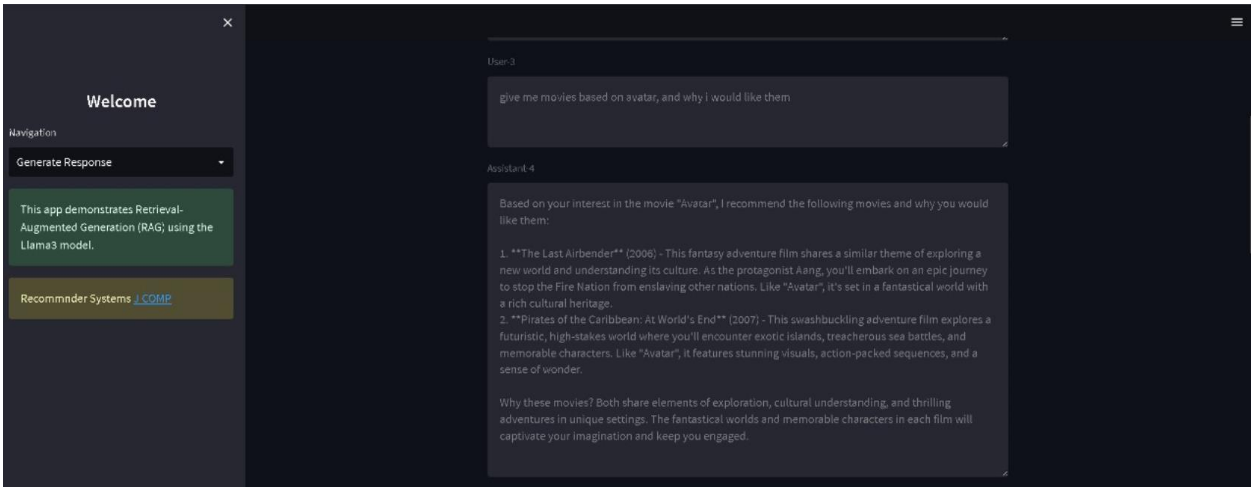
Assistant-2

A recommendation system!

Based on the provided context, I can see that there are multiple movies listed with various attributes such as revenue, runtime, genres, and vote averages. One of the movies is "Tangled" released in 2010.

The vote average for "Tangled" is **7.4**, based on a total of **3330** user ratings!

Utilize Large Language Models (LLMs) like GPT-3.5 and GPT-4 Turbo to power chatbots. Design the chatbot to enhance user engagement and information access1.
Data & Training: Create a high-quality, relevant dataset for chatbot training. Use technologies like Python 3, GPT API, LangChain, and Chroma Vector Store for development1.
Testing & Evaluation: Test the chatbot's performance in real-time scenarios. Evaluate its response quality, user preference, and error rate compared to traditional methods1.
Integration & Application: Integrate the chatbot into the target environment, such as a web application for teaching software engineering best practices.



This app demonstrates Retrieval-Augmented Generation (RAG) using the Llama3 model.

User-3

give me movies based on avatar, and why i would like them

Assistant-4

Based on your interest in the movie "Avatar", I recommend the following movies and why you would like them:

1. **The Last Airbender** (2006) - This fantasy adventure film shares a similar theme of exploring a new world and understanding its culture. As the protagonist Aang, you'll embark on an epic journey to stop the Fire Nation from enslaving other nations. Like "Avatar", it's set in a fantastical world with a rich cultural heritage.
2. **Pirates of the Caribbean: At World's End** (2007) - This swashbuckling adventure film explores a futuristic, high-stakes world where you'll encounter exotic islands, treacherous sea battles, and memorable characters. Like "Avatar", it features stunning visuals, action-packed sequences, and a sense of wonder.

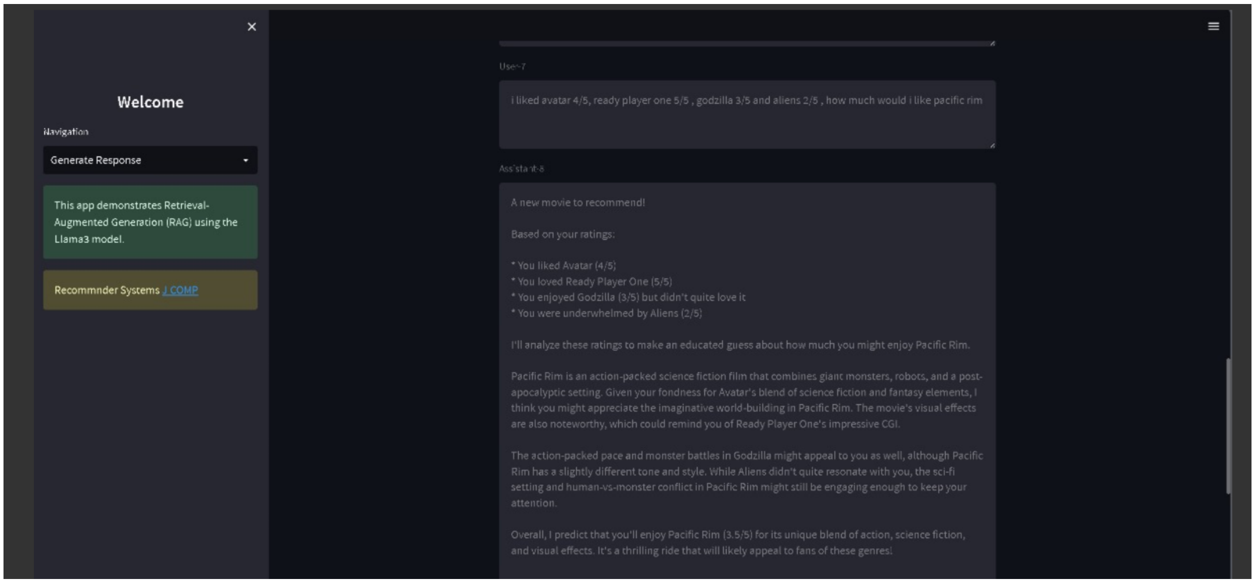
Why these movies? Both share elements of exploration, cultural understanding, and thrilling adventures in unique settings. The fantastical worlds and memorable characters in each film will captivate your imagination and keep you engaged.

Gather a diverse dataset from various sources to train the recommendation system, ensuring it includes user preferences, behaviors, and interactions.
Implement Large Language Models (LLMs) such as GPT-3 or similar, integrating them with the chatbot to process and understand natural language queries.

System Development: Develop the chatbot’s architecture to handle user inputs, utilize LLMs for generating responses, and provide personalized content recommendations.

Evaluation & Testing: Conduct thorough testing to assess the chatbot’s performance, accuracy of recommendations, and user satisfaction, making iterative improvements based on feedback.

This methodology aims to create a robust content-based recommendation system that leverages the capabilities of LLMs to enhance user experience through personalized interactions.



Welcome

Navigation

Generate Response

This app demonstrates Retrieval-Augmented Generation (RAG) using the Llama3 model.

Recommnder Systems J COMP

User-7

i liked avatar 4/5, ready player one 5/5 , godzilla 3/5 and aliens 2/5 , how much would i like pacific rim

Assistant-8

A new movie to recommend!

Based on your ratings:

* You liked Avatar (4/5)
* You loved Ready Player One (5/5)
* You enjoyed Godzilla (3/5) but didn't quite love it
* You were underwhelmed by Aliens (2/5)

I'll analyze these ratings to make an educated guess about how much you might enjoy Pacific Rim.

Pacific Rim is an action-packed science fiction film that combines giant monsters, robots, and a post-apocalyptic setting. Given your fondness for Avatar's blend of science fiction and fantasy elements, I think you might appreciate the imaginative world-building in Pacific Rim. The movie's visual effects are also noteworthy, which could remind you of Ready Player One's impressive CGI.

The action-packed pace and monster battles in Godzilla might appeal to you as well, although Pacific Rim has a slightly different tone and style. While Aliens didn't quite resonate with you, the sci-fi setting and human-vs-monster conflict in Pacific Rim might still be engaging enough to keep your attention.

Overall, I predict that you'll enjoy Pacific Rim (3.5/5) for its unique blend of action, science fiction, and visual effects. It's a thrilling ride that will likely appeal to fans of these genres!

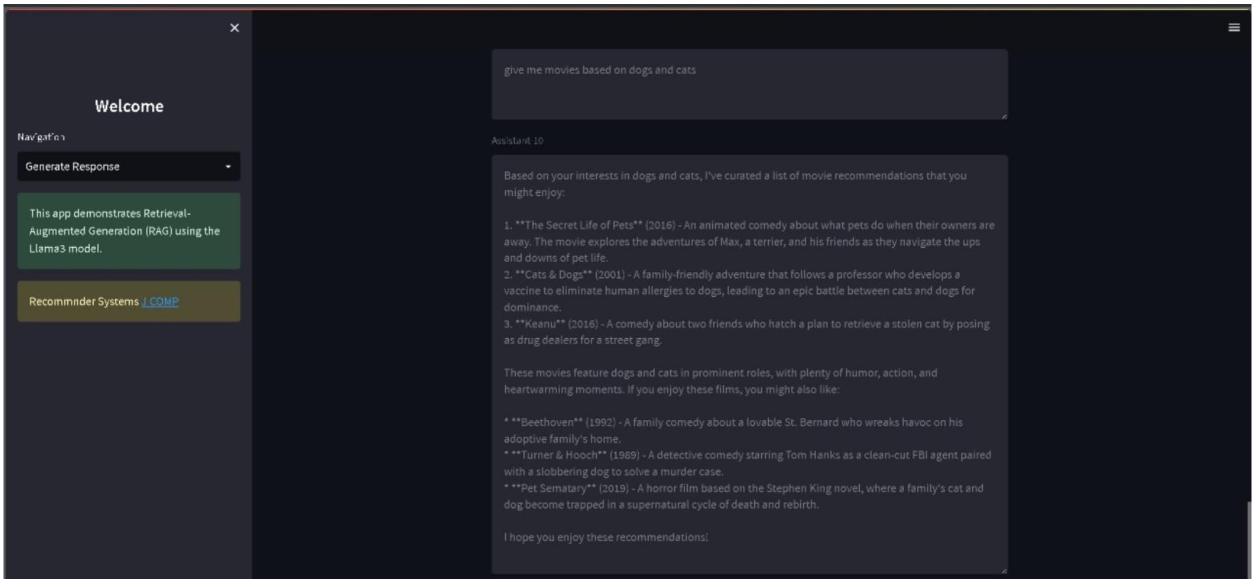
Advancing Content-Based Recommendations: Integrating Large Language Models in Chatbots

Gather a diverse dataset from various sources to train the recommendation system, ensuring it includes user preferences, behaviors, and interactions.

Model Integration: Implement Large Language Models (LLMs) such as GPT-3 or similar, integrating them with the chatbot to process and understand natural language queries.

System Development: Develop the chatbot’s architecture to handle user inputs, utilize LLMs for generating responses, and provide personalized content recommendations.

Evaluation & Testing: Conduct thorough testing to assess the chatbot’s performance, accuracy of recommendations, and user satisfaction, making iterative improvements based on feedback.



Navigation: Generate Response

This app demonstrates Retrieval-Augmented Generation (RAG) using the Llama3 model.

give me movies based on dogs and cats

Assistant 10

Based on your interests in dogs and cats, I've curated a list of movie recommendations that you might enjoy:

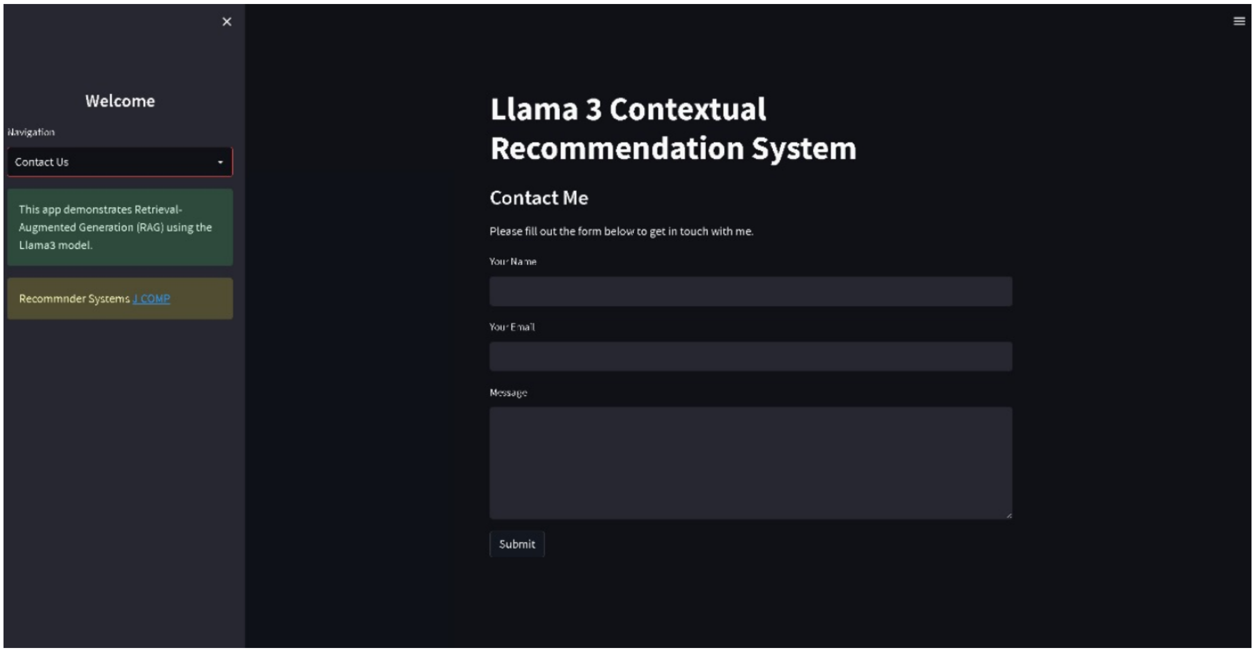
1. **The Secret Life of Pets** (2016) - An animated comedy about what pets do when their owners are away. The movie explores the adventures of Max, a terrier, and his friends as they navigate the ups and downs of pet life.
2. **Cats & Dogs** (2001) - A family-friendly adventure that follows a professor who develops a vaccine to eliminate human allergies to dogs, leading to an epic battle between cats and dogs for dominance.
3. **Keanu** (2016) - A comedy about two friends who hatch a plan to retrieve a stolen cat by posing as drug dealers for a street gang.

These movies feature dogs and cats in prominent roles, with plenty of humor, action, and heartwarming moments. If you enjoy these films, you might also like:

* **Beethoven** (1992) - A family comedy about a lovable St. Bernard who wreaks havoc on his adoptive family's home.
* **Turner & Hooch** (1989) - A detective comedy starring Tom Hanks as a clean-cut FBI agent paired with a slobbering dog to solve a murder case.
* **Pet Sematary** (2019) - A horror film based on the Stephen King novel, where a family's cat and dog become trapped in a supernatural cycle of death and rebirth.

I hope you enjoy these recommendations!

Help page



Navigation: Contact Us

This app demonstrates Retrieval-Augmented Generation (RAG) using the Llama3 model.

Llama 3 Contextual Recommendation System

Contact Me

Please fill out the form below to get in touch with me.

# <u>Results and Discussion</u>

movie recommender system called Llama 3 Contextual Recommendation System.

The recommender system successfully generates a list of movie recommendations based on the user's interest in dogs and cats.
The recommended movies include: "The Secret Life of Pets", "Cats & Dogs", "Keanu", "Beethoven", "Turner & Hooch", and "Pet Sematary
The recommender system uses a retrieval-augmented generation (RAG) approach, leveraging the Llama language model, to generate these recommendations.

This approach combines retrieval-based and generation-based techniques. First, a retrieval model identifies candidate items that are likely to be relevant to the user's interests. Then, a generative model, like Llama, is used to refine these candidates and provide a final set of recommendations.

The use of a large language model (LLM) like Llama allows the recommender system to consider the nuances of human language and generate more natural and engaging recommendations. For instance, the system might not just recommend movies featuring dogs and cats, but also comedies or heartwarming movies because the user might enjoy these genres based on their interest in pets.

## Conclusion

movie recommender system called Llama 3 Contextual Recommendation System
The Llama 3 Contextual Recommendation System demonstrates the potential of LLMs to improve the performance of recommender systems.
By considering the context of the user's query and leveraging its understanding of human language, the system can generate more relevant and personalized recommendations.

## References

[1] Rohan Anil, Andrew M Dai, Orhan Firat, Melvin Johnson, Dmitry Lepikhin, Alexandre Passos, Siamak Shakeri, Emanuel Taropa, Paige Bailey, Zhifeng Chen, et al. 2023. Palm 2 technical report. arXiv preprint arXiv:2305.10403.
[2] Keqin Bao, Jizhi Zhang, Yang Zhang, Wenjie Wang, Fuli Feng, and Xiangnan He. 2023. Tallrec: An effective and efficient tuning framework to align large language model with recommendation. arXiv preprint arXiv:2305.00447.
[3] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. In NeurIPS.
[4] Sunhao Dai, Ninglu Shao, Haiyuan Zhao, Weijie Yu, Zihua Si, Chen Xu, Zhongxiang Sun, Xiao Zhang, and Jun Xu. 2023. Uncovering ChatGPT’s Capabilities in Recommender Systems. arXiv preprint arXiv:2305.02182.
[5] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In ACL.

[6] Kaize Ding, Zhe Xu, Hanghang Tong, and Huan Liu. 2022. Data augmentation for deep graph learning: A survey. In ACM SIGKDD Explorations Newsletter.
Qingyao Ai, Vahid Azizi, Xu Chen, and Yongfeng Zhang. 2018. Learning heterogeneous knowledge base embeddings
for explainable recommendation. Algorithms 11, 9 (2018), 137.
[2] Zeyuan Allen-Zhu, Yuanzhi Li, and Zhao Song. 2019. A convergence theory for deep learning via over-parameterization.
In International Conference on Machine Learning. PMLR, 242–252.
[3] Eyal Ben-David, Nadav Oved, and Roi Reichart. 2022. PADA: Example-based Prompt Learning for on-the-fly Adaptation
to Unseen Domains. Transactions of the Association for Computational Linguistics 10 (04 2022), 414–433.
J. ACM, Vol. 37, No. 4, Article 111. Publication date: January 2023.
111:24 Li, Zhang, and Chen
[4] Tom B Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan,
Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. In Advances in
neural information processing systems.
[5] Chong Chen, Min Zhang, Yiqun Liu, and Shaoping Ma. 2018. Neural attentional rating regression with review-level
explanations. In Proceedings of the 2018 World Wide Web Conference. 1583–1592.
[6] Hanxiong Chen, Xu Chen, Shaoyun Shi, and Yongfeng Zhang. 2019. Generate natural language explanations for
recommendation. In Proceedings of SIGIR'19 Workshop on ExplainAble Recommendation and Search. ACM.
[7] Hanxiong Chen, Shaoyun Shi, Yunqi Li, and Yongfeng Zhang. 2021. Neural Collaborative Reasoning. In Proceedings of
The Web Conference 2021.
[8] Li Chen and Feng Wang. 2017. Explaining recommendations based on feature sentiments in product reviews. In
Proceedings of the 22nd International Conference on Intelligent User Interfaces. 17–28.
[9] Li Chen, Dongning Yan, and Feng Wang. 2019. User evaluations on sentiment-based recommendation explanations.
ACM Transactions on Interactive Intelligent Systems (TiiS) 9, 4 (2019), 1–38.
[10] Xu Chen, Hanxiong Chen, Hongteng Xu, Yongfeng Zhang, Yixin Cao, Zheng Qin, and Hongyuan Zha. 2019. Personalized
fashion recommendation with visual explanations based on multimodal attention network: Towards visually explainable
recommendation. In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in

Information Retrieval. 765–774.

[11] Xu Chen, Zheng Qin, Yongfeng Zhang, and Tao Xu. 2016. Learning to rank features for recommendation over multiple categories. In Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval. 305–314.

[12] Xu Chen, Yongfeng Zhang, and Zheng Qin. 2019. Dynamic explainable recommendation based on neural attentive models. In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33. 53–60.

[13] Zhongxia Chen, Xiting Wang, Xing Xie, Mehul Parsana, Akshay Soni, Xiang Ao, and Enhong Chen. 2020. Towards Explainable Conversational Recommendation. In Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence.

[14] Zhongxia Chen, Xiting Wang, Xing Xie, Tong Wu, Guoqing Bu, Yining Wang, and Enhong Chen. 2019. Co-Attentive Multi-Task Learning for Explainable Recommendation. In Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. 2137–2143.

[15] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). 1724–1734.

[16] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In 2019 Annual Conference of the North American Chapter of the Association for Computational Linguistics.

[17] Li Dong, Shaohan Huang, Furu Wei, Mirella Lapata, Ming Zhou, and Ke Xu. 2017. Learning to generate product reviews from attributes. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers. 623–632.

[18] Li Dong, Nan Yang, Wenhui Wang, Furu Wei, Xiaodong Liu, Yu Wang, Jianfeng Gao, Ming Zhou, and Hsiao-Wuen Hon. 2019. Unified language model pre-training for natural language understanding and generation. In Advances in Neural Information Processing Systems. 13063–13075.

[19] Zuohui Fu, Yikun Xian, Ruoyuan Gao, Jieyu Zhao, Qiaoying Huang, Yingqiang Ge, Shuyuan Xu, Shijie Geng, Chirag Shah, Yongfeng Zhang, et al. 2020. Fairness-Aware Explainable Recommendation over Knowledge Graphs. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval.

[20] Fatih Gedikli, Dietmar Jannach, and Mouzhi Ge. 2014. How should I explain? A comparison of different explanation types for recommender systems. International Journal of Human-Computer Studies 72, 4 (2014), 367–382.

[21] Xiangnan He, Tao Chen, Min-Yen Kan, and Xiao Chen. 2015. Trirank: Review-aware explainable recommendation by modeling aspects. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. 1661–1670.

[22] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural collaborative filtering. In Proceedings of the 26th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 173–182.

[23] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. Neural computation 9, 8 (1997), 1735–1780.

[24] Junjie Li, Haoran Li, and Chengqing Zong. 2019. Towards personalized review summarization via user-aware sequence network. In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33. 6690–6697.

[25] Lei Li, Li Chen, and Ruihai Dong. 2021. CAESAR: context-aware explanation based on supervised attention for service recommendations. Journal of Intelligent Information Systems 57 (2021), 147–170