
MOVIE ASSISTANT CHATBOT
Data Mining | Software Engineering Master's Degree | Universidade do Minho
Abhimanyu Aryan PG51632 | Millena Santos PG54107 | Ricardo Oliveira PG54177

Table of contents


01
Introduction

Context

- Movies are a powerful medium of storytelling that have a profound impact on society. They entertain, educate, and inspire audiences, transcending geographical and cultural boundaries;
- Personalized recommendations help users discover movies that match their preferences, enhancing their viewing experience;
- Hollywood films often set trends in filmmaking and popular culture. The industry is renowned for producing a wide variety of genres, including action, drama, science fiction, and animation.
- Based in Mumbai, India, Bollywood is the largest film industry in the world in terms of the number of films produced and tickets sold.
- The cultural influence of Bollywood is immense, with its movies and stars being adored across South Asia and among the Indian diaspora worldwide.

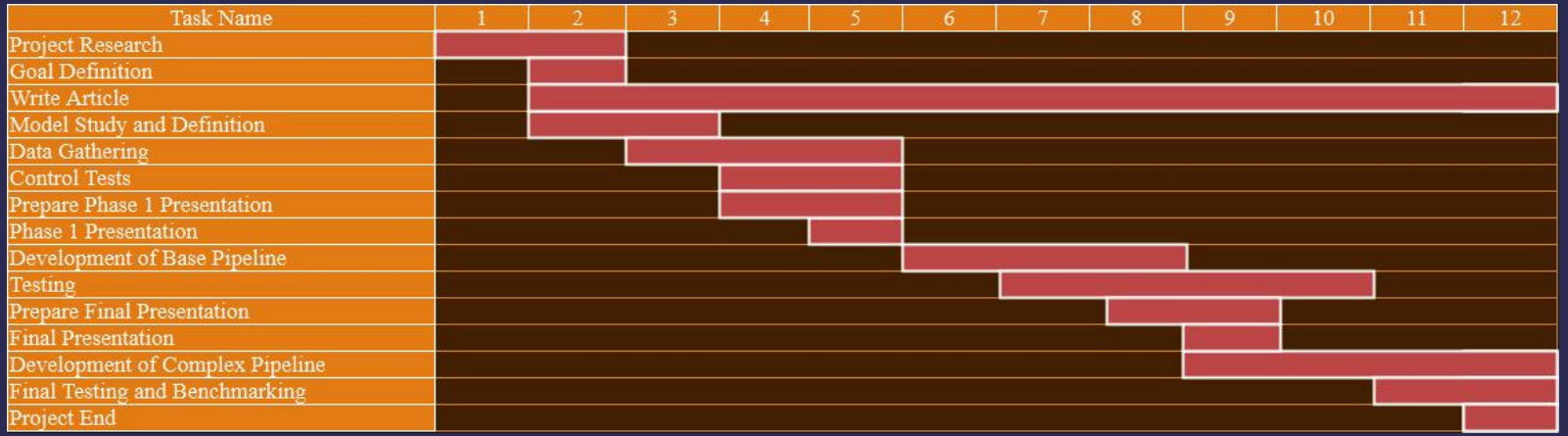
Objectives

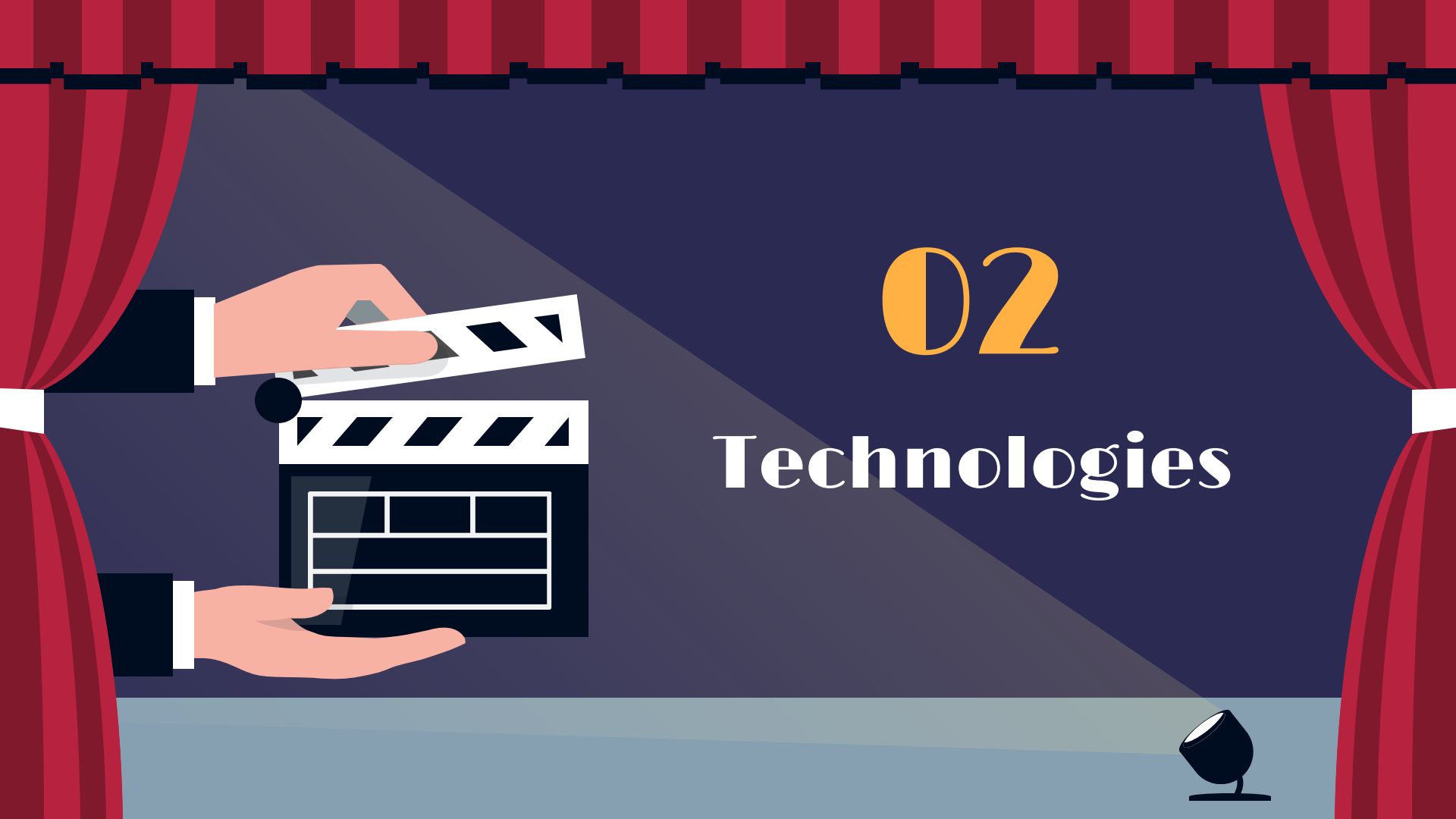
Better recommendation for movies based on more context

Specific questions about movies

Up to date with the latest movies

Planning


Task Name
1
2
3
4
5
6
7
8
9
10
11
12
Project Research
Goal Definition
Write Article
Model Study and Definition
Data Gathering
Control Tests
Prepare Phase 1 Presentation
Phase 1 Presentation
Development of Base Pipeline
Testing
Prepare Final Presentation
Final Presentation
Development of Complex Pipeline
Final Testing and Benchmarking
Project End


02
Technologies

Technologies

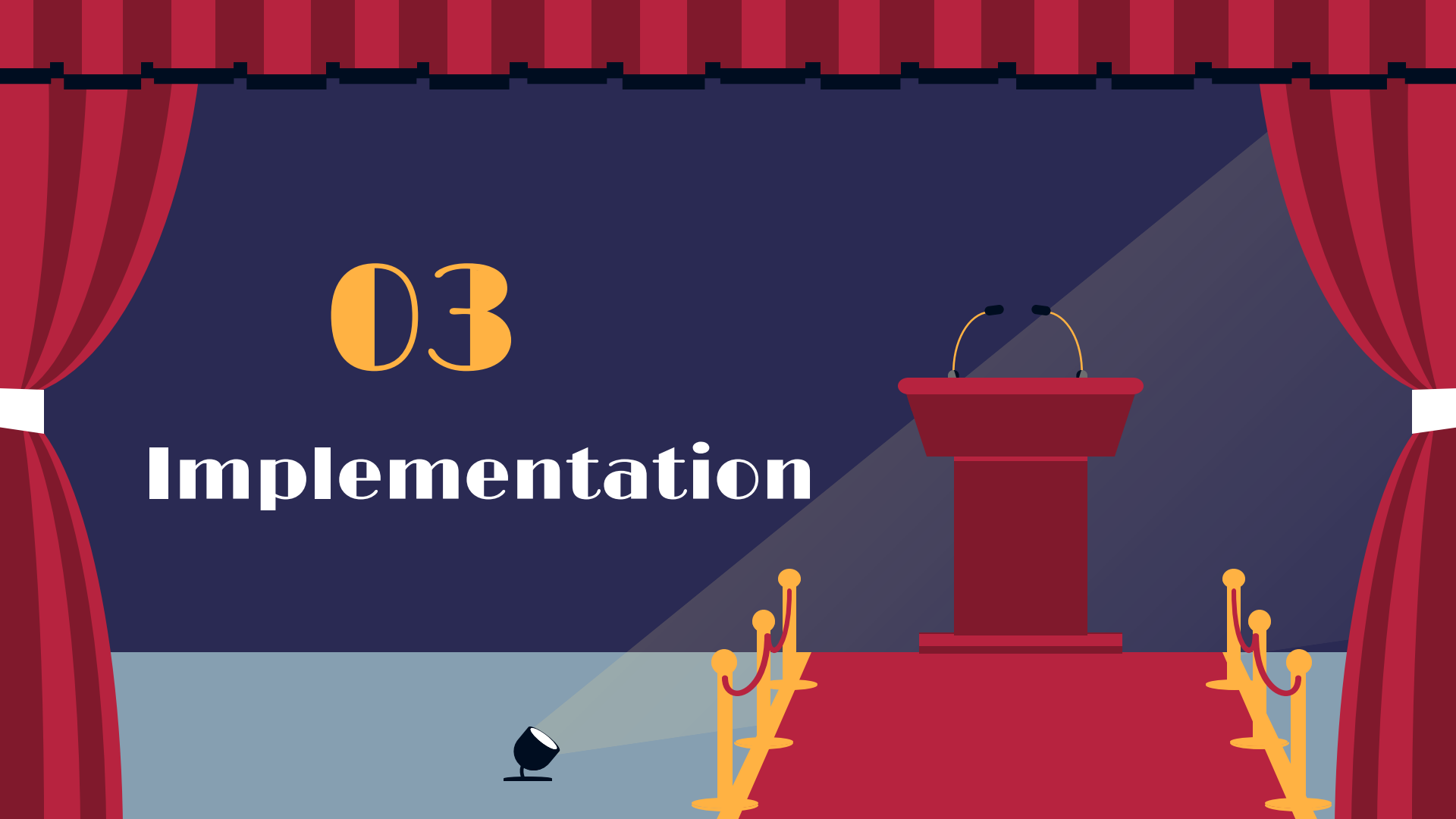
Text-embedding
ada-002 v2

CHATGPT 3.5
Turbo 0125

Llama index

03

Implementation



Implementation Process

01 Data Gathering

02 Information Indexing

03 Pipeline Definition

04 Test the LLM

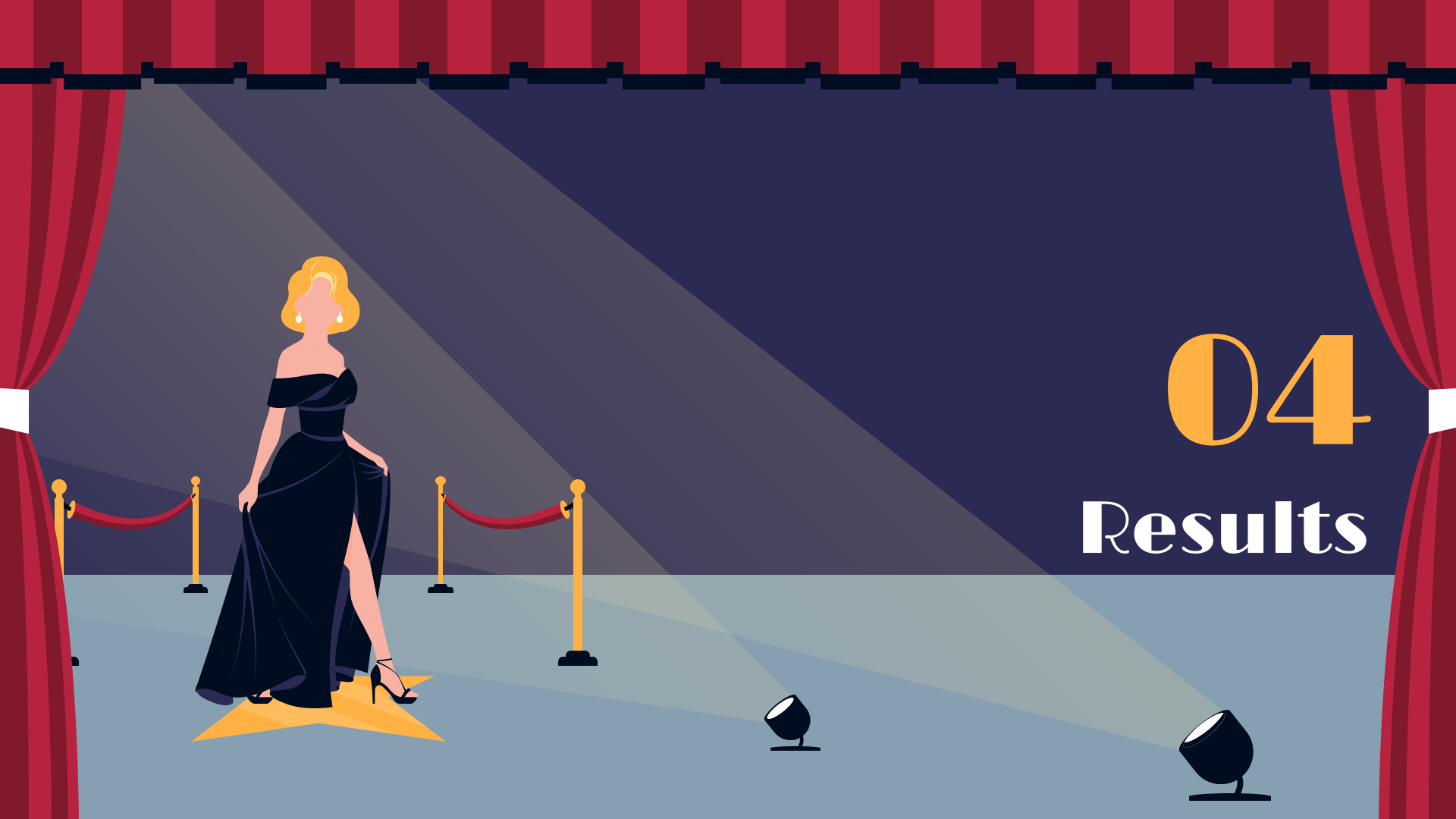
05 Evaluate the Results

Data Gathering

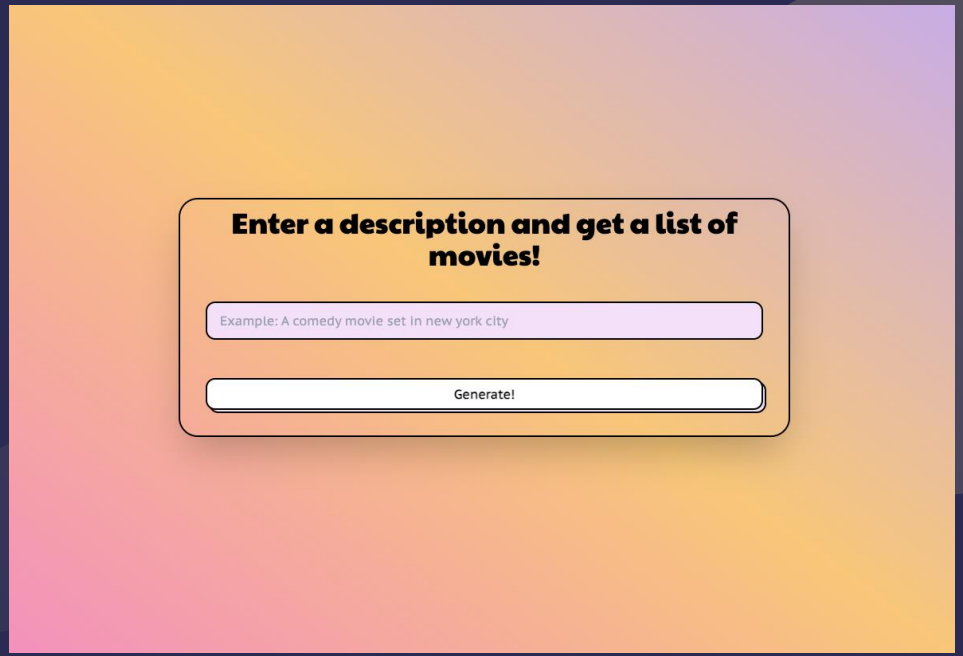
- All movies from the 2020's section in the page *"Lists of Hindi films"*;
- All movies from the 2020's section in the page *"Lists of American films"*;
- Used Wikipedia API to extract all content from each movie's page;
- Contents saved as PDF.

RAG Pipeline

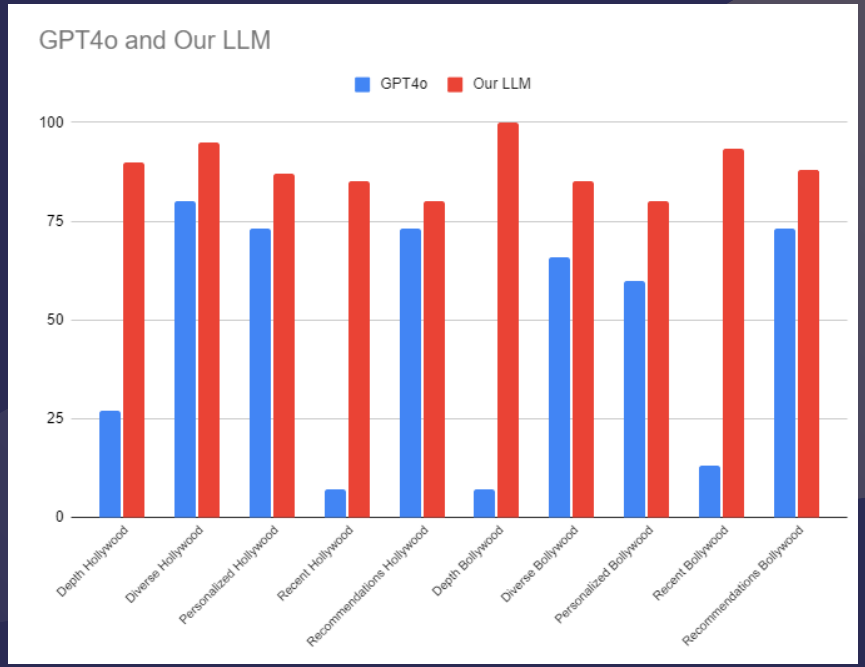
- Prompt Engineering;
- Indexing and Embedding;
- Embedding Storage.


04
Results

User Interface



Comparison with GPT-4o



05

Conclusion

Limitations and Future Work

- Difficulty in handling "tricky" questions;
- Evaluation Pipeline
  - Sentence-window retrieval;
  - Auto-merging retrieval.


MOVIE ASSISTANT CHATBOT
Data Mining | Software Engineering Master's Degree | Universidade do Minho
Abhimanyu Aryan PG51632 | Millena Santos PG54107 | Ricardo Oliveira PG54177