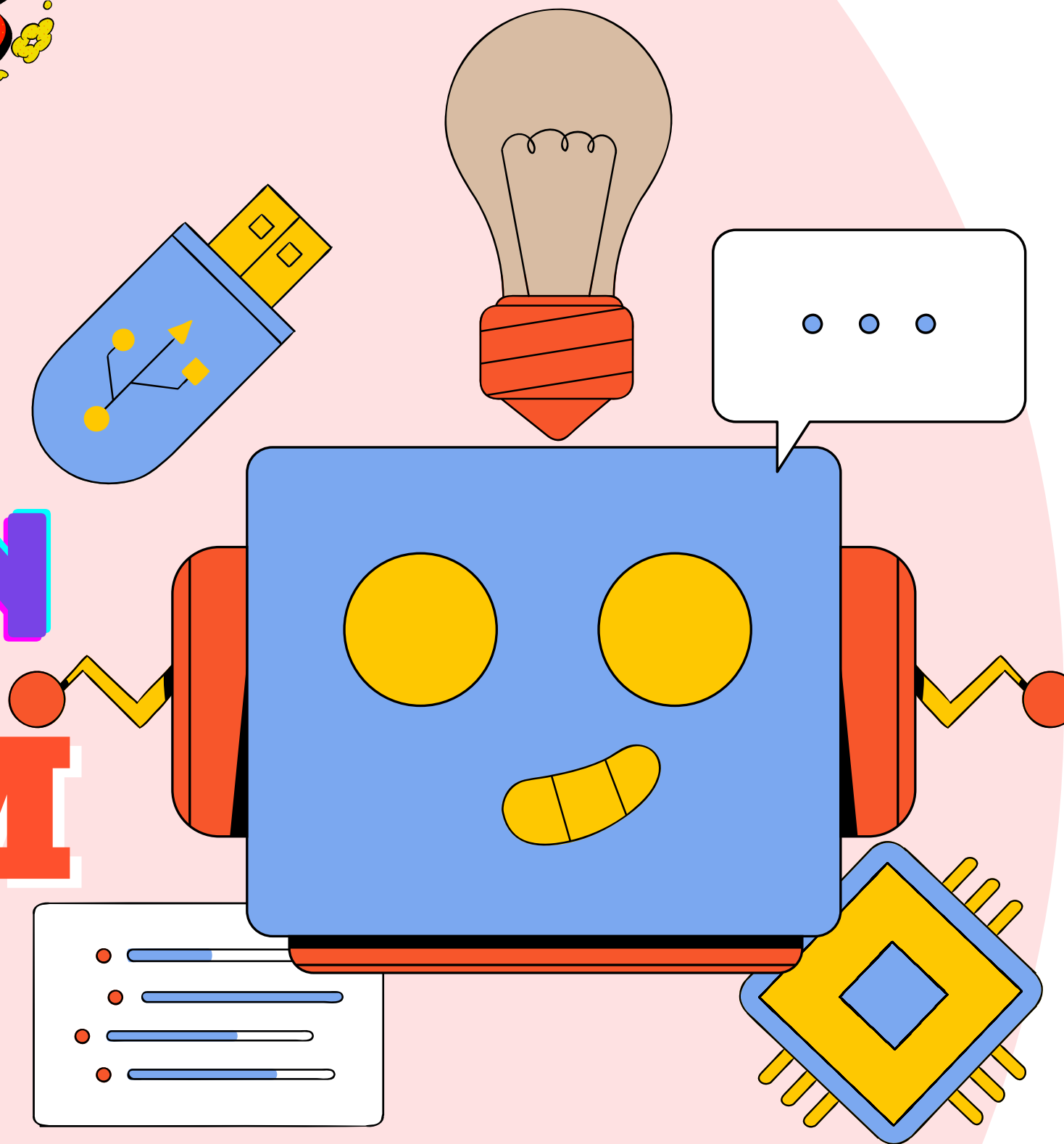


**D S 1 0 8**

**NHÓM 24**

# **RECOMMENDATION MOVIE SYSTEM**

**GVHD:** TS.Nguyễn Gia Tuấn Anh  
CN.Trần Quốc Khánh



# INTRODUCTION

## Ý TƯỞNG ĐỀ TÀI

- Hệ thống đưa ra những bộ phim gợi ý dựa theo sở thích tìm kiếm cá nhân của người dùng
- Dùng hai mô hình TF-IDF+KNN và TF-IDF+Cosin Similarity

## KỲ VỌNG

- Giúp người dùng tiết kiệm thời gian cho việc tìm kiếm và khám phá những bộ phim khác họ có thể yêu thích

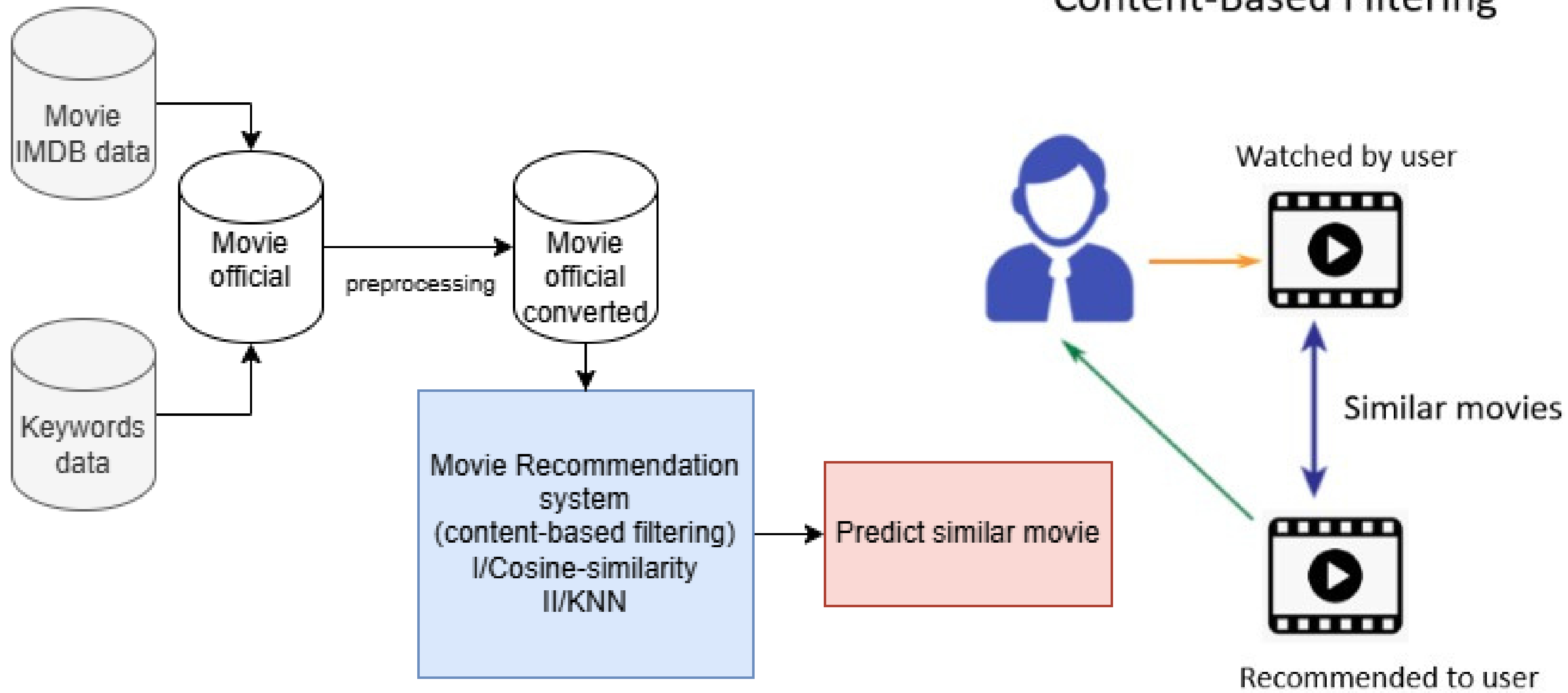
## Ý NGHĨA, PHÁT TRIỂN HỆ THỐNG

- Có thể cải thiện tối ưu hóa hơn, nếu có thêm nhiều dữ liệu nạp vào giúp nó đưa ra dự đoán chính xác và đa dạng hơn
- Kết hợp được nhiều phương pháp khác.



# MÔ TẢ NGHIÊN CỨU

## Content-Based Filtering



# PREPROCESSING

## Chuyển đổi

Chuyển đổi dữ liệu từ dạng **dictionary** sang dạng **chuỗi** có thể giảm chiều dữ, tăng khả năng tương tác giữa các đặc trưng liệu, giúp cải thiện hiệu suất và độ chính xác cho **KNN và Cosine Similarity**

```
[{'id': 8828, 'name': 'marvel comic'}, {'id': 9715, 'name': 'superhero'}, {'id': 9717, 'name': 'based on comic'}, {'id': 179430, 'name': 'aftercreditsstinger'}, {'id': 179431, 'name': 'duringcreditsstinger'}, {'id': 180547, 'name': 'marvel cinematic universe'}, {'id': 209714, 'name': '3d'}]
```

marvel comic, superhero, based on comic,  
aftercreditsstinger, duringcreditsstinger, marvel cinematic  
universe, 3d

# PREPROCESSING

## The Avengers (2012)

new york, shield, marvel comic, superhero, based on comic, alien invasion, superhero team, aftercreditsstinger, duringcreditsstinger, marvel cinematic universe

## Avengers: Age of Ultron (2015)

marvel comic, sequel, superhero, based on comic, vision, superhero team, duringcreditsstinger, marvel cinematic universe, 3d

## Avengers: Infinity War (2018)

mutant, supernatural powers, marvel comic, superhero, based on comic, superhuman, apocalypse, superhero team, world domination, aftercreditsstinger, 1980s

## Avengers: Endgame (2019)

civil war, war, marvel comic, sequel, superhero, based on comic, imax, aftercreditsstinger, duringcreditsstinger, marvel cinematic universe, 3d

## Dự đoán keywords

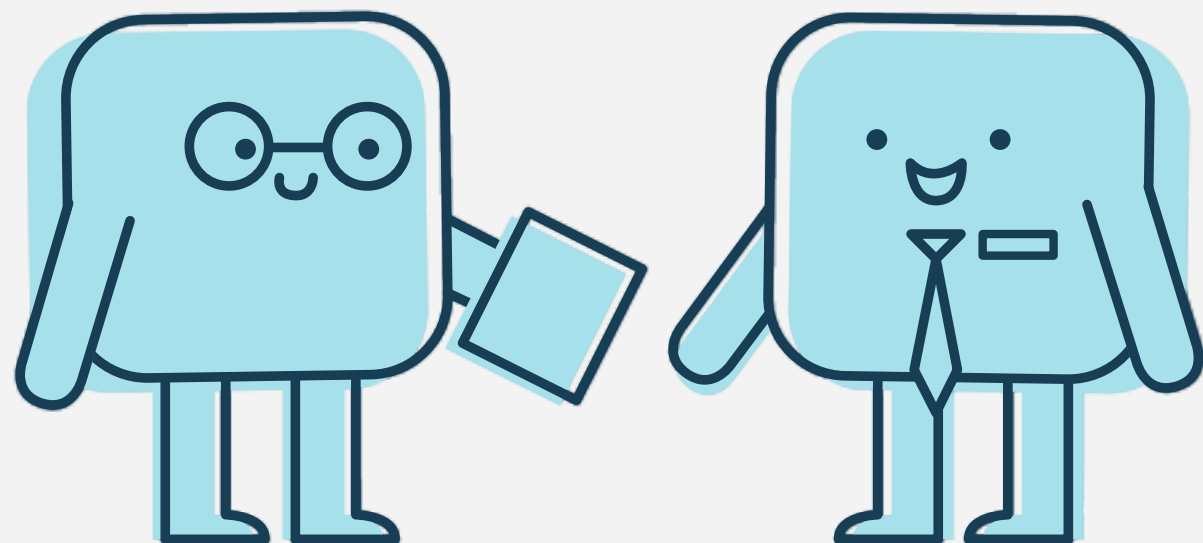
Dựa trên keywords đã có sẵn của 770 bộ phim, sẽ sử dụng **TF-IDF+Cosine Similarity** dự đoán keywords cho 230 bộ phim sau 2017



# ĐẦU VÀO

---

# ĐẦU RA



## Lệnh tìm kiếm truy vấn ( từ hoặc cụm từ có nghĩa):

- Tên của bộ phim hoặc các thuộc tính có liên quan đến phim mà người dùng quan tâm đến
  - Thuộc tính liên quan có thể là: thể loại phim, các từ khóa liên quan đến phim,....
- 

Danh sách top 10 phim có liên quan đến lệnh truy vấn

**TF-IDF+KNN:** khoảng cách từ lệnh truy vấn đến các bộ phim tương tự

**TF-IDF+Cosine Similarity:** đánh giá độ tương đồng so với lệnh truy vấn

## “car racing” with cosine similarity

Name of movie	Year of release	Genre	Similarity Score
The Fast and the Furious	2001	Action, Crime, Thriller	0.392096
2 Fast 2 Furious	2003	Action, Crime, Thriller	0.380914
Fast & Furious 6	2013	Action, Adventure, Crime	0.346331
Fast X	2023	Action, Adventure, Crime	0.288712
Furious 7	2015	Action, Crime, Thriller	0.261671
Fast & Furious	2009	Action, Crime, Thriller	0.261211
Fast Five	2011	Action, Crime, Thriller	0.247087
The Final Destination	2009	Horror, Thriller	0.205030
Gone in 60 Seconds	2000	Action, Crime, Thriller	0.202658
Ford v Ferrari	2019	Action, Biography, Drama	0.198483



# “car racing Fast & Furious Action” with cosine similarity

Name of movie	Year of release	Genre	Similarity Score
Fast & Furious	2009	Action, Crime, Thriller	0.488757
2 Fast 2 Furious	2003	Action, Crime, Thriller	0.449952
The Fast and the Furious	2001	Action, Crime, Thriller	0.430217
Fast & Furious 6	2013	Action, Adventure, Crime	0.409101
Furious 7	2015	Action, Crime, Thriller	0.304113
Fast & Furious Presents: Hobbs & Shaw	2019	Action, Adventure, Thriller	0.293546
Fast X	2023	Action, Adventure, Crime	0.275236
Fast Five	2011	Action, Crime, Thriller	0.235554
The Fate of the Furious	2017	Action, Crime, Thriller	0.200873
Gone in 60 Seconds	2000	Action, Crime, Thriller	0.139081



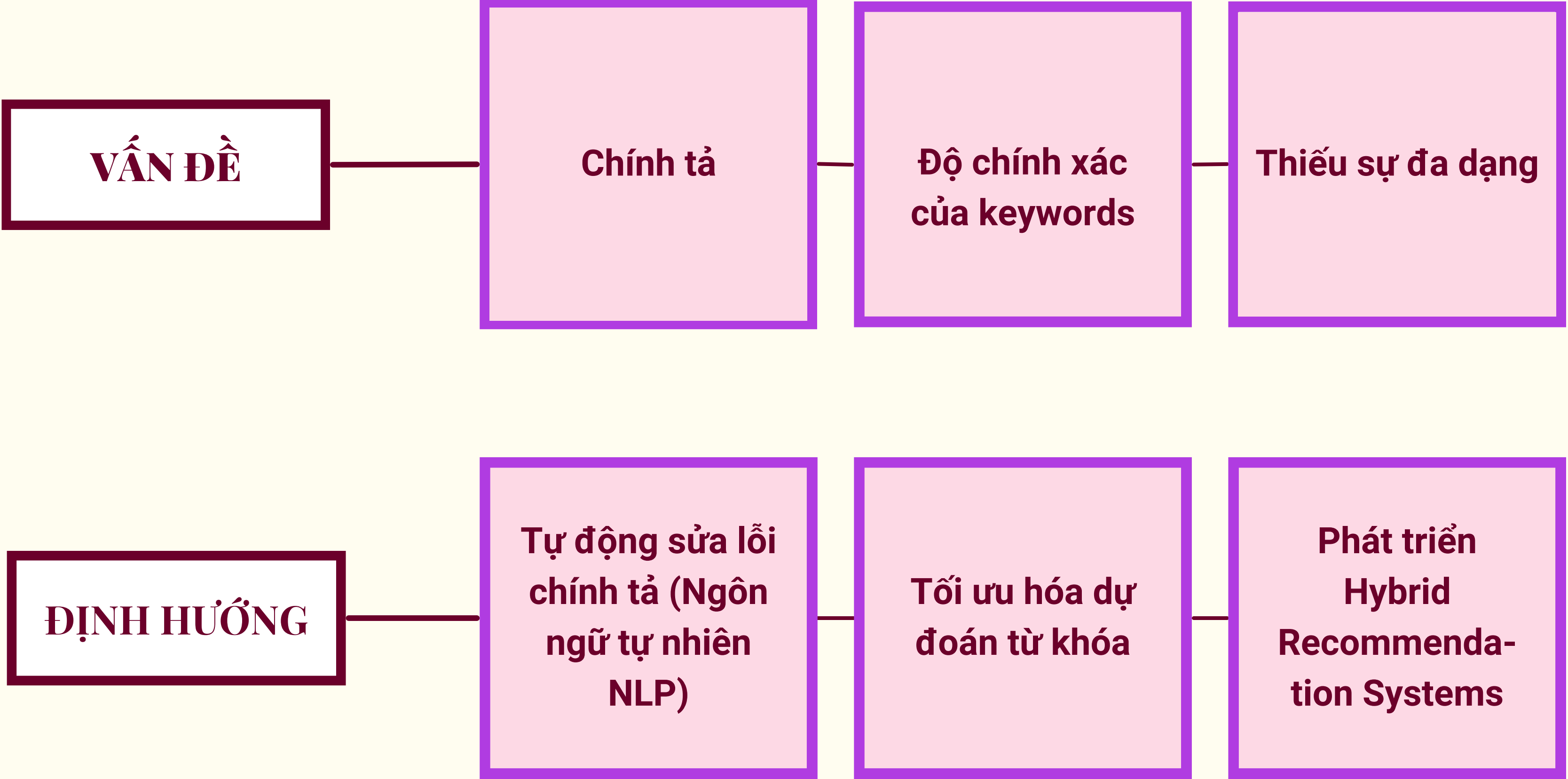
## “Robot” with KNN

Name of movie	Year of release	Genre	Distance
Robots	2005	Animation, Adventure, Comedy	0.947110
I, Robot	2004	Action, Mystery, Sci-Fi	0.960347
Real Steel	2011	Action, Drama, Sci-Fi	1.051595
Transformers: Age of Extinction	2014	Action, Adventure, Sci-Fi	1.158090
Big Hero 6	2014	Animation, Action, Adventure	1.236624
The Day the Earth Stood Still	2008	Adventure, Drama, Sci-Fi	1.253967
WALL-E	2008	Animation, Adventure, Family	1.270625
Transformers	2007	Action, Adventure, Sci-Fi	1.296347
Transformers: Rise of the Beasts	2023	Action, Adventure, Sci-Fi	1.306453
Sing 2	2021	Animation, Adventure, Comedy	1.308129

## “Robot Transformers Sci-Fi” with KNN

Name of movie	Year of release	Genre	Distance
Transformers: Age of Extinction	2014	Action, Adventure, Sci-Fi	0.927604
Transformers: Rise of the Beasts	2023	Action, Adventure, Sci-Fi	0.966285
Transformers: Revenge of the Fallen	2009	Action, Adventure, Sci-Fi	1.024485
Transformers	2007	Action, Adventure, Sci-Fi	1.058810
I, Robot	2004	Action, Mystery, Sci-Fi	1.134221
Transformers: The Last Knight	2017	Action, Adventure, Sci-Fi	1.137672
Real Steel	2011	Action, Drama, Sci-Fi	1.141398
Transformers: Dark of the Moon	2011	Action, Adventure, Sci-Fi	1.142060
Moon Man	2022	Comedy, Sci-Fi	1.211183
Sing 2	2021	Animation, Adventure, Comedy	1.260679

# Vấn đề và định hướng tương lai



## “Transformers” with cosine similarity

	Name of movie	Year of release	Genre	Similarity Score
1000	Snow White and the Seven Dwarfs	1937	Animation, Adventure, Family	0.0
328	22 Jump Street	2014	Action, Comedy, Crime	0.0
341	Dracula Untold	2014	Action, Drama, Fantasy	0.0
340	Annabelle	2014	Horror, Mystery, Thriller	0.0
339	Stand by Me Doraemon	2014	Animation, Comedy, Drama	0.0
338	Into the Woods	2014	Adventure, Comedy, Drama	0.0
337	Ode to My Father	2014	Drama, War	0.0
336	Fury	2014	Action, Drama, War	0.0
335	The Fault in Our Stars	2014	Drama, Romance	0.0
334	The Equalizer	2014	Action, Crime, Thriller	0.0

## “Tranformer” with cosine similarity

	Name of movie	Year of release	Genre	Similarity Score
609	Transformers	2007	Action, Adventure, Sci-Fi	0.0
205	Transformers: The Last Knight	2017	Action, Adventure, Sci-Fi	0.0
474	Transformers: Dark of the Moon	2011	Action, Adventure, Sci-Fi	0.0
344	Transformers: Age of Extinction	2014	Action, Adventure, Sci-Fi	0.0
5	Transformers: Rise of the Beasts	2023	Action, Adventure, Sci-Fi	0.0
557	Transformers: Revenge of the Fallen	2009	Action, Adventure, Sci-Fi	0.0
1000	Snow White and the Seven Dwarfs	1937	Animation, Adventure, Family	0.0
328	22 Jump Street	2014	Action, Comedy, Crime	0.0
341	Dracula Untold	2014	Action, Drama, Fantasy	0.0
340	Annabelle	2014	Horror, Mystery, Thriller	0.0

# FUTURE WORK

## Tự động sửa lỗi chính tả (Ngôn ngữ tự nhiên NLP)

- Các thư viện như NLTK hoặc SpaCy có thể giúp tách từ, kiểm tra chính tả và đề xuất các từ đúng chính tả.
- Kiểm tra từ gần giống nhất: Khi người dùng nhập sai chính tả, hệ thống có thể kiểm tra từ gần đúng nhất và đề xuất từ đúng. Ví dụ: nếu người dùng nhập "Tranformer," hệ thống có thể đề xuất "Transformers."
- Xây dựng mô hình học máy để nhận diện và sửa lỗi chính tả: Các mô hình như Naive Bayes, SVM

## Tối ưu hóa dự đoán từ khóa

- Sử dụng mô hình học máy: Xây dựng các mô hình như Random Forest hoặc Gradient Boosting để dự đoán mức độ quan trọng của từ khóa dựa trên các đặc trưng của từ khóa và ngữ cảnh.
- Kết hợp các phương pháp: Sử dụng TF-IDF để xác định từ khóa ban đầu, sau đó tinh chỉnh bằng attention scores từ BERT hoặc TextRank.

## Phát triển hệ thống gợi ý lai (Hybrid Recommendation System)

- Kết hợp nhiều phương pháp: Hệ thống gợi ý lai kết hợp Content-Based, Collaborative Filtering

## Phát triển hệ thống gợi ý lai (Hybrid Recommendation System)

- Kết hợp nhiều phương pháp: Hệ thống gợi ý lai kết hợp Content-Based, Collaborative Filtering



# KẾT LUẬN

Hệ thống gợi ý phim dựa trên nội dung (Content-Based Recommendation System) thực sự là một công cụ hữu ích để cá nhân hóa đề xuất phim cho người dùng

Khả năng hoạt động với ít thông tin: Vì hệ thống chủ yếu dựa vào dữ liệu về các bộ phim, nó có thể hoạt động hiệu quả ngay cả khi có ít thông tin về người dùng.







# THANK YOU!



**GVHD: TS.Nguyễn Gia Tuấn Anh**  
**CN.Trần Quốc Khánh**

**Trương Nhật Quang(22521207)**  
**Đỗ Tuấn Trức (22521548)**  
**Trương Lưu Song Tâm(22521294)**  
**Nguyễn Đức Tài(22521277)**

