

CHAT-REC: TOWARDS INTERACTIVE AND EXPLAINABLE LLMS-AUGMENTED RECOMMENDER SYSTEM

TOMÁS FOUYET - SEBASTIÁN PÉREZ - TOMÁS VERGARA

CONTEXTO DE LOS LLMS

Capacidades

- Inferencia Compleja
- Inferencia de conocimiento
- Robustez externa



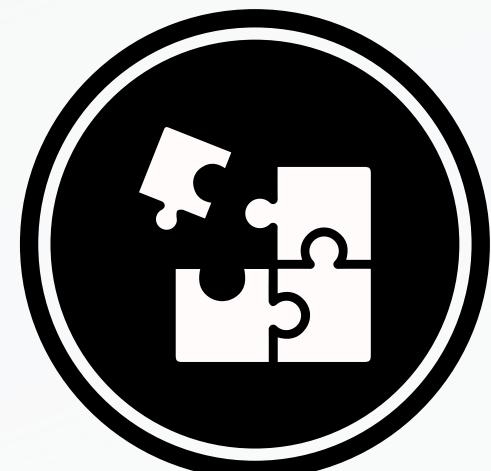
Objetivo

- Facilitar el proceso de aprendizaje
- Mantener parámetros sin ajustar
- Proporcionar contexto adecuado para su modelo



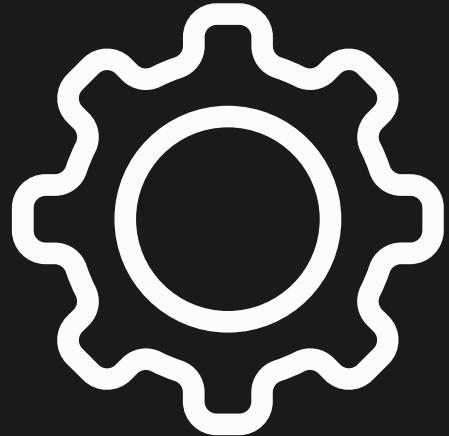
Técnicas

- In-Context Learning
- Chain-of-Thought
- Instruct Learning



LIMITACIONES ACTUALES

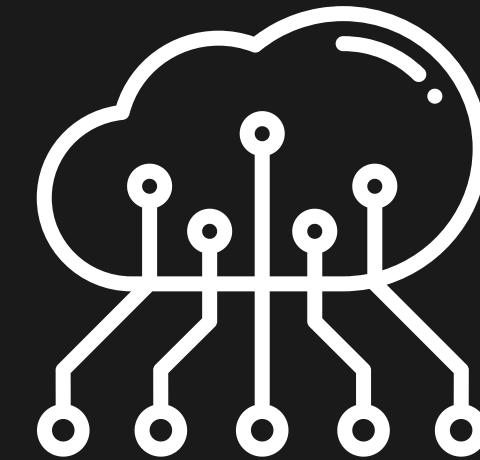
Funcionalidad



Cold Start



Multiples Dominios



- Interactividad
- Explicabilidad
- Retroalimentacion

- Problemas con usuarios y elementos nuevos

- Utilizar material externo
- Un set de entrenamiento extenso
- Conocimiento profundo sobre el tema en cuestión

RECOMENDACIÓN NLP



- Modelos ya entrenados a partir de grandes conjuntos de datos
- Aprenden a comprender el lenguaje en un nivel profundo, sin contexto

QUE SON ?



- Superar el problema de falta de datos, lo que evita el cuello de botella en los modelos
- Utilizar sus conocimientos para mejorar recomendaciones

PARA QUE ?



- Entender como los diferentes tipos de datos y entradas afectan y mejoran los RecSys.
- Como la elección de datos influye en el rendimiento

CHAT - REC

ChatGPT Augmented Recommender System

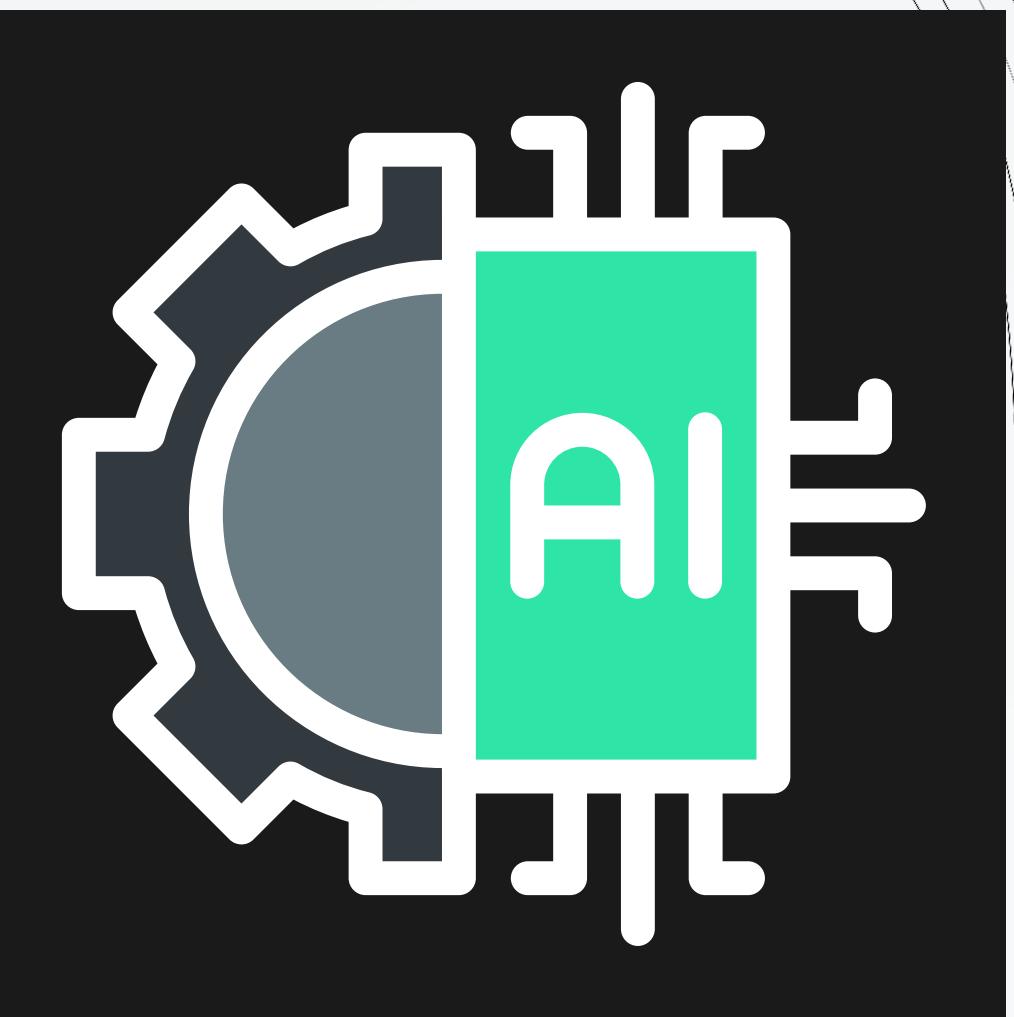
No requiere
entrenamiento

Aprender del
usuario
interactuando

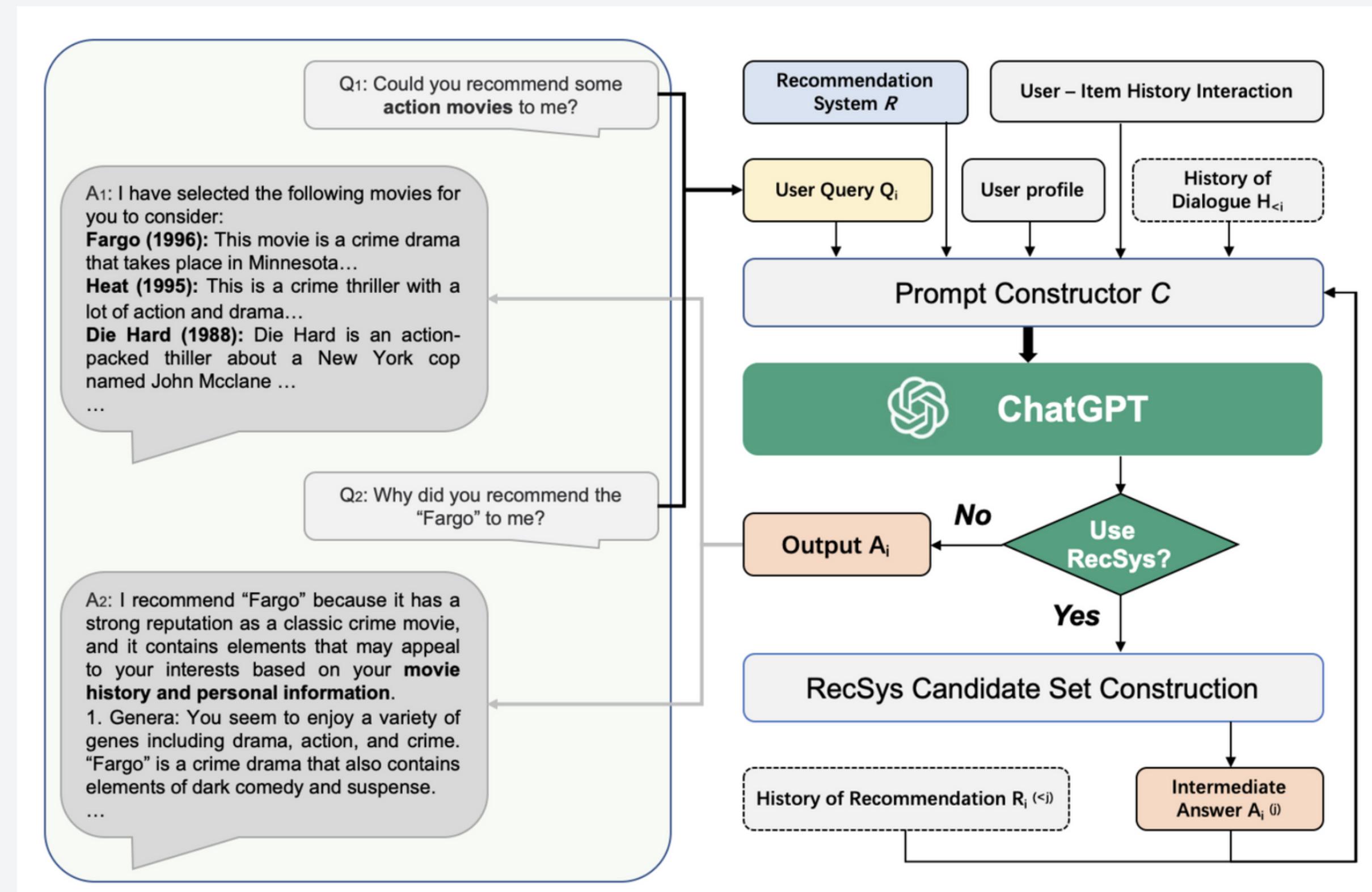
Buena
capacidad
entre dominios

ALMS

- Utilización de Herramientas externas
- Capacidad de razonamiento
- Mayor potencial de escalabilidad
- Toma de decisiones
- Mayor capacidad de interpretabilidad
- No necesitan de heurísticas



MÉTODO

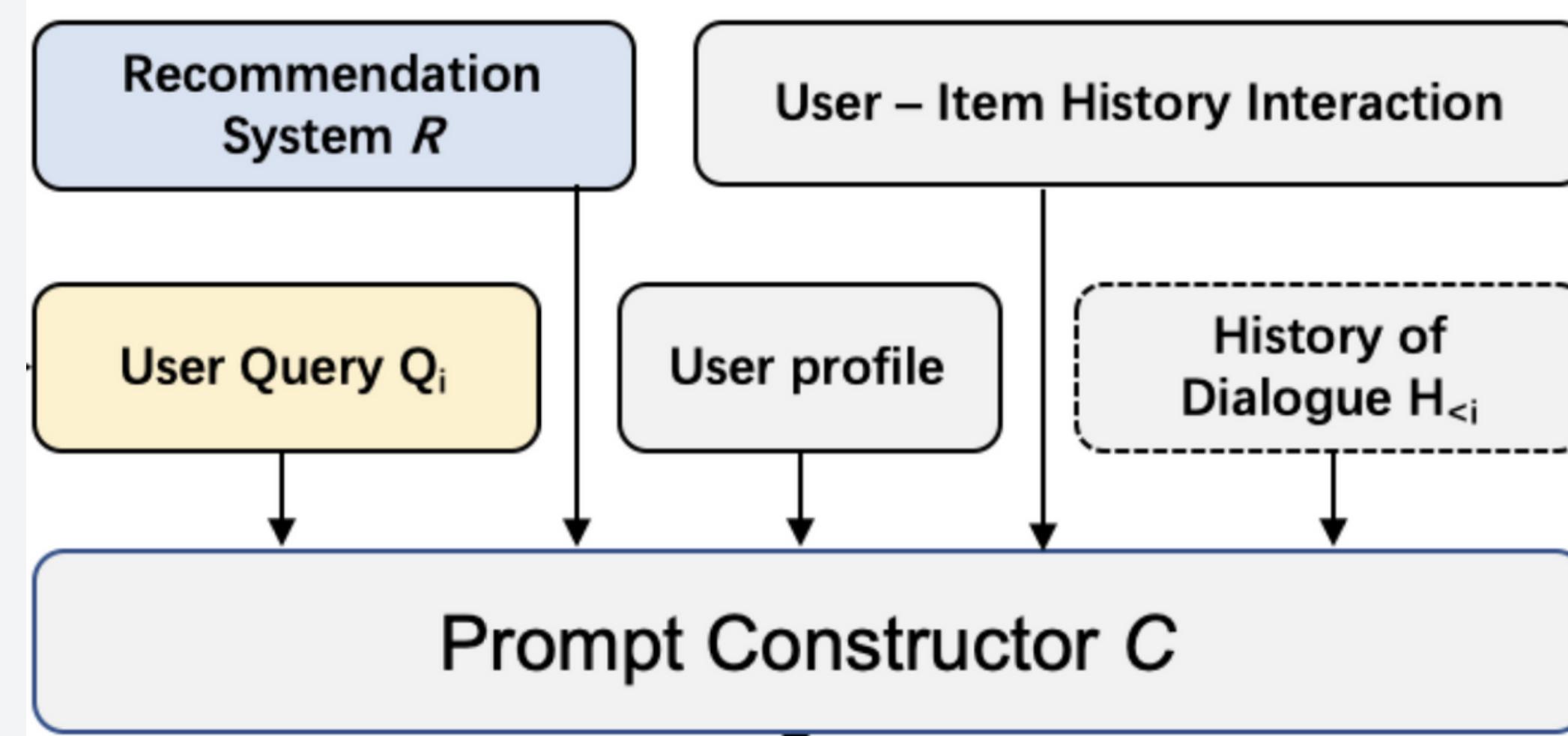


MÉTODO

Primero, llega una pregunta del usuario:

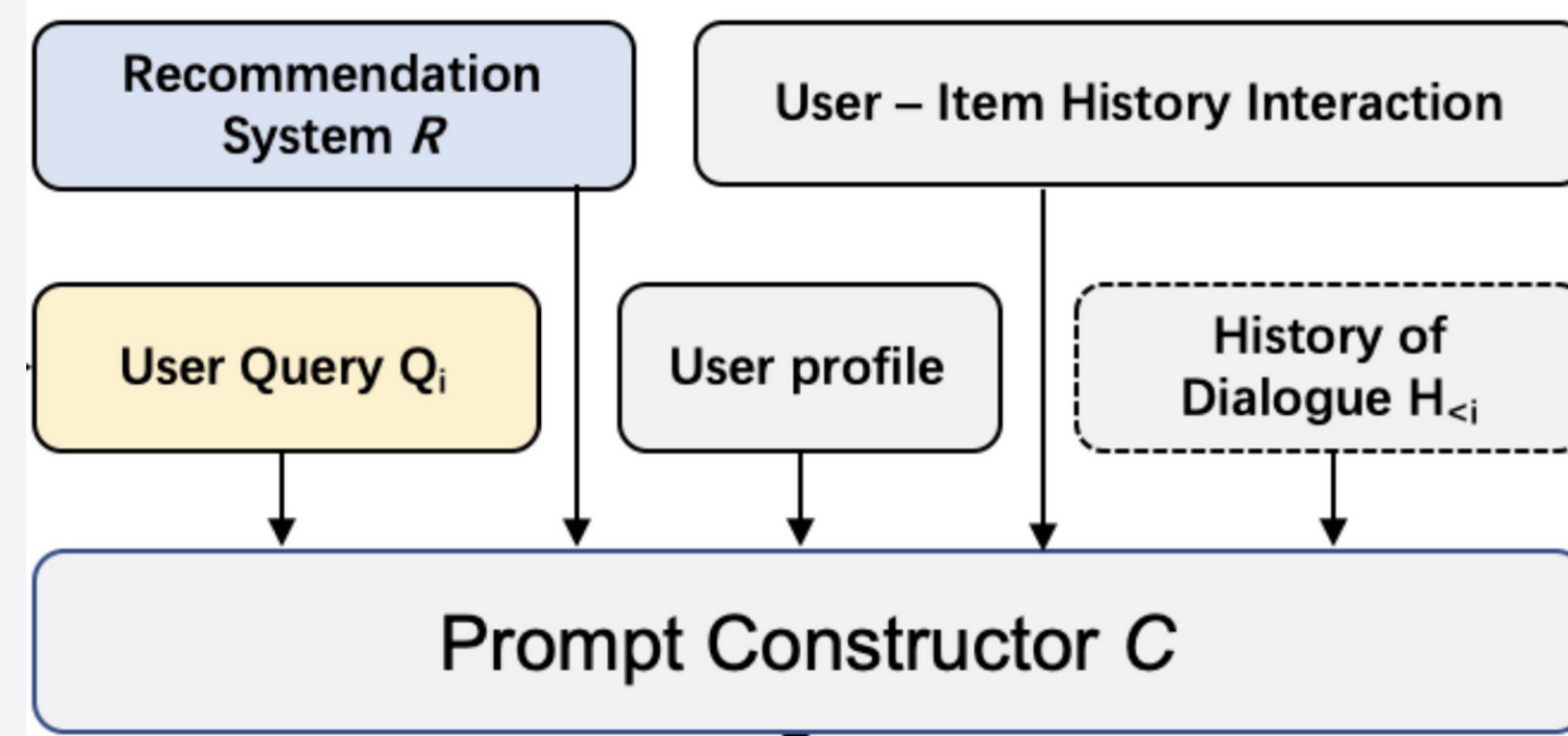
"Could you recommend some action movies to me?"

MÉTODO



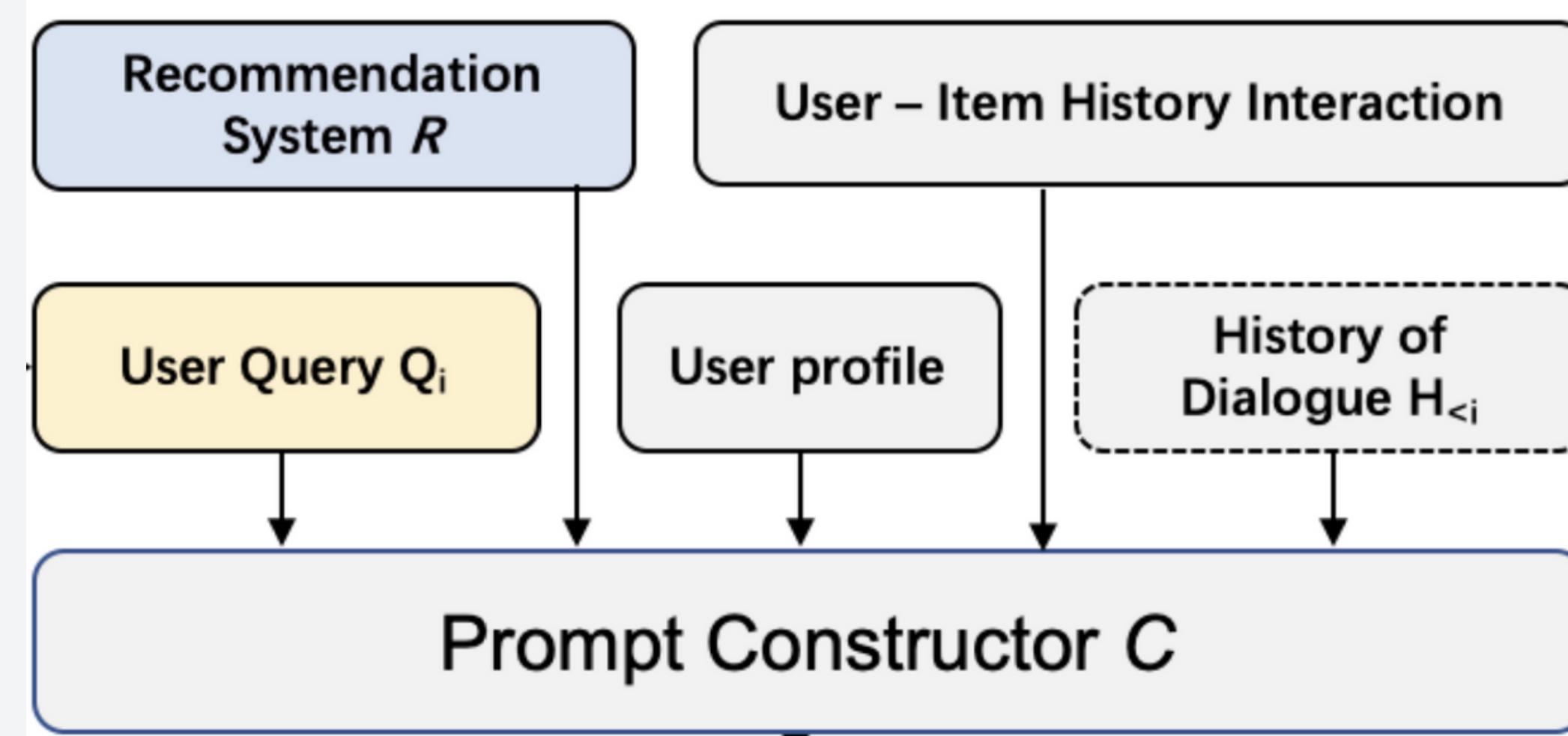
User Query: la pregunta del usuario

MÉTODO



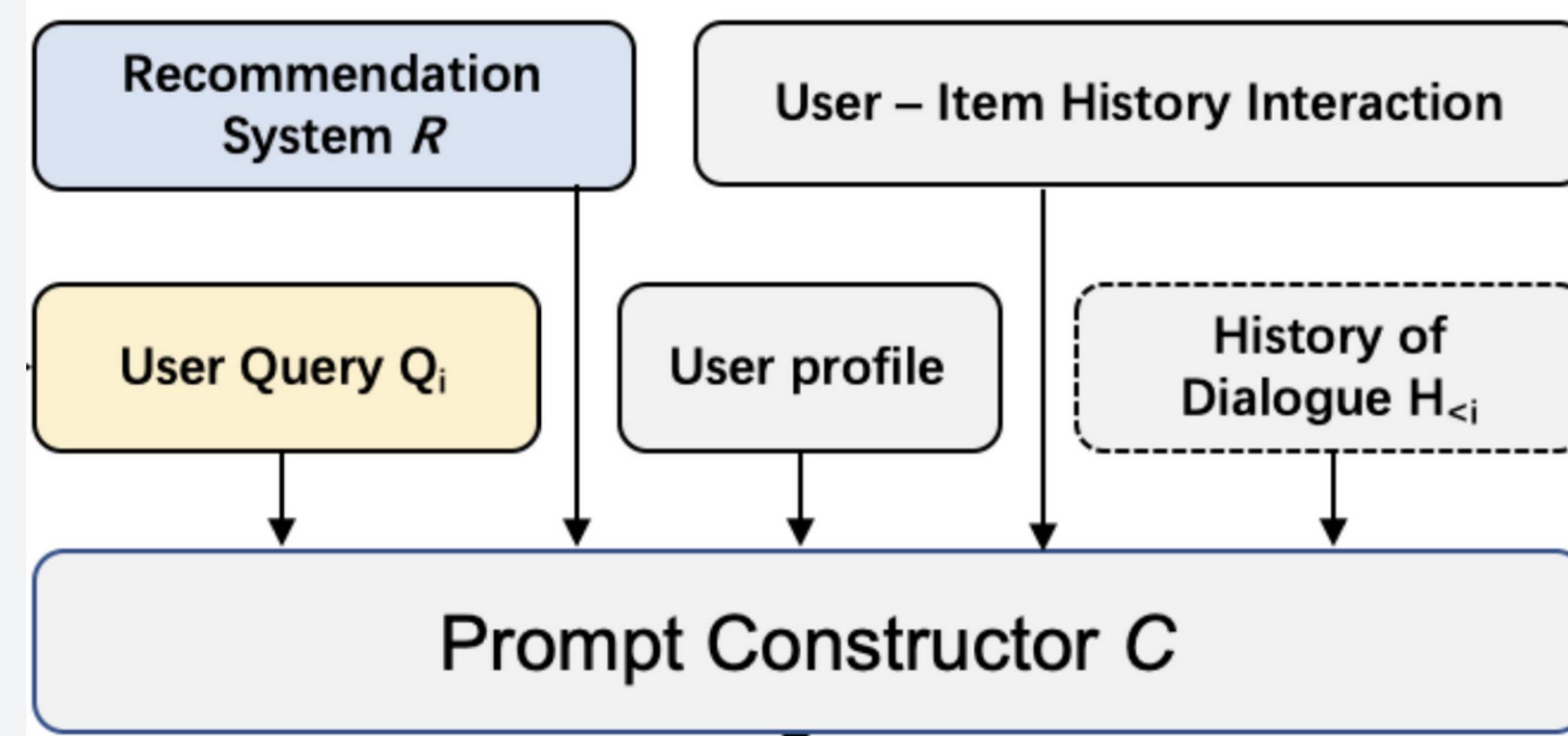
Recommendation System: Sistema de recomendación tradicional (MF, User-Based, etc.)

MÉTODO



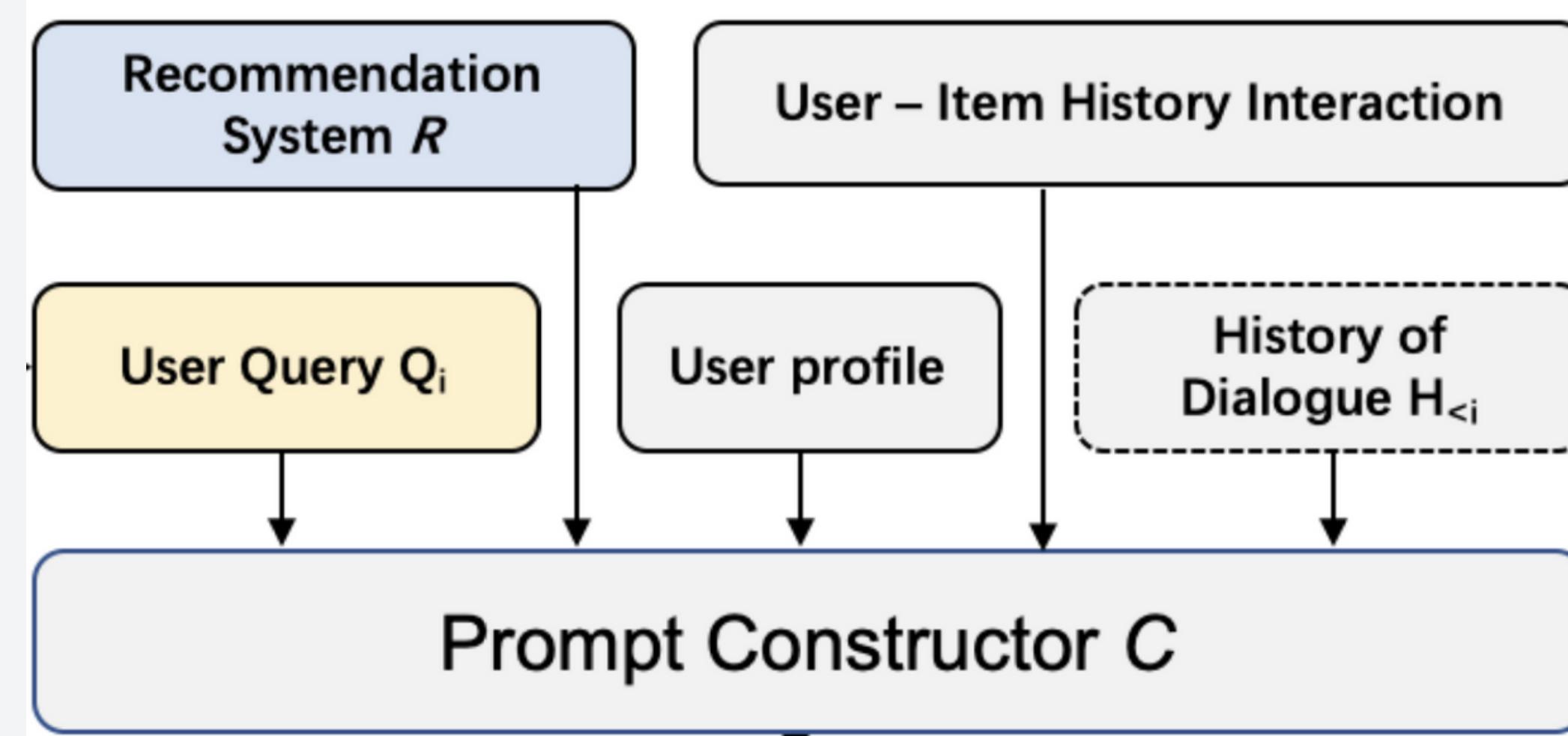
User Profile: Información que tengamos del usuario (edad, geografía, etc.).

MÉTODO



History of interactions: Historial de este mismo usuario

MÉTODO



History of dialogue: Historial de este mismo chat

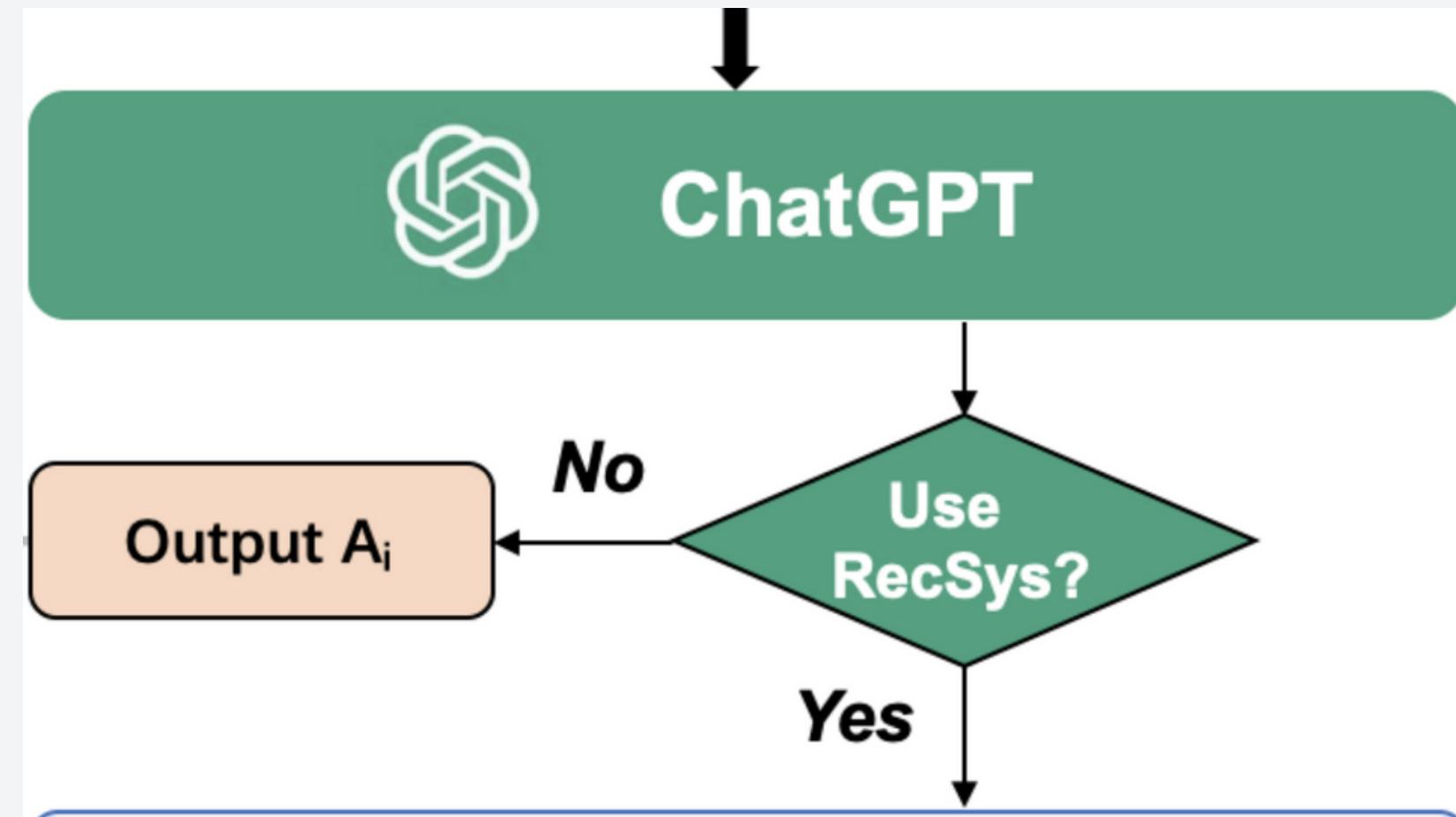
MÉTODO

Prompt construido: Dispone de toda la información para que ChatGPT pueda recomendar

"El usuario John Doe, **chileno** de 22 años, ha **visto la película X** le dio **5 estrellas**....

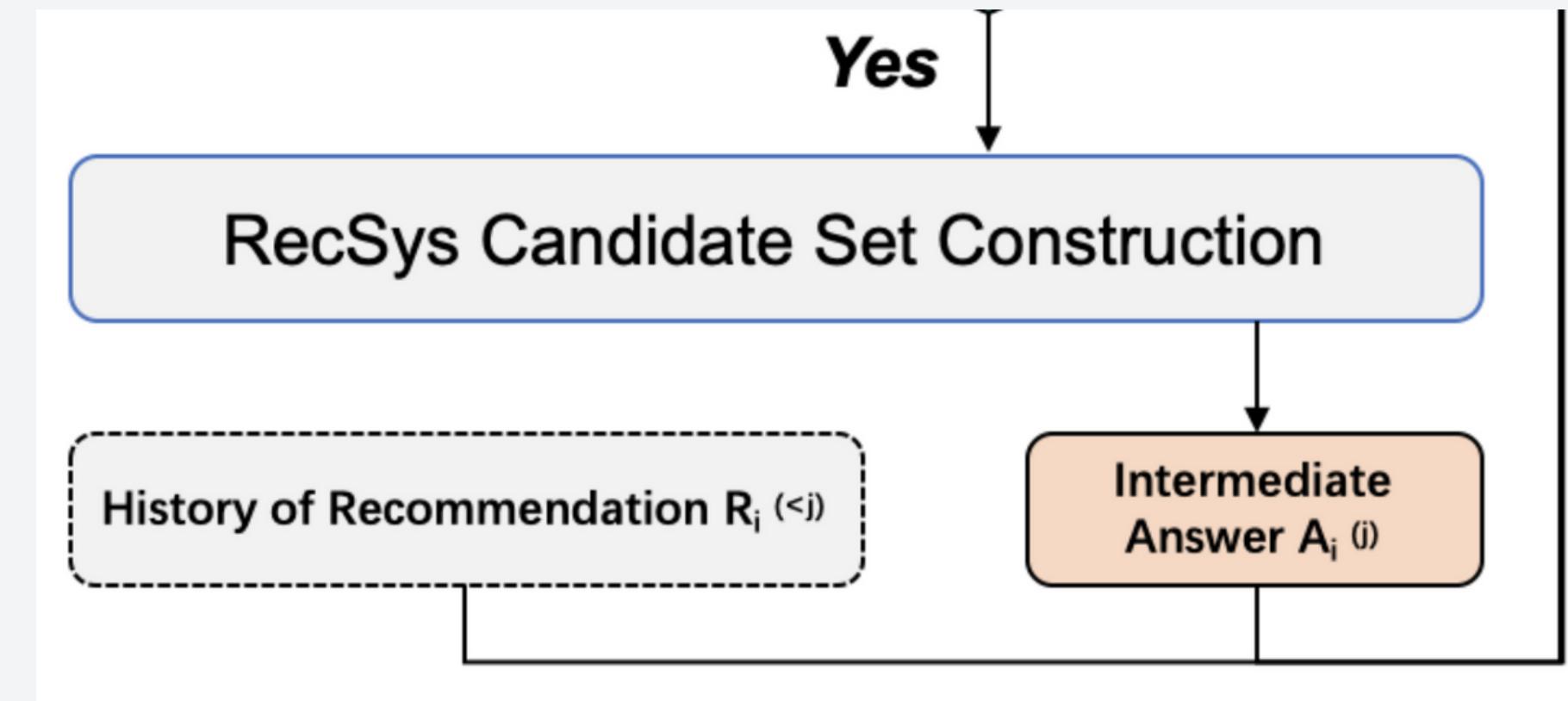
Ahora preguntó: ¿Qué películas de acción me recomiendas?"

MÉTODO



El modelo decide si debe o no usar el recomendador

MÉTODO



Si decide usar el recomendador, obtiene
(una gran cantidad) recomendaciones de
manera tradicional

MÉTODO

Estas recomendaciones se vuelven a pasar por ChatGPT, para filtrar y ordenar

"John Doe, chileno 22 años, que vio las películas A, B, C..."

Considerando las películas D, E, F...

¿Cuáles 5 de estas películas le recomendarías y por qué?"

MÉTODO

A1: I have selected the following movies for you to consider:

Fargo (1996): This movie is a crime drama that takes place in Minnesota...

Heat (1995): This is a crime thriller with a lot of action and drama...

Die Hard (1988): Die Hard is an action-packed thriller about a New York cop named John McClane ...

...

MÉTODO

Q1: Could you recommend some **action movies** to me?

A1: I have selected the following movies for you to consider:

Fargo (1996): This movie is a crime drama that takes place in Minnesota...

Heat (1995): This is a crime thriller with a lot of action and drama...

Die Hard (1988): Die Hard is an action-packed thriller about a New York cop named John McClane ...

...

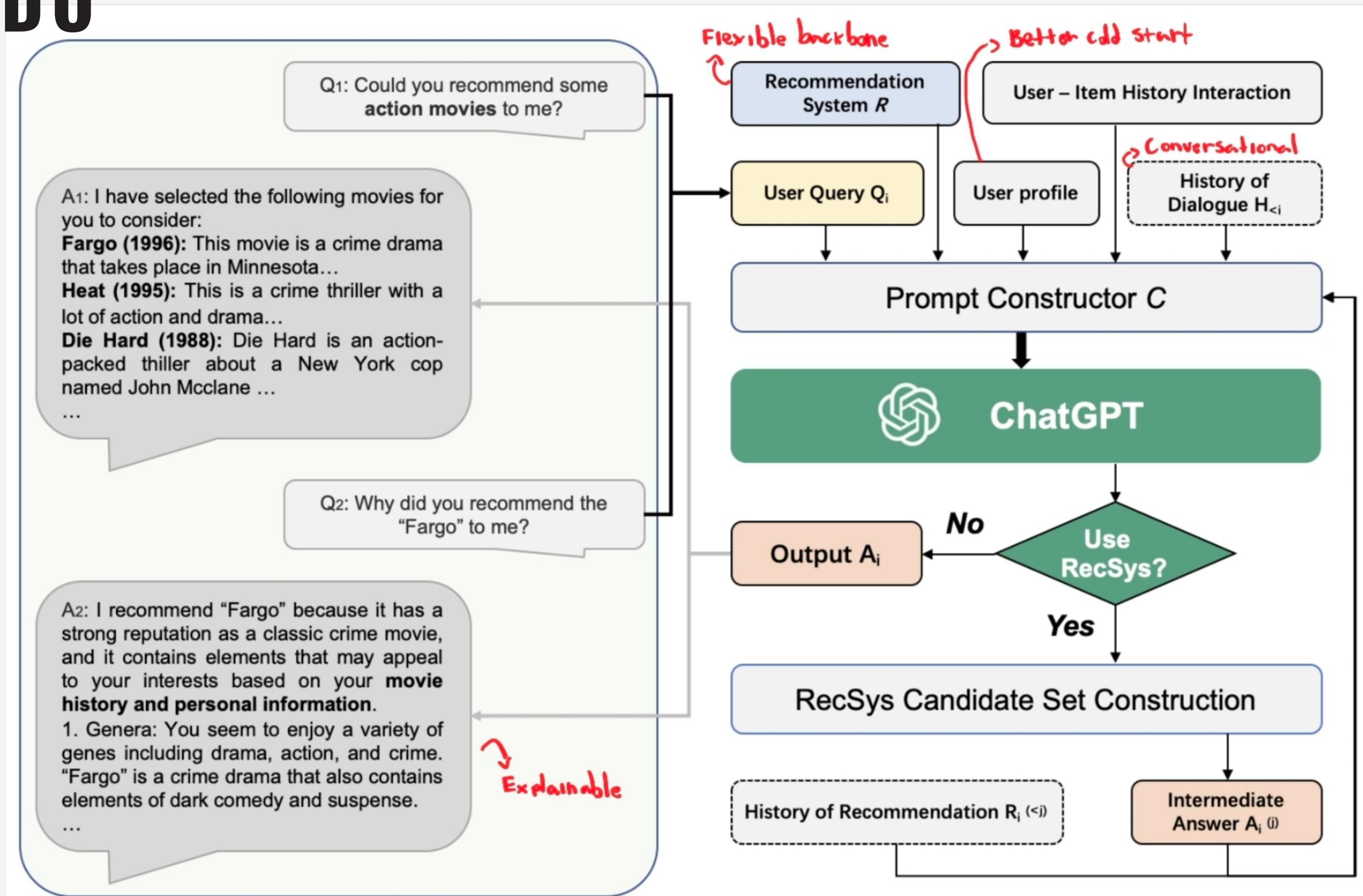
Q2: Why did you recommend the “Fargo” to me?

A2: I recommend “Fargo” because it has a strong reputation as a classic crime movie, and it contains elements that may appeal to your interests based on your **movie history and personal information**.

1. Genera: You seem to enjoy a variety of genres including drama, action, and crime. “Fargo” is a crime drama that also contains elements of dark comedy and suspense.

...

MÉTODO



MÉTODO

Incluso permite **Cross-Domain Recommendation**

“El usuario vio las **películas A, B, C...**

¿Qué **libros** recomiendas?”

EXPERIMENTO

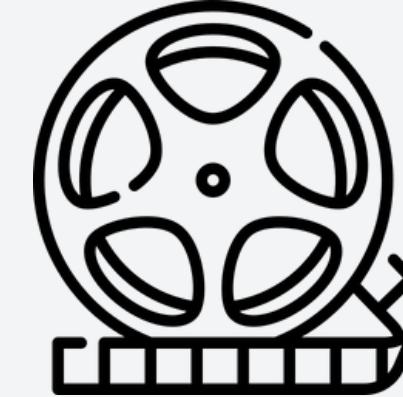
DATASET



MovieLens 100K



943
usuarios



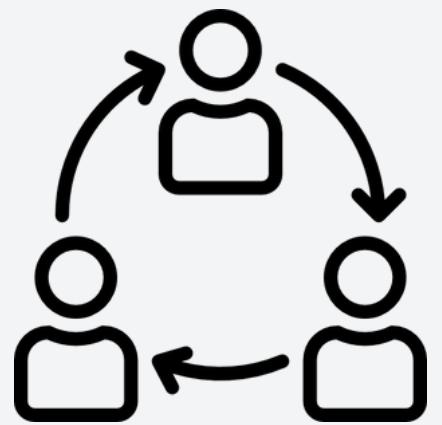
1682
películas



Ratings del
1 al 5

LINEA BASE

TOP-K



Light FM



Light GCN



Item-KNN

Predictión

$$\begin{Bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{Bmatrix}$$

MF

LINEA BASE



gpt-3.5-turbo

text-davinci-002

text-davinci-003

MÉTRICAS

TOP-K

Precisión

NDCG

Recall

Predicción

RMSE

MAE

RESULTADOS

TOP-K

Models	Precision	Recall	NDCG
LightFM	0.2830	0.1410	0.2846
LightGCN	0.3030	0.1455	0.3425
CHAT-REC (gpt-3.5-turbo)	0.3103	0.1279	0.3696
CHAT-REC (text-davinci-003)	0.3240 (+6.93%)	0.1404 (-3.51%)	0.3802 (+11.01%)
CHAT-REC (text-davinci-002)	0.3031	0.1240	0.3629

text-davinci-003 supera al resto en 2/3 métricas

CHAT-REC demuestra mejoras significativas frente a recomendadores clásicos

PREDICCIÓN

Models	RMSE	MAE
MF	0.988	0.771
Item-KNN	0.933	0.734
CHAT-REC (gpt-3.5-turbo)	0.969	0.756
CHAT-REC (text-davinci-003)	0.785	0.593
CHAT-REC (text-davinci-002)	0.8309	0.6215

text-davinci-003 supera al resto en 2/2 métricas

CHAT-REC demuestra mejoras significativas frente a recomendadores clásicos

OTRAS CONCLUSIONES

El rendimiento de **gpt-3.5-turbo** en tareas de predicción numérica es **más débil** que los otros dos modelos

La habilidad más importante de CHAT-REC es **optimizar el conjunto de candidatos** del sistema de recomendación

ESTUDIO CON DAVINCI-003

DISTINTOS PROMPTS

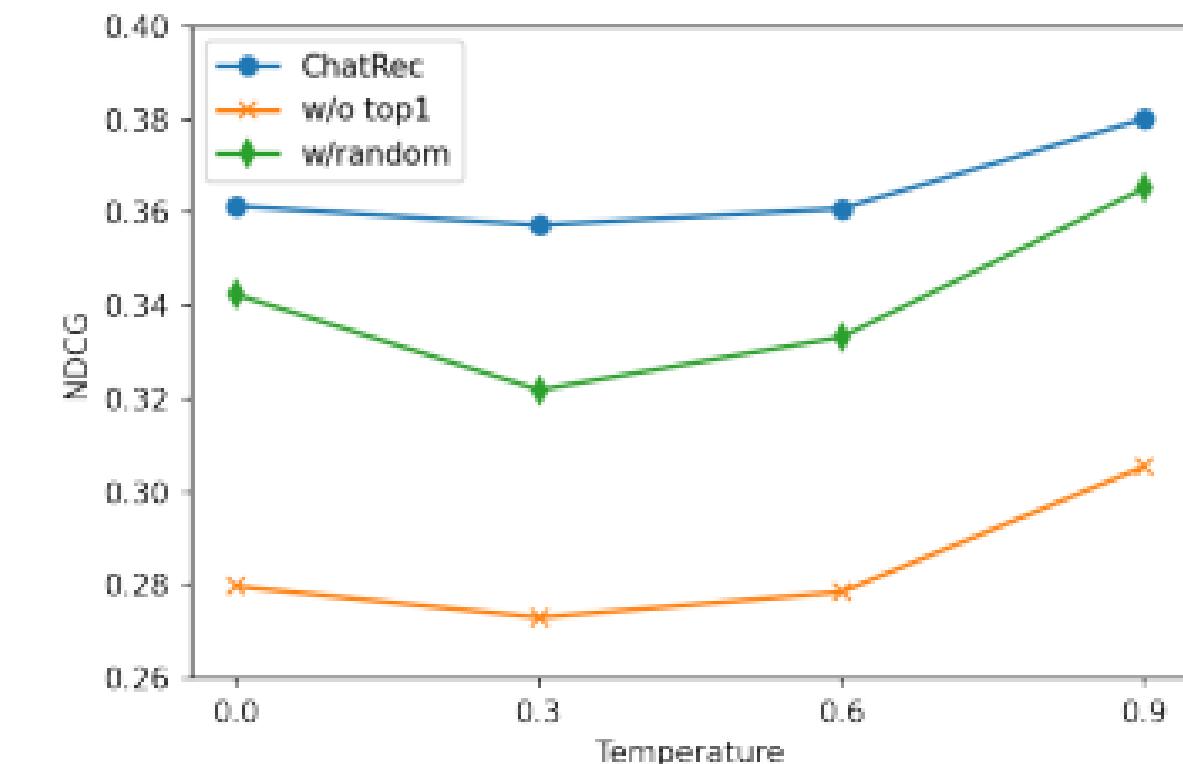
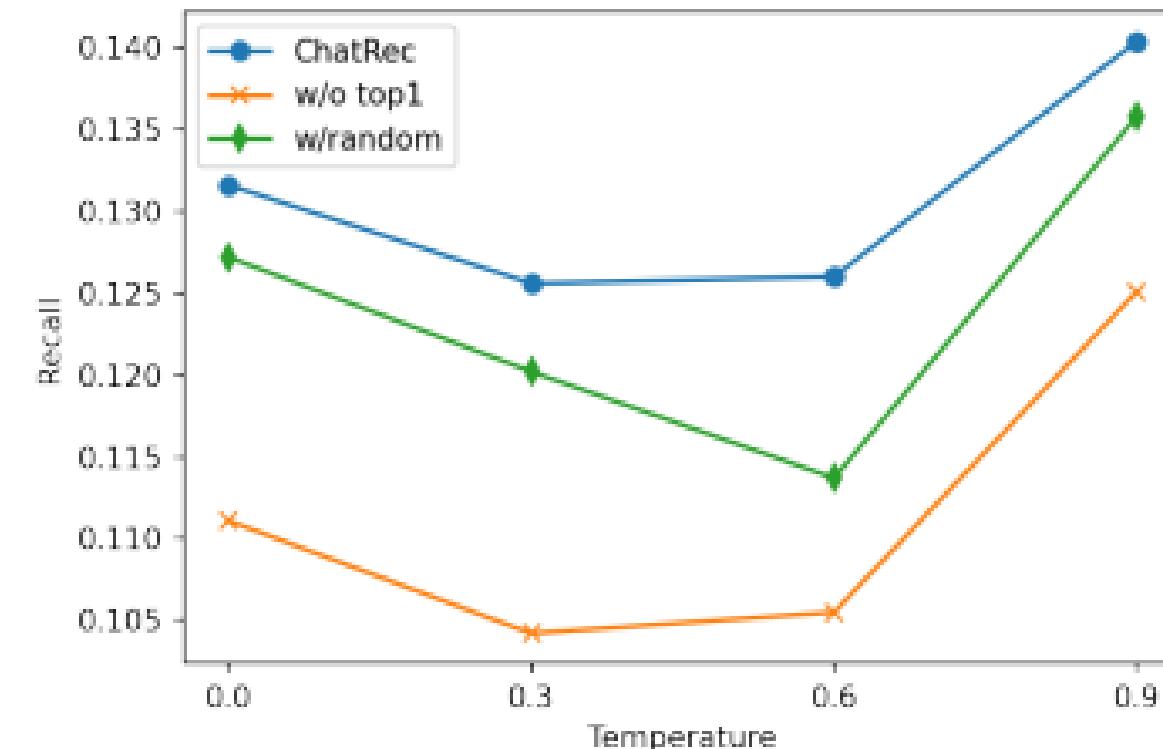
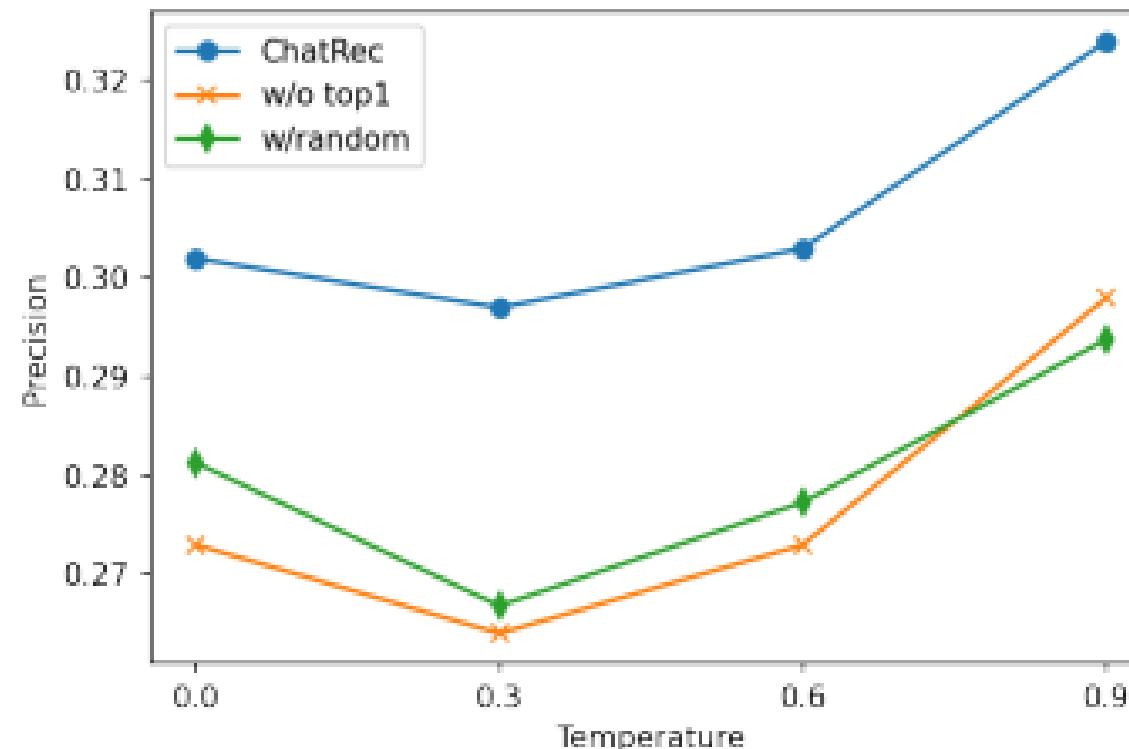
w/random

w/top1

DISTINTAS TEMPERATURAS

0.0 - 0.9

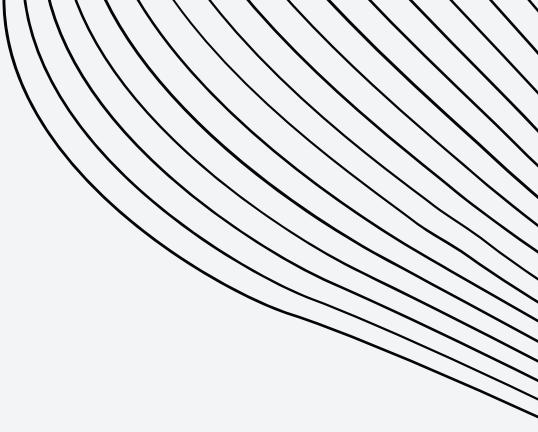
ESTUDIO CON DAVINCI-003



Mejor temperatura 0.9

Prompts sin top 1 o aleatorias
empeoran rendimiento

CONCLUSIÓN



CONCLUSIÓN

**CHAT-REC produce mejoras significativas en
recomendación top-K y predicción**

CONCLUSIÓN

CHAT-REC produce mejoras significativas en recomendación top-K y predicción

Potencial LLM's: Mejora interactividad, explicabilidad y recomendaciones cruzadas.

CONCLUSIÓN

CHAT-REC produce mejoras significativas en recomendación top-K y predicción

Potencial LLM's: Mejora interactividad, explicabilidad y recomendaciones cruzadas.

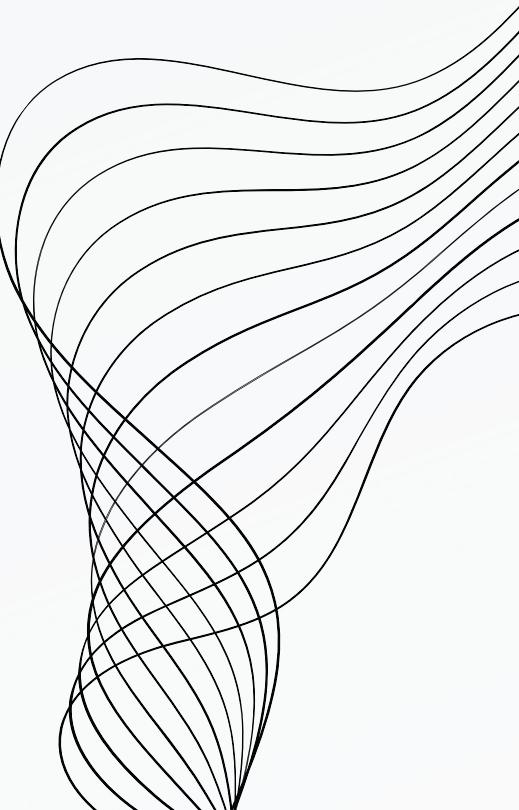
El prompt es fundamental para el rendimiento

CRÍTICAS

No mencionan qué sistema recomendador usan por atrás

Faltan razones de por qué seleccionaron esos 4 modelos clásicos para comparar

Algunos detalles de redacción y presentación



CHAT-REC: TOWARDS INTERACTIVE AND EXPLAINABLE LLMS-AUGMENTED RECOMMENDER SYSTEM

TOMÁS FOUYET - SEBASTIÁN PÉREZ - TOMÁS VERGARA