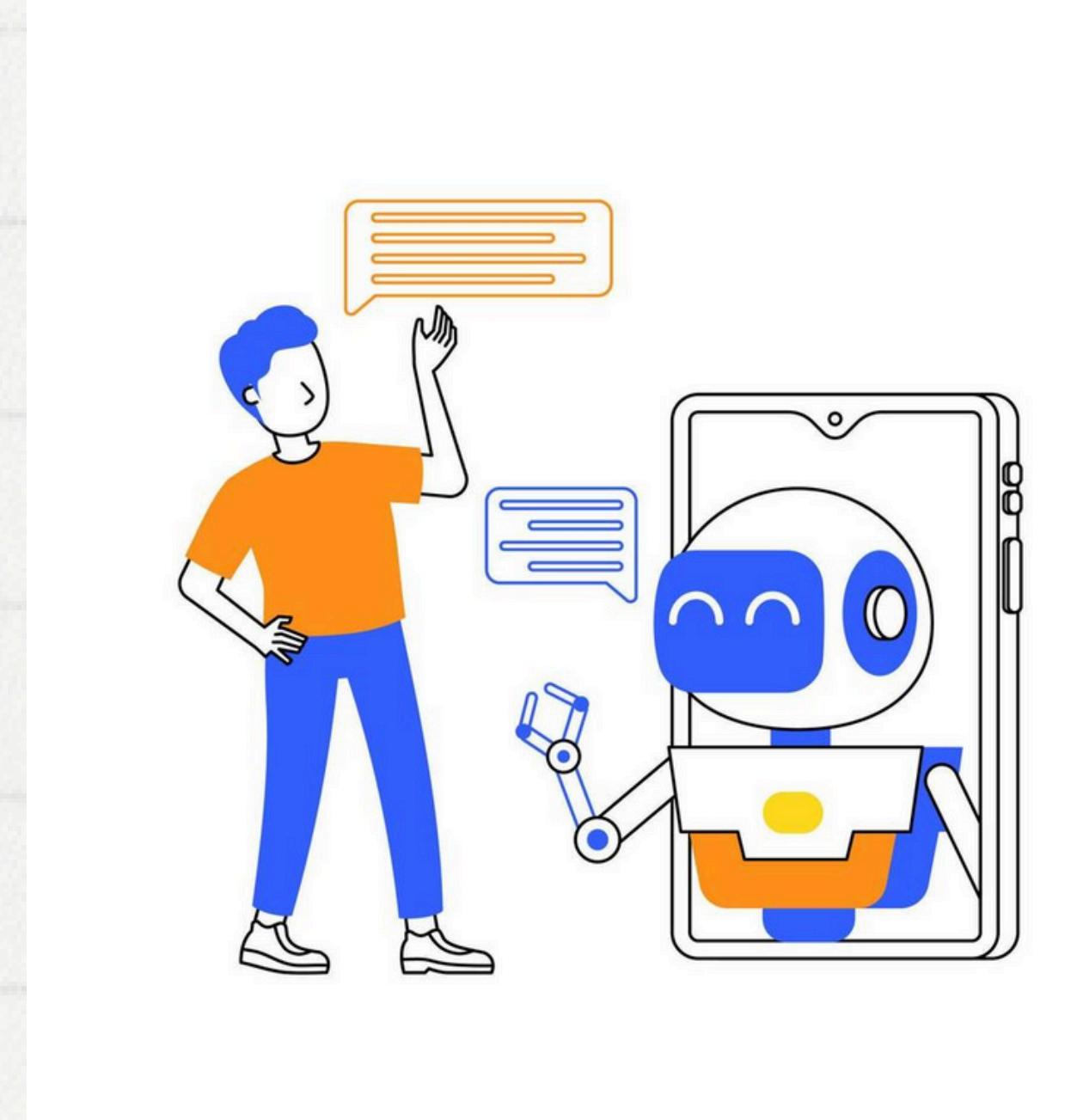


PEARL

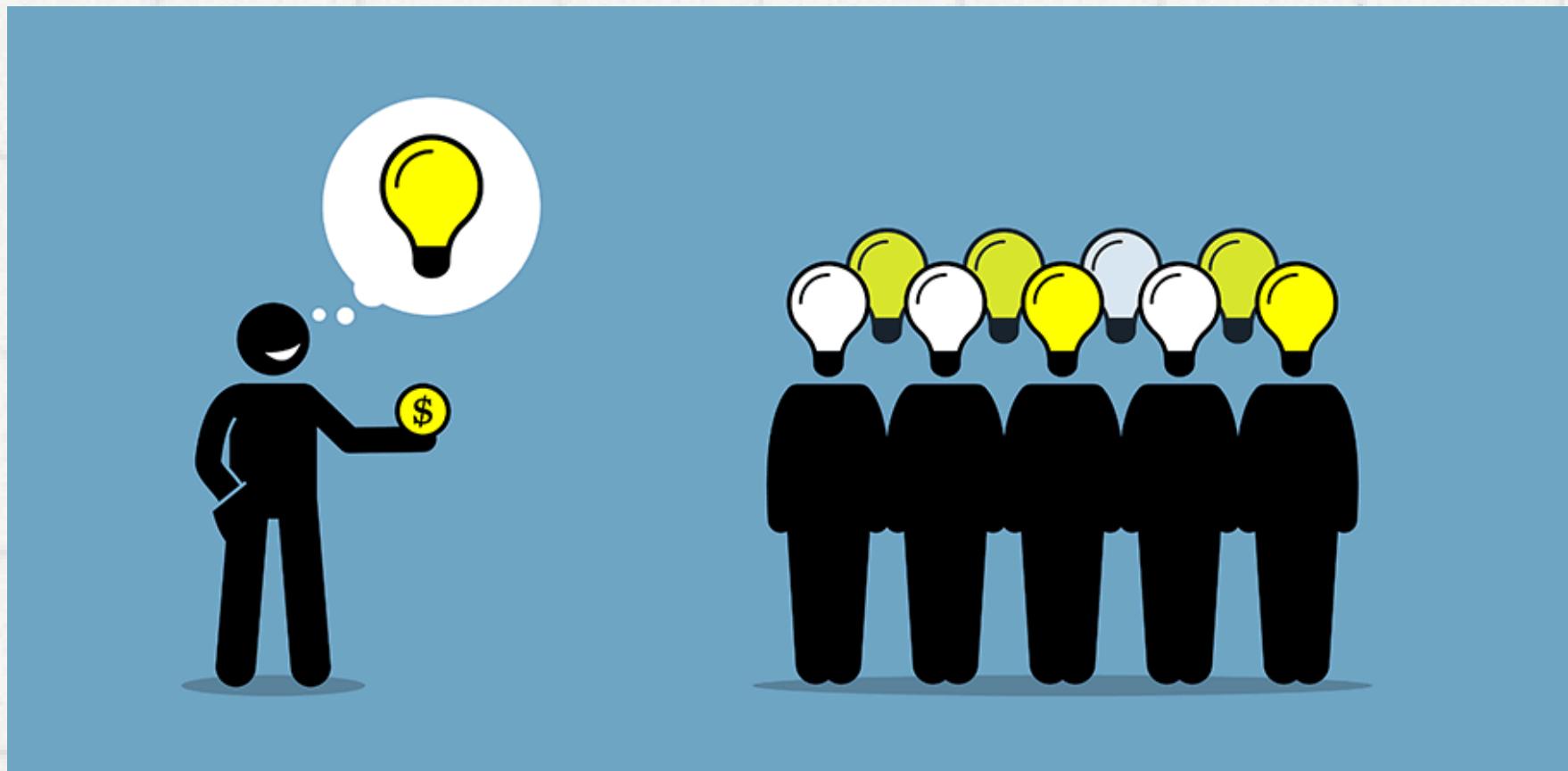
A Review-Driven Persona-Knowledge Grounded
Conversational Recommendation Dataset

Jorge Guzmán
Blanca Lamarca
Sebastián Silva

Contexto



¿Qué es el Crowdsourcing?



Problema de recomendación



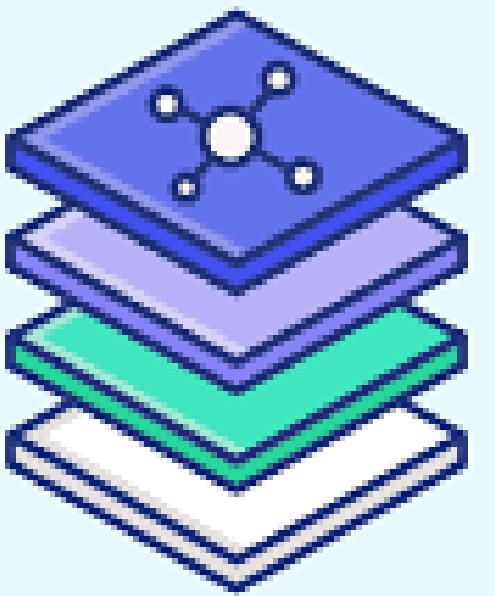
dreamlike tense comic book
zombies black and white world war ii serial killer
based on a true story bittersweet
nonlinear thought-provoking high school
cinematography time travel
historical parody visually appealing tim burton criterion
violence animation
adventure dark comedy imdb top 250
superhero nudity (topless) murder
horror disturbing satir crim
boring aliens
drugs space
humorous suspense war magic japan
british pixar
bd-r bill murray
music musical
psychological
religion robots
friendship mystery politics
sci-fi family
histor anim
psychology
twist ending
multiple storylines
predictable
quentin tarantino
black comedy
cult film
social commentary
johnny depp
nudity (full frontal)
artificial intelligence
thriller
tom hanks
mental illness
oscar (best picture)
coming of age
drama
bruce willis

romance action fantasy
atmospheric
comedy quirky
stylized sci-fi dystopia
psychology twist ending
classic adapted from book
multiple storylines
predictable
quentin tarantino
black comedy
cult film
social commentary
johnny depp
nudity (full frontal)
artificial intelligence
thriller
tom hanks
mental illness
oscar (best picture)
coming of age
drama
bruce willis

based on a book
atmospheric
comedy quirky
stylized sci-fi dystopia
psychology twist ending
classic adapted from book
multiple storylines
predictable
quentin tarantino
black comedy
cult film
social commentary
johnny depp
nudity (full frontal)
artificial intelligence
thriller
tom hanks
mental illness
oscar (best picture)
coming of age
drama
bruce willis

Contribución

PEARL



GPT-3.5

Estado del arte y marco teórico

Términos importantes

- Sistemas recomendadores conversacionales (CRS)
 - Conversación natural
 - Personalización
- Large Language Models (LLM)
- Persona
- Simulador de usuario y recomendador
- Utterance
- N-gram specificity



Trabajos relacionados

- Diálogos obtenidos con Crowdsourcing no son escalables
- Baja calidad

01.

ReDial (Li et al., 2018) y
INSPIRED (Hayati et al., 2020)
• Conversaciones y
diálogos de
recomendación

02.

Zhou et al. (2020b)
• Conversaciones human-machine

03.

He et al. (2023)
• Scraping de Reddit
Lu et al. (2023)
• Data-to-Text

Trabajos relacionados

- Diálogos sintetizados
- Uso de LLM

04.

Kim et al., 2022; Lu et al., 2023; Chen et al., 2023; Kim et al., 2023; Chae et al., 2023).

- Sintetizar diálogo con LLM para datasets

05.

Kim et al., 2022

- Blended Skill BotsTalk

06.

Chen et al., 2023

- Conversación diádicas y multipartite

Comparación datasets

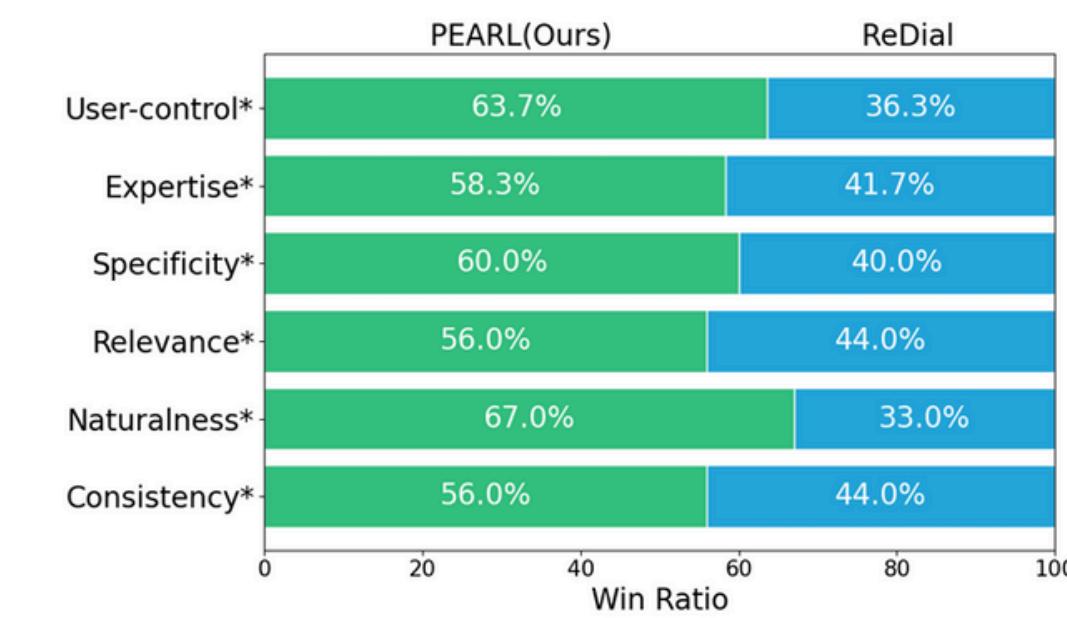


Figure 3: Results of human evaluation on head-to-head comparison between conversations sampled from PEARL and those from ReDial. (*: p-value < 0.05)

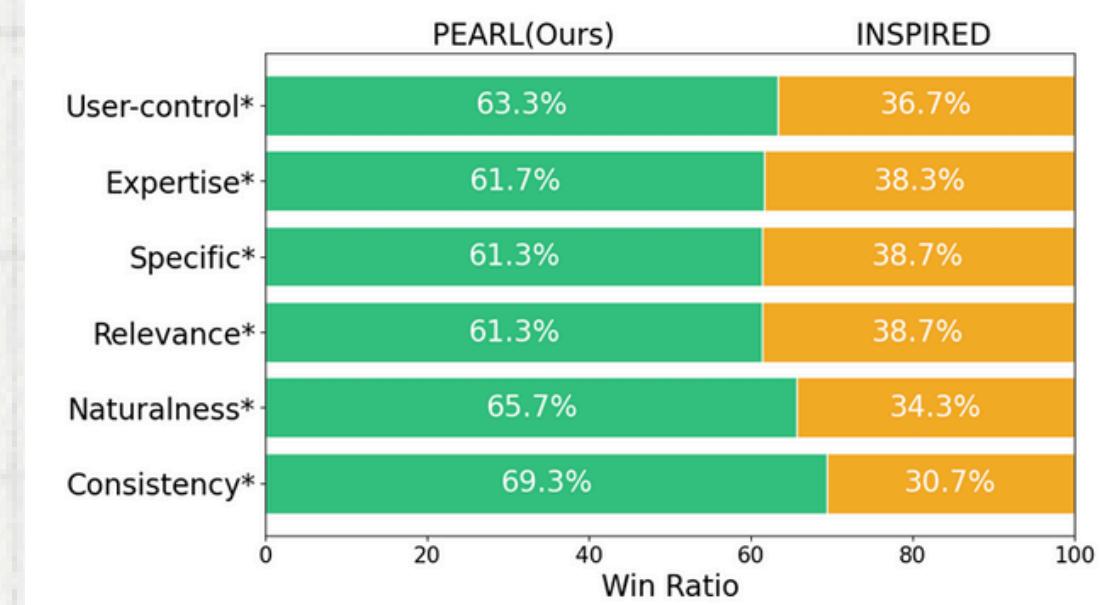


Figure 4: Results of human evaluation on head-to-head comparison between conversations sampled from PEARL and those from INSPIRED. (*: p-value < 0.05)

Evaluación humana

Comparación datasets

	ReDial	INSPIRED	PEARL
# of dialogues	10,006	1,001	57,277
# of utterances	182,150	35,811	548,061
2-gram specificity	65.44	119.56	141.79
3-gram specificity	65.97	123.01	149.75
4-gram specificity	65.37	122.81	153.00

Table 3: Statistics of PEARL compared to ReDial and INSPIRED. The table shows the number of dialogues, utterances, and n-gram specificities for each dataset.

Inter-dialogue similarity of user utterances	
Ours	0.1900
w/o Persona	0.1962

Table 4: Inter-dialogue similarity of PEARL and ablated dialogues generated without persona.

	ReDial	INSPIRED	PEARL
# of words	11.01	14.62	38.81

Table 5: Average number of words per recommender utterance in ReDial, INSPIRED, and PEARL.

Análisis de preferencia de usuarios
Análisis de recomendaciones Knowledge-Augmented
Escala de datos
Tiempo de obtención de datos

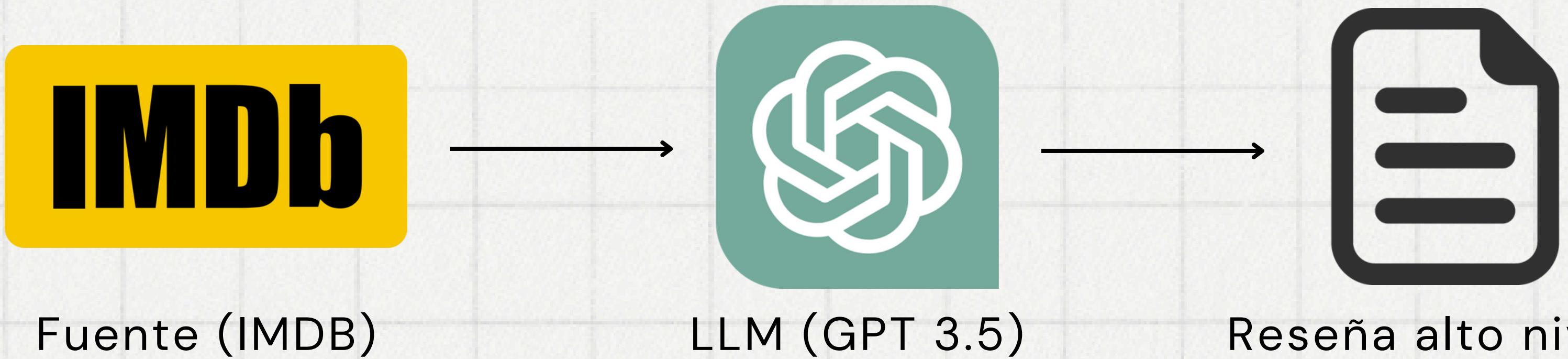
Detalle solución

Construcción de Pearl

- ❖ Recopilación de reseñas reales
- ❖ Generación de un simulador de usuarios
- ❖ Generación de un simulador de recomendador
- ❖ Generación de diálogos entre los simuladores
- ❖ Filtrado de los diálogos

Recopilación de reseñas reales

Datos para simuladores de usuario y recomendador



Simulador de Usuarios

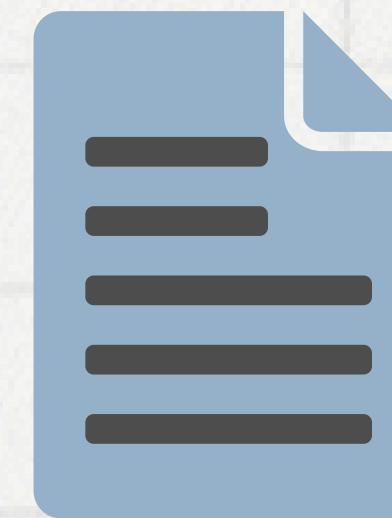
Simula los gustos de un usuario real y expresa los atributos que espera ver en la siguiente película

tres reseñas aleatorias



Simular gustos

reseña con calificación superior a ocho



Atributos deseados

Simulador de Recomendador

Utilizando un retriever y un LLM busca dar una recomendación acertada según las características solicitadas por el usuario

busca los items más relevantes



Retriever

según gustos del usuario
elige que item recomendar



LLM (GPT 3.5)

Generación de diálogos

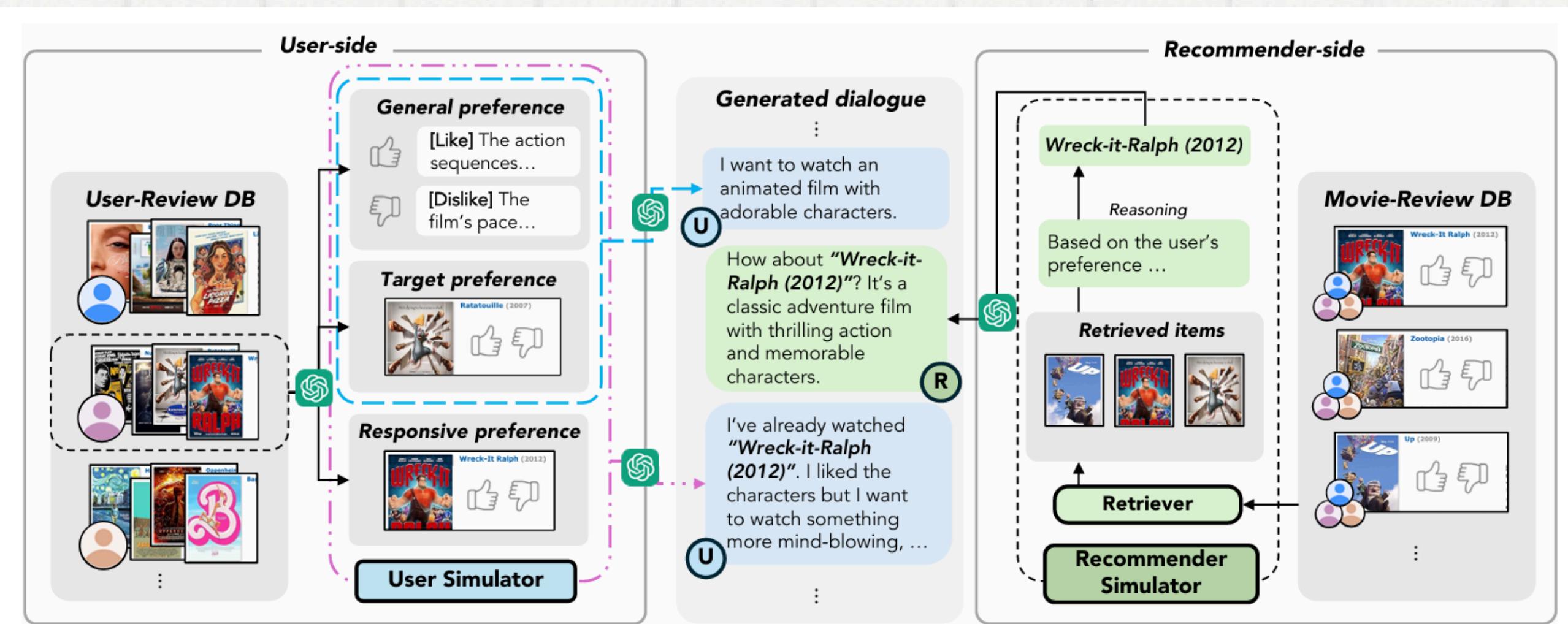
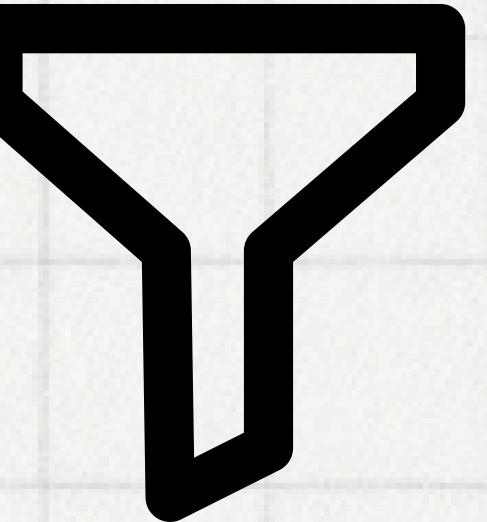


Figure 2: The overview of PEARL construction method. We synthesize recommendation dialogues with review-driven persona-knowledge grounded simulators. Specifically, our user simulator is equipped with persona and our recommender simulator is augmented by knowledge derived from reviews.

Filtrado de diálogos

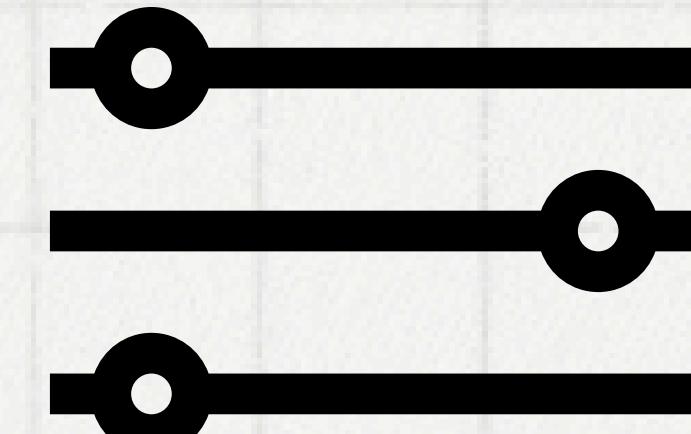
Para garantizar la calidad del dataset se filtraron los diálogos según distintos criterios

filtra diálogos con sesgos o errores



Filtros básicos

según gustos del usuario
filtra diálogos desacertados



Filtros de preferencia

Benchmarks

Evaluación automática muestra alta diversidad por parte del recomendador, pero un bajo rendimiento al medir el *recall*

Model	Dist-3	Dist-4
BART-ReDial	0.6220	0.5057
BART-PEARL	0.9241	0.8861
UniCRS-ReDial	0.5413	0.3667
UniCRS-PEARL	0.9338	0.9007
PECRS-ReDial	0.6798	0.5906
PECRS-PEARL	0.9132	0.8947
GPT-3.5	0.9256	0.8910

Table 7: Response generation performances on INSPIRED.

Model	R@1	R@10	R@50
BERT-PEARL	0.0018	0.0208	0.0736
UniCRS-PEARL	0.0310	0.0697	0.1202
PECRS-PEARL	0.0151	0.0339	0.0798
GPT-3.5	0.0071	0.0355	0.0709

Table 8: Recommendation performances on PEARL.

Diversidad

Recall

Referencias

- Hyungjoo Chae, Yongho Song, Kai Ong, Taeyoon Kwon, Minjin Kim, Youngjae Yu, Dongha Lee, Dongyeop Kang, and Jinyoung Yeo. 2023. Dialogue chain-of-thought distillation for commonsense-aware conversational agents. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, pages 5606–5632.
- Maximillian Chen, Alexandros Papangelis, Chenyang Tao, Seokhwan Kim, Andy Rosenbaum, Yang Liu, Zhou Yu, and Dilek Hakkani-Tur. 2023. PLACES: Prompting language models for social conversation synthesis. In Findings of the Association for Computational Linguistics: EACL 2023, pages 844–868, Dubrovnik, Croatia. Association for Computational Linguistics.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Shuyu Guo, Shuo Zhang, Weiwei Sun, Pengjie Ren, Zhumin Chen, and Zhaochun Ren. 2023. Towards explainable conversational recommender systems. In Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 2786–2795.

Anexos

Figuras

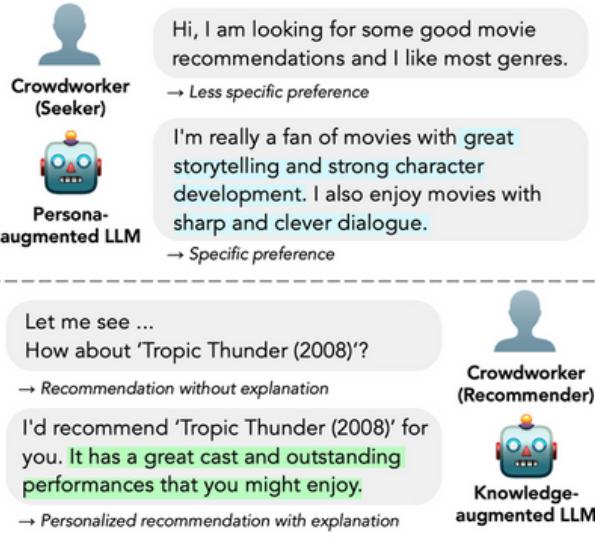


Figure 1: An example comparing utterances of crowdworkers and our persona-knowledge augmented LLM simulators.

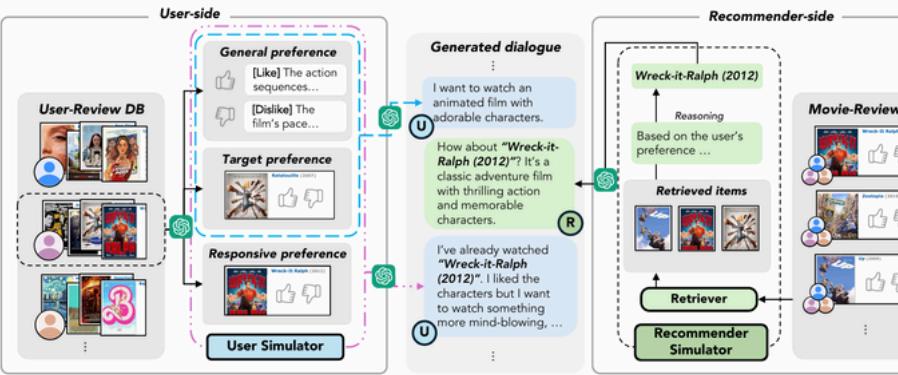


Figure 2: The overview of PEARL construction method. We synthesize recommendation dialogues with review-driven persona-knowledge grounded simulators. Specifically, our user simulator is equipped with persona and our recommender simulator is augmented by knowledge derived from reviews.

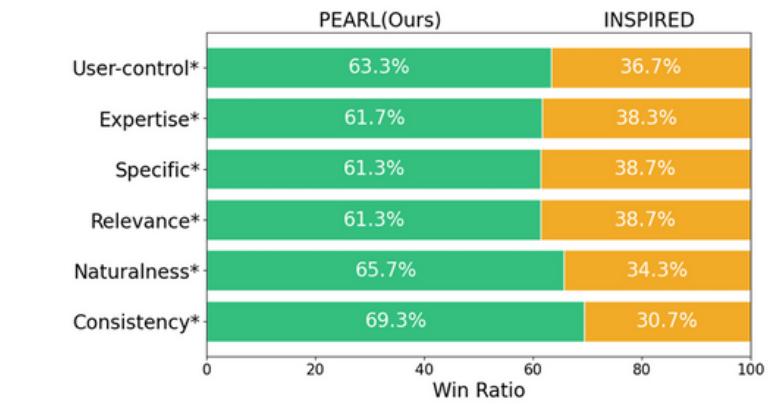


Figure 4: Results of human evaluation on head-to-head comparison between conversations sampled from PEARL and those from INSPIRED. (*: p-value < 0.05)

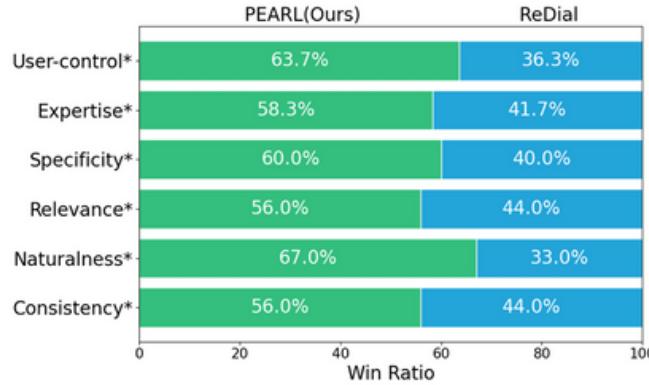


Figure 3: Results of human evaluation on head-to-head comparison between conversations sampled from PEARL and those from ReDial. (*: p-value < 0.05)

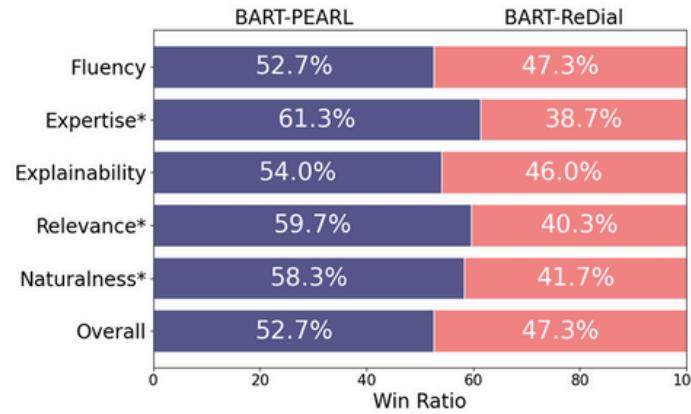


Figure 5: Results of head-to-head comparison human evaluation between responses generated from BART trained on PEARL and on ReDial. (*: p-value < 0.05)

Tablas

	ReDial	INSPIRED	PEARL
# of dialogues	10,006	1,001	57,277
# of utterances	182,150	35,811	548,061
2-gram specificity	65.44	119.56	141.79
3-gram specificity	65.97	123.01	149.75
4-gram specificity	65.37	122.81	153.00

Table 3: Statistics of PEARL compared to ReDial and INSPIRED. The table shows the number of dialogues, utterances, and n-gram specificities for each dataset.

	ReDial	INSPIRED	PEARL
# of words	11.01	14.62	38.81

Table 5: Average number of words per recommender utterance in ReDial, INSPIRED, and PEARL.

	PEARL (this work)	ReDial (Li et al., 2018)	INSPIRED (Hayati et al., 2020)	TG-ReDial (Zhou et al., 2020b)	DuRecDial 2.0 (Liu et al., 2021)
Collection method	Synthesized	Crowdsourced	Crowdsourced	Human-Machine	Crowdsourced
Real-world persona	✓	✗	✗	✓	✗
Explained recommendation	✓	✗	✗	✗	✗
Number of dialogues	57,277	10,006	1,001	10,000	16,482
Number of users	4,680	956	1,594	1,482	2,714
Number of utterances	548,061	182,150	35,811	129,392	255,346

Table 1: A comparison of our synthesized dataset to notable conversational recommendation datasets.

Model	R@1	R@10	R@50
BERT-PEARL	0.0018	0.0208	0.0736
UniCRS-PEARL	0.0310	0.0697	0.1202
PECRS-PEARL	0.0151	0.0339	0.0798
GPT-3.5	0.0071	0.0355	0.0709

Table 8: Recommendation performances on PEARL.

Dialogue context
...
Seeker: I like both, but I'm looking for something super hero related, like Guardians of the Galaxy.
Recommender: Oh! I have a few ideas of what you have in mind. Are comic book characters interesting at all?
Seeker: Yes! I love comic book characters, like Spider man, Ironman and Captain America.
Recommender: Okay. Would you like to watch this trailer for the Avengers?
Seeker: Oh that would be great!

Model response
GPT-3.5: I would recommend you watch the trailer for "Avengers: Endgame." It features all your favorite comic book characters like Iron Man, Captain America, and Spider-Man, and it's a super hero-related movie just like Guardians of the Galaxy. Enjoy watching the trailer!
BART-ReDial: Okay. I would also recommend 'The Avengers (2012)'

BART-PEARL (Ours): How about 'The Avengers (2012)'? It's an action-packed superhero movie with a great cast and some really interesting characters.

Ground-truth: I haven't actually watched it yet. It is the one with that Thanos character and the Justice league, right?

Table 10: Lemon-picked sample responses from BART-ReDial and BART-PEARL to a context in INSPIRED. The full example is in Table 17.

Seeker: Hi there! I'm in the mood to watch a movie. Can you recommend something?
Recommender: Absolutely! What kind of movie are you in the mood for? Any specific genre or theme?
Seeker: I'm generally a fan of movies with strong performances by the cast and enthralling cinematic accomplishments. I also appreciate well-written plots and screenplays.
Recommender: I think you would really enjoy 'Darkest Hour' (2017). It has a strong performance by Gary Oldman as Winston Churchill and the attention to detail in bringing the period to life is enthralling. Plus, the focus on Churchill's point of view and how he came to power makes for a compelling and well-written plot.
Seeker: I can see how 'Darkest Hour' fits some of my preferences, but I'm really in the mood for something with more action and combat. I also enjoy movies with historical accuracy and realistic depictions of challenges faced by characters in combat situations. Do you have anything that fits those criteria?
Recommender: How about watching 'Godzilla (1998)'? It has great action sequences involving Godzilla and the movie stays faithful to the original 1950's monster concept. The special effects are impressive and well done, and the film balances action, sci-fi, and thriller elements. Plus, it's based on historical events and provides a realistic depiction of characters facing challenges in combat situations.
Seeker: I appreciate the recommendation, but 'Godzilla (1998)' isn't quite what I'm in the mood for right now. I'm really looking for something with a focus on the life of an Army crew and the challenges they face in combat. Do you have anything else in mind that fits that description?
Recommender: I think 'Fury (2014)' would be a perfect fit for what you're looking for. It provides a realistic depiction of the challenges faced by an Army Sherman tank crew in combat situations. The performances by the cast are strong, and the screenplay is well-written, delivering a compelling and enthralling cinematic experience.
Seeker: That sounds like exactly what I'm looking for! I'll definitely give 'Fury (2014)' a watch. Thanks for the recommendation!

Table 2: Turn-by-turn dialogue sample from PEARL.

Inter-dialogue similarity of user utterances		
Ours	0.1900	
w/o Persona	0.1962	

Table 4: Inter-dialogue similarity of PEARL and ablated dialogues generated without persona.

Model	Dist-3	Dist-4
BART-ReDial	0.6220	0.5057
BART-PEARL	0.9241	0.8861
UniCRS-ReDial	0.5413	0.3667
UniCRS-PEARL	0.9338	0.9007
PECRS-ReDial	0.6798	0.5906
PECRS-PEARL	0.9132	0.8947
GPT-3.5	0.9256	0.8910

Table 7: Response generation performances on INSPIRED.

Implementación

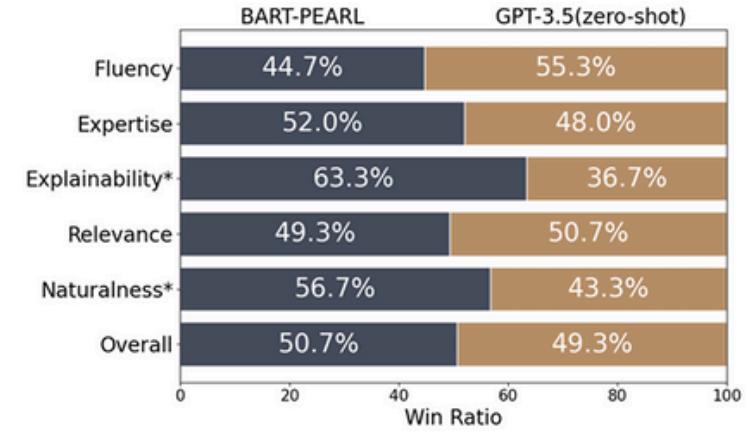


Figure 6: Results of head-to-head comparison human evaluation between responses generated from BART trained on PEARL and from GPT-3.5 zero-shot inference. (*: p-value < 0.05)

Model	ROUGE-1	ROUGE-2	Dist-1	Dist-2	Dist-3	Dist-4
BART-ReDial	0.1370	0.0264	0.9826	0.7596	0.6208	0.4964
BART-PEARL	0.1642	0.0241	0.8803	0.9594	0.9365	0.9047
UniCRS-ReDial	0.0925	0.0097	0.9545	0.7916	0.6288	0.4635
UniCRS-PEARL	0.2153	0.0218	0.7551	0.8997	0.9352	0.9027
PECRS-ReDial	0.1979	0.0573	0.7995	0.7649	0.6801	0.6039
PECRS-PEARL	0.2619	0.0497	0.7637	0.9058	0.9182	0.9074
GPT-3.5	0.2256	0.0330	0.8990	0.9620	0.9354	0.9046

Table 11: Response generation performances on E-ReDial (Guo et al., 2023) Test-Rec subset.

Prompt
You are a Seeker who interacts with a Recommender to get a movie recommendation that suits your preferences well. You will be given a dialogue context, and you must follow the instructions below to interact with the Recommender:
- The recommender may ask for your preference or recommend a movie to you.
- In the beginning, express your general preference on movies based on your past reviews about movies you have watched before.
- If you are recommended a movie which is not {gt movie title}, you should reject it with a reason based on your thought about the recommended movie. Also, express some common features of the movies you have watched before and you should be recommended (DO NOT explicitly mention the movie title!).
- If you are recommended {gt movie title}, you should accept it as if you haven't watched it before, and end the conversation by generating [EOD] token.
- Continue the conversation for at least six turns.
Here are your reviews about movies you have watched before: {user persona}
Some features of the movie you should be recommended: {gt abstract}
{rec movie abstract}
Here is the dialogue context: {dialogue context}

Table 14: The prompt for Seeker simulator.

Prompt
You are a Recommender who recommends a Seeker a movie that he/she will enjoy, among the three candidates and based on the dialogue context given. Follow the instructions below to complete the task:
- In the beginning of the conversation, engage with the Seeker to discover his/her movie preferences (regardless of the given three candidates). Follow this format: Think: (think about which question to ask the seeker) Recommender: (utterance that asks for the seeker's movie preference)
- After some interactions, choose and suggest a movie from the three candidates and make the recommendation in the specified format: Think: (think about the seeker's movie preference based on the context) Movie: (Movie title (Year)) Recommender: (response to the seeker's utterance)
- Do not recommend a movie that has been mentioned before in the dialogue context.
Here are the three movie candidates: {k movies info}
Here is the dialogue context: {dialogue context}
Think:

Table 15: The prompt for Recommender simulator.