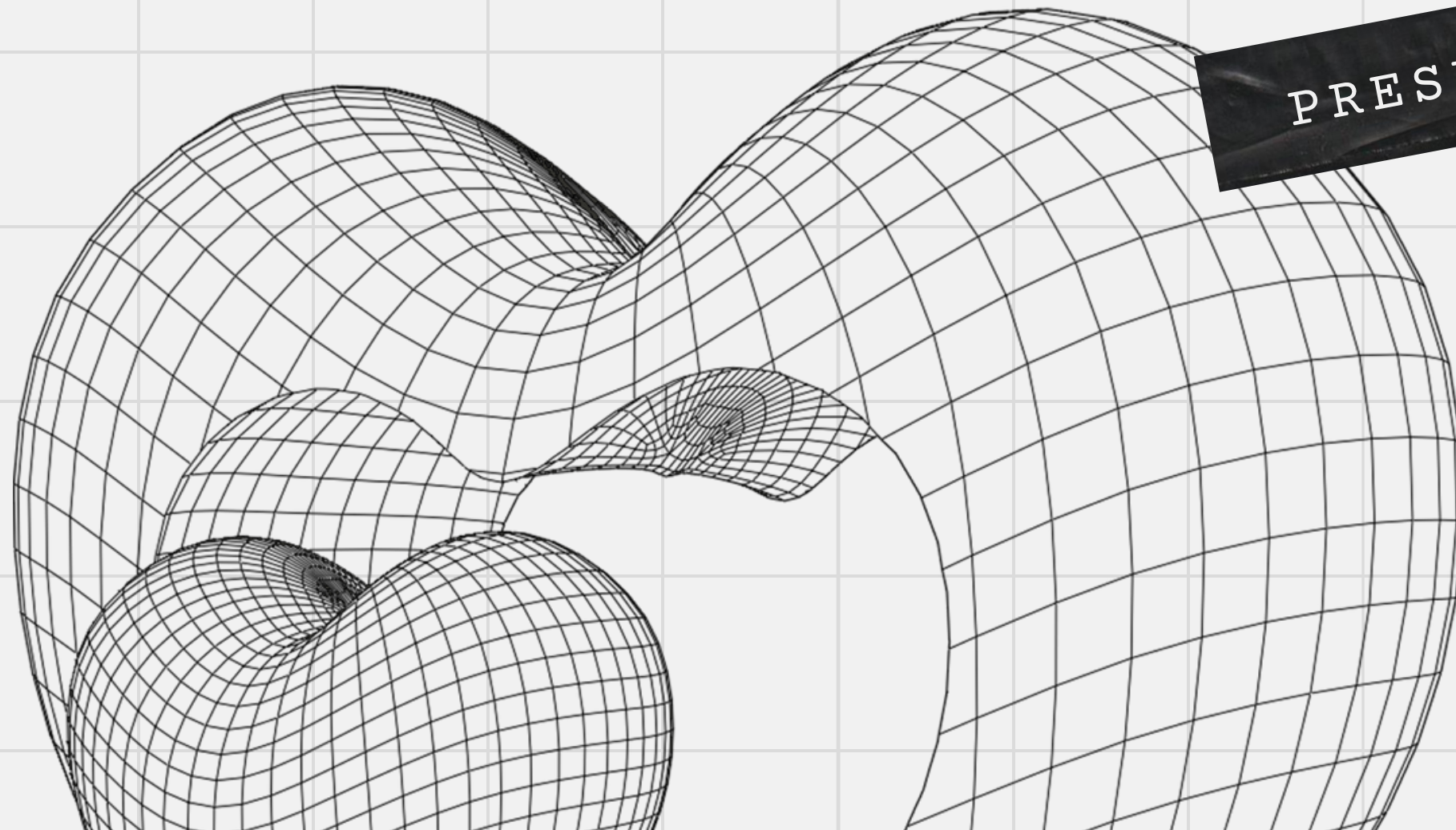


LEARNING PERSONALIZED STORY EVALUATION



PRESENTATION



THÀNH VIÊN NHÓM

2 1 1 2 7 5 7 7 - T R I N H H O À N G A N

2 1 1 2 7 0 5 6 - L Â M T H I Ê Æ U H U Y

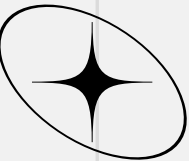
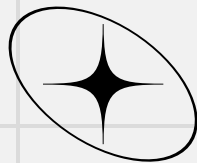


TABLE OF CONTENT

1	INTRODUCTION
2	RELATED WORK
3	PERSE: PERSONALIZED STORY EVALUATION MODEL
4	EXAMPLE OF PERSE

5	PERSONALIZED STORY EVALUATION DATASET
6	EXPERIENCES
7	CONCLUSION AND DISCUSSION





INTRODUCTION

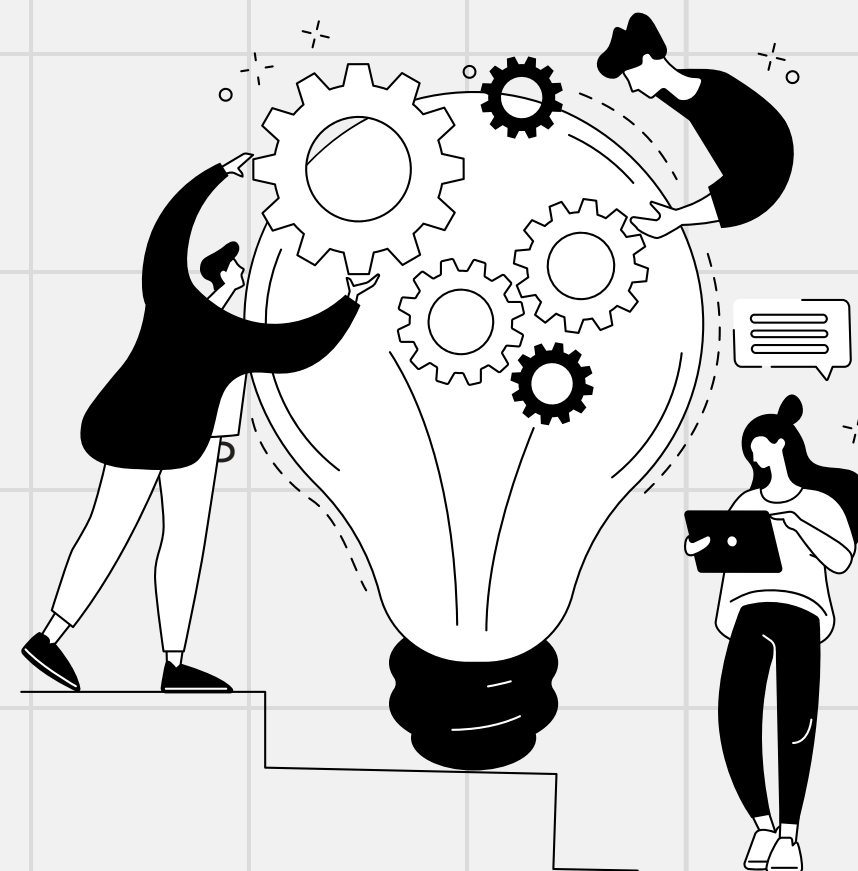
HIỆU SUẤT CỦA LLMS

**THÁCH THỨC TRONG
ĐÁNH GIÁ VĂN BẢN MỞ**

ĐÁNH GIÁ CÁ NHÂN HÓA

**MÔ HÌNH HÓA DỮ LIỆU
VÀ ƯU TIÊN:**

HIỆU SUẤT CỦA PERSE



**QUAN SÁT VỀ LLMS VÀ
CÁ NHÂN HÓA:**



lonesharks



RELATED WORK

1. Đánh Giá Câu Chuyện Tự Động

Các Chỉ Số Tham Chiếu và Không Tham Chiếu:

Sử Dụng LLMs Trong Đánh Giá: :

2.Đánh Giá Câu Chuyện Bởi Con Người:

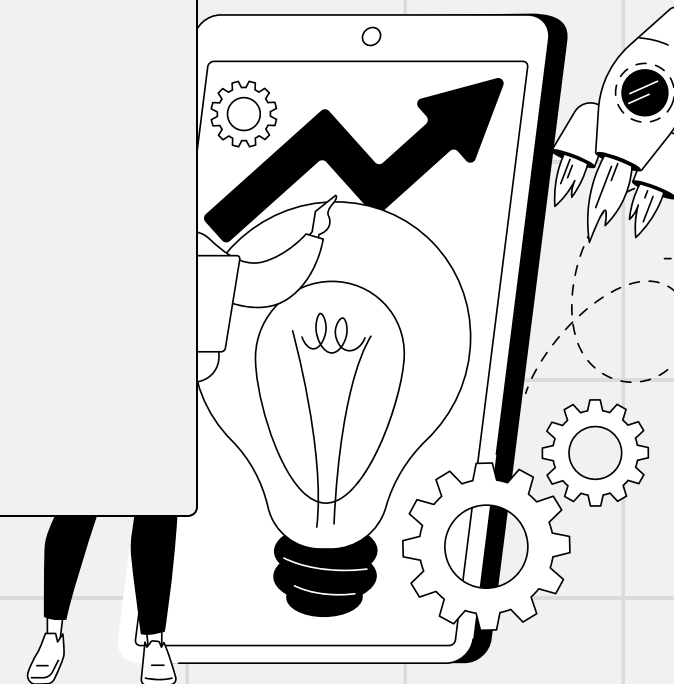
Các Khía Cạnh Đánh Giá:

Thách Thức

3. Cá Nhân Hóa Trong Xử Lý Ngôn Ngữ Tự Nhiên:

Tầm quan trọng của cá nhân hóa

Các Nghiên cứu gần đây:



MÔ HÌNH PERSE:

Mô hình PERSE (Personalized Story Evaluation Model) là một mô hình đánh giá câu chuyện được thiết kế để xác định và phản ánh sở thích cá nhân của người đánh giá dựa trên các đánh giá trước đó của họ. Dưới đây là một số thông tin cụ thể về mô hình PERSE:

1.Xác Định Sở Thích Cá Nhân

2.Tạo Đánh Giá Cá Nhân Hóa

3.Học từ Phản Hồi:

4.Đánh Giá So Sánh Tinh Tế:

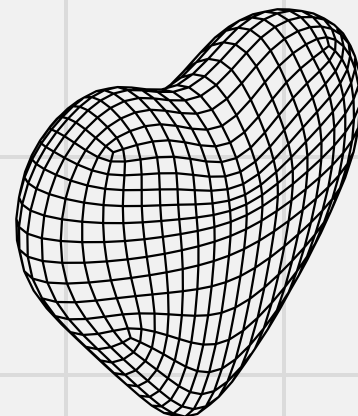
CÁCH THỨC HOẠT ĐỘNG:

Bước	Mô tả	Ví dụ
1	Xác Định Sở Thích Cá Nhân	Người đánh giá đã cung cấp các đánh giá trước đó cho các câu chuyện khác nhau. Sở thích cá nhân được suy luận từ các mô tả và điểm số trong những đánh giá này.
2	Tạo Cốt Truyện Mới	Cung cấp một cốt truyện mới để đánh giá. Cốt truyện này không nằm trong số các câu chuyện đã được đánh giá trước đó bởi người đánh giá.
3	Áp Dụng Sở Thích vào Đánh Giá	Sử dụng PERSE để tạo ra một đánh giá mới cho cốt truyện dựa trên sở thích cá nhân đã được xác định. Đánh giá bao gồm một điểm số và phản hồi chi tiết.
4	So Sánh và Phân Tích	So sánh đánh giá mới từ PERSE với đánh giá thực từ người đánh giá (nếu có) để xác định mức độ chính xác của mô hình trong việc hiểu sở thích cá nhân.
5	Cải Thiện Mô Hình	Sử dụng phản hồi để cải thiện khả năng suy luận sở thích của PERSE, nhằm tăng cường chất lượng đánh giá cá nhân hóa trong tương lai.

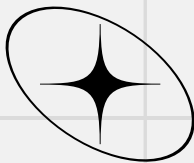
ỨNG DỤNG CỦA PERSE:

Trong thực tế, PERSE có thể hỗ trợ trong việc cá nhân hóa các hệ thống gợi ý câu chuyện, cung cấp phản hồi tự động cho các nhà văn, hoặc được sử dụng trong các ứng dụng AI khác nhau để hiểu và tương tác tốt hơn với sở thích cá nhân của người dùng.

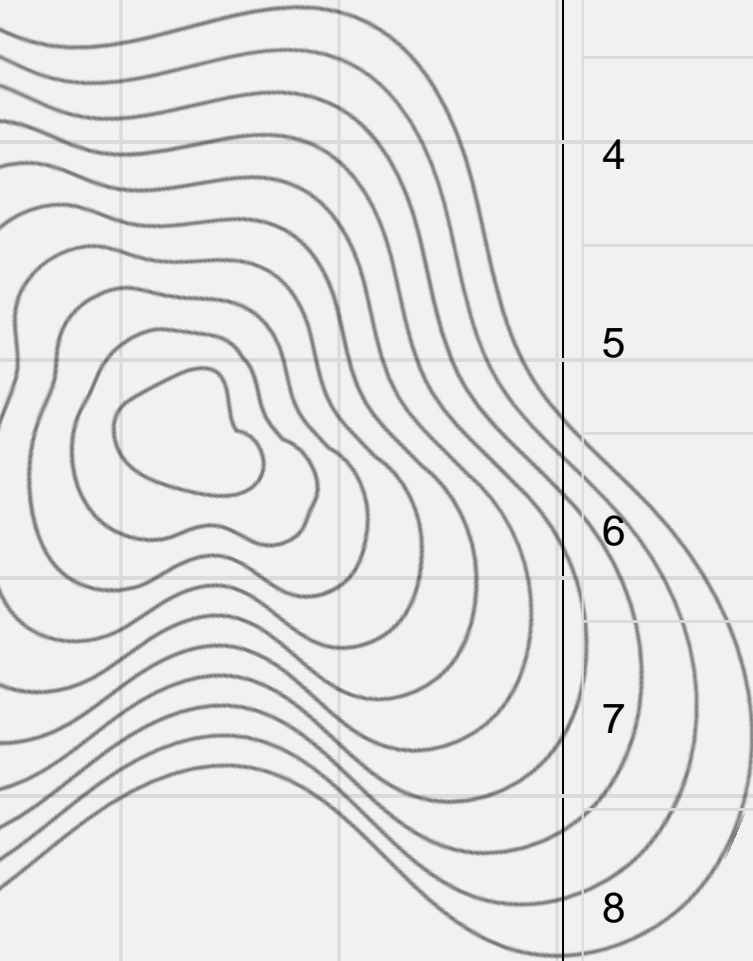
PERSE: PERSONALIZED STORY EVALUATION MODEL



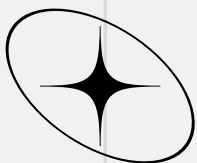
EXAMPLE OF PERSE



Bước	Mô tả	Ví dụ Vòng Lặp 1	Ví dụ Vòng Lặp 2
1	Xác Định Sở Thích Cá Nhân	Người đánh giá A thích câu chuyện có nhân vật sâu sắc và tình tiết phức tạp.	Người đánh giá A được hiểu là thích thêm yếu tố hài hước trong câu chuyện.
2	Tạo Cốt Truyện Mới (1)	Cốt truyện 1: Một bí mật gia đình được tiết lộ tại hội nghị tái hợp.	Cốt truyện 3: Cuộc phiêu lưu rừng nhiệt đới với nhiều tình huống hài hước.
3	Đánh Giá từ PERSE (1)	PERSE dự đoán: "Cốt truyện đầy kịch tính và nhân vật phong phú. Điểm 8/10."	PERSE dự đoán: "Một hành trình hài hước và nhẹ nhàng. Điểm 7/10."
4	Phản Hồi từ Người Đánh Giá (1)	Người đánh giá A: "Cốt truyện hay, nhưng thiếu hài hước. Điểm 6/10."	Người đánh giá A: "Cười không ngớt, nhưng cốt truyện hơi rối. Điểm 7/10."
5	Cải Thiện Mô Hình PERSE (1)	Cập nhật PERSE với thông tin: Người đánh giá A đánh giá cao kịch tính nhưng cũng mong đợi yếu tố hài hước.	Cập nhật PERSE: Tăng cường mô hình hóa yếu tố hài hước theo sở thích mới của người đánh giá A.
6	Tạo Cốt Truyện Mới (2)	Cốt truyện 2: Một cuộc phiêu lưu không gian với cuộc chiến chống lại quái vật.	Cốt truyện 4: Một cuộc đua xe đầy thách thức và những tình huống bất ngờ.
7	Đánh Giá từ PERSE (2)	PERSE dự đoán: "Phần không gian thú vị, nhưng quái vật không ấn tượng. Điểm 5/10."	PERSE dự đoán: "Tốc độ và hài hước kết hợp tốt. Điểm 8/10."
8	Phản Hồi từ Người Đánh Giá (2)	Người đánh giá A: "Cốt truyện không gian tuyệt vời, quái vật độc đáo. Điểm 9/10."	Người đánh giá A: "Hài hước nhưng cần một cốt truyện mạch lạc hơn. Điểm 6/10."
9	Cải Thiện Mô Hình PERSE (2)	Cập nhật PERSE: Điều chỉnh mô hình để tăng cường đánh giá sự độc đáo và đặc sắc của cốt truyện.	Cập nhật PERSE: Cải thiện khả năng đánh giá sự mạch lạc của cốt truyện để phản ánh sở thích cụ thể hơn của người đánh giá A.



PERSONALIZED STORY EVALUATION DATASET



CONTAMINATION IN EVALUATING EXISTING STORIES

Table 1: GPT-4 in comparing two movies.
The plot with the correct predicted title
is viewed as a known plot by GPT-4.

Accu. \bar{y} Cons. \bar{y} Bias First \bar{y} Percent

Both	Raw	0.714	63.0%		16.5%	91.0%
	Anonymized	0.712	60.7%	Known	17.8%	73.0%
	Summarized	0.753	73.4%		12.9%	42.5%
One Known	Raw	0.778	78.9%	-11.1%	-6.5%	9.0%
	Anonymized	0.804	71.7%		1.5%	23.0%
	Summarized	0.632	82.4%			34.0%
Neither Known	Raw	/	/	/		0.0%
	Anonymized	0.500	62.5%	25.0%	Summarized	4.0%
		0.660	85.1%	4.3%		23.5%

SỬ DỤNG BỘ DỮ LIỆU IMDB

KIỂM TRA SỰ Ô NHIỄM

TIẾN HÀNH THÍ NGHIỆM

Kết Quả và Quan Sát

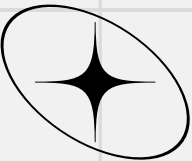


Table 3: Pearson, Spearman and Kendall correlations of each method with human ratings on Per-MPST.

	PEARSON	SPEARMAN	KENDALL
SIMPLE BASELINE	0.301	0.302	0.315
LLAMA-2-7B	0.146	0.117	0.315
LLAMA-2-13B	0.172	0.182	0.315
LLAMA-2-70B	0.214	0.232	0.315
GPT-4	0.315	0.312	0.315
PERSEIND-7B	0.307	0.329	0.263
PERSEIND-13B	0.345	0.368	0.293

EXPERIENCES

MAIN RESULTS



EXPERIENCES

ANALYSIS

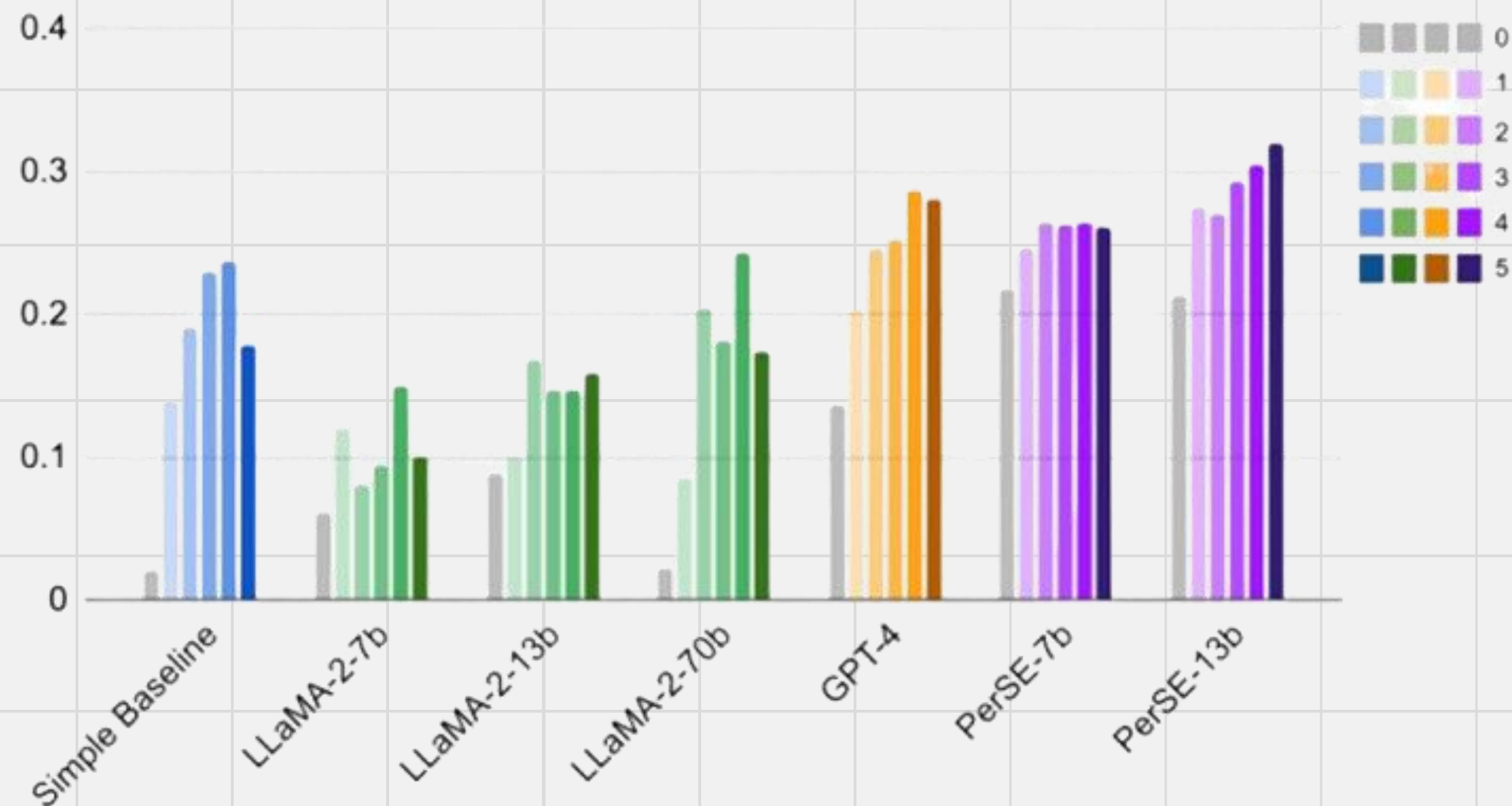


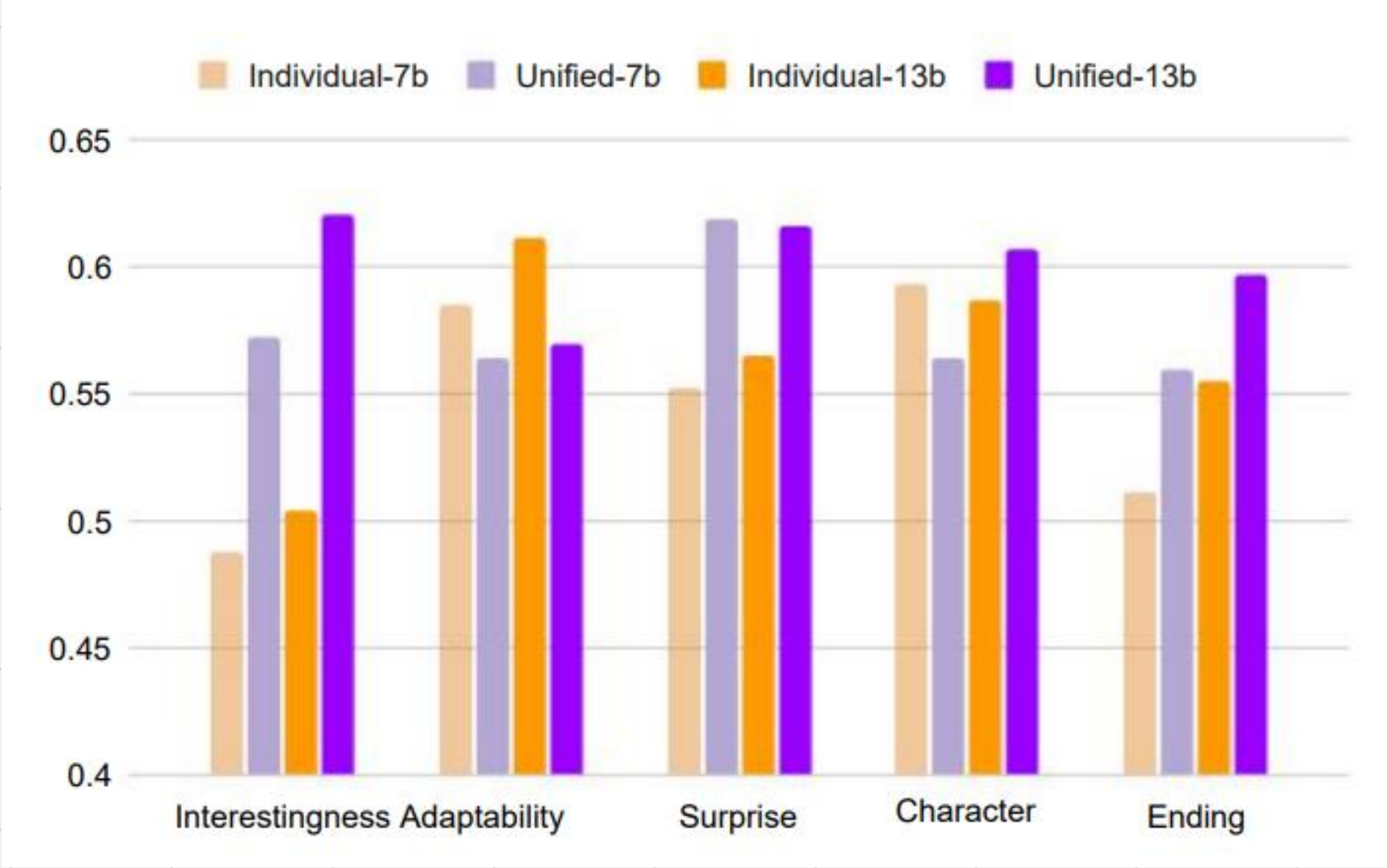
Figure 3: Kendall correlation on Per-MPST with different numbers of reviews (k) in reviewer history. Having more reviews benefits PERSE-13b, but the increased complexity may harm the performance of LLaMA.

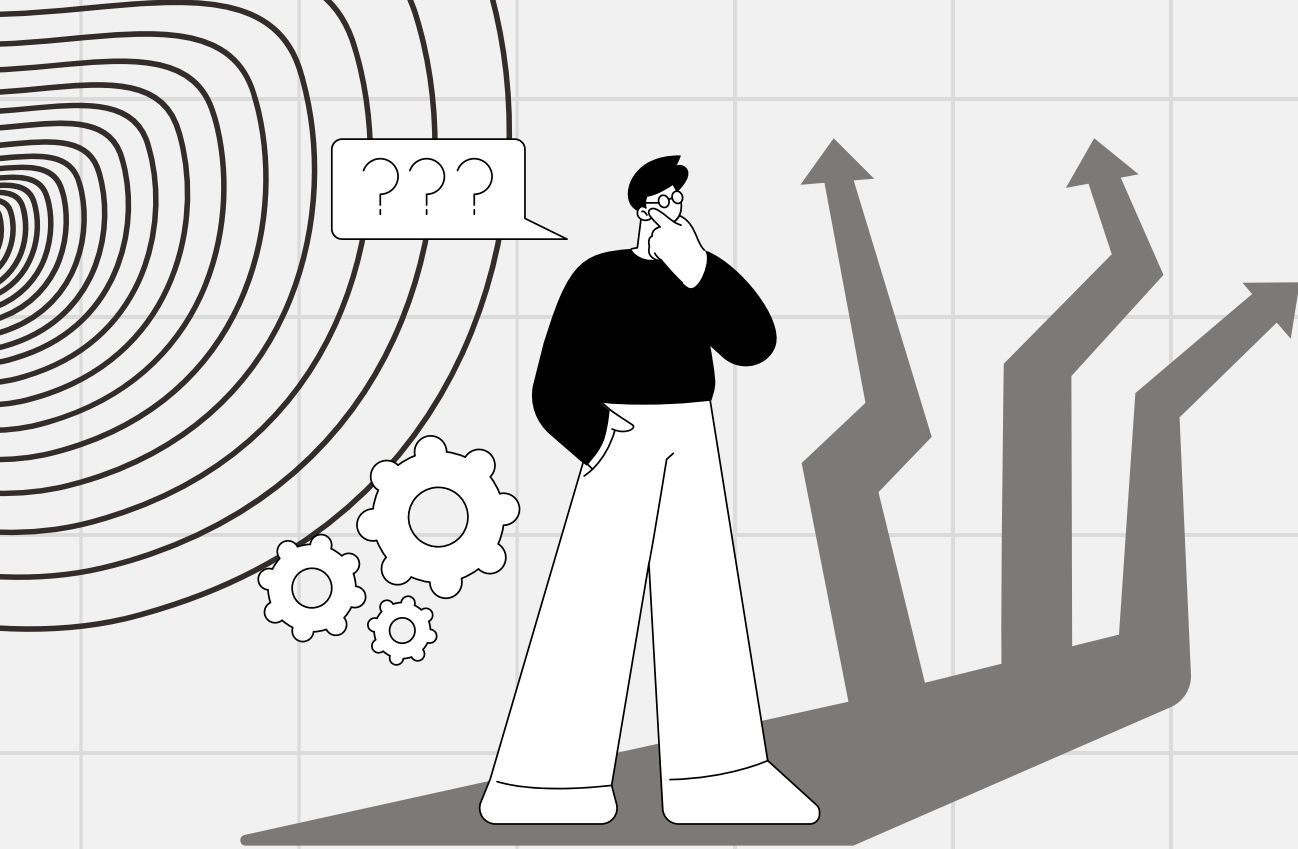
EXPERIENCES

ANALYSIS



Figure 5: Accuracy of the unified models and individual models on Per-DOC.



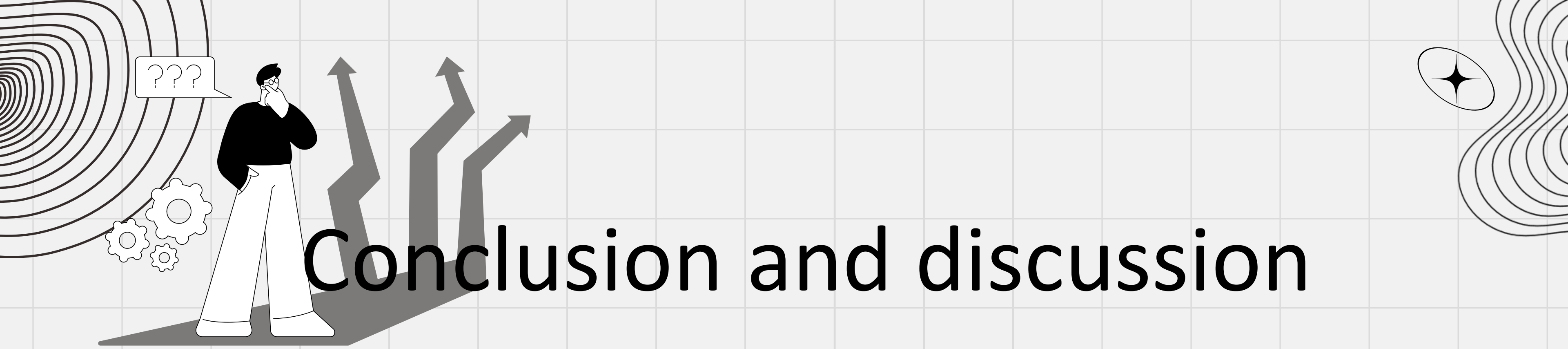


EXPERIENCES

ANALYSIS

Review Preference
<p>[The Start of Plot 0] A newscast is shown about a year after a prehistoric piranha attack on Lake Victoria. The lake is now uninhabitable, and the town has been abandoned due to the drying up of the main revenue source, tourism...</p> <p>[The End of Plot 0] [Review] Quote The Hoff: Welcome to rock-bottom.. features everything that gives horror , as well as sequels in general , a bad reputation [Score] 3</p> <p>[The Start of Plot 1] Evelyn Summers is a shy and unattractive but intelligent girl who likes witchcraft and has a crush on Bill Natolly, her handsome English teacher...</p> <p>[The End of Plot 1] [Review] honey, I think we should see other people... Dont you just hate it when this happens? Now that you finally found yourself a gorgeous looking high school sweetheart, she turns out to be a murderous witch, capable of making you eat glass beer bottles with the touch of her hand! [Score] 5</p>
Plot Query
<p>Professor Jack Hanson and his colleagues go to the Gibraltar ice shelf to take ice core samples and discover evidence of global warming through a presentation to the UN that is poorly received. However, when multiple buoys in the North Atlantic show a sharp drop in ocean temperature, Professor Christopher Rapson believes Jack's theories and works with him to build a forecast model....</p> <p>At the US refugee camp in Austinia, President Elizabeth Blake died when her motorcade was caught in the superstorm, and Vice President Elizabeth James Becker became the new president. The next morning, Sam's group realized the storm is over, and they began their journey journey to the evacuation center.</p>
Reference
<p>[Review] I mean, you can't judge a film negatively because its far fetched. The whole thing may be scientifically unjustified, but it has to be remains a little appealing to the audience, no? The terrifyingness of this film lies in the same old, disastrous cliché that keeps on getting repeated over and over again. [Score] 2</p>
GPT4 Response
<p>[Review] This film offers a chilling, literally and metaphorically, journey into the devastating effects of global warming, a pressing issue that's often brushed aside... Nevertheless, the film effectively uses the fear of the unknown future of our planet to craft an engaging narrative. [Score] 7</p>
LLaMA-2-70b Response
<p>[Review] I really am amazed at how steady and consistently terrible the movies are. Oh well, at least the disaster itself was kind of cool to watch, although even that had its flaws and clearly visible budget restrictions. Too bad they didn't spend a little more money on it, and actually made it a decent movie. [Score] 3</p>
PerSE Response
<p>[Review] I'm just not sure if it all adds up to a compelling story. I mean, the characters are pretty one-dimensional, and the plot feels like it's all over the place. [Score] 4</p>

Figure 6: An example for evaluating the individual story from the given reviewer's preference. The reference is the ground-truth reviews given by this reviewer. Some content has been abridged due to space considerations. More cases are shown in the Appendix.



Conclusion and discussion

•Tổng Kết Nghiên Cứu:

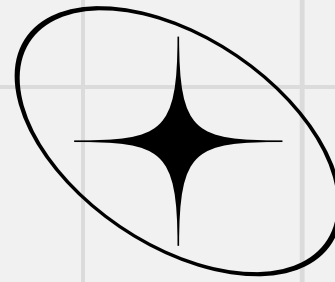
- Nghiên cứu tập trung vào đánh giá cá nhân hóa của tạo văn bản mở cuối cùng.
- Phương pháp nghiên cứu bao gồm việc tạo ra các bộ dữ liệu cá nhân hóa không thiên vị từ các nguồn dữ liệu hiện có, giúp giảm thiểu vấn đề về việc mô hình đánh giá LLMs tiếp xúc với các bộ dữ liệu đã tồn tại.
- Hai bộ dữ liệu đánh giá câu chuyện, Per-MPST và Per-DOC, được tái sử dụng để đánh giá cá nhân hóa toàn diện trong các tình huống khác nhau.
- Mô hình PERSE được đề xuất cho đánh giá câu chuyện cá nhân hóa và đã cho thấy hiệu suất cao, vượt trội so với GPT-4 trong cả đánh giá chung và so sánh tinh tế giữa các cặp.

•Hạn Chế của Nghiên Cứu:

- PERSE sử dụng đánh giá từ người đánh giá để nắm bắt sở thích của họ, với giả định rằng sở thích này không thay đổi theo thời gian, điều này có thể không phản ánh chính xác các tình huống thực tế.
- Hạn chế về độ dài ngữ cảnh của các mô hình ngôn ngữ lớn hiện tại có thể ảnh hưởng đến việc hiểu biết toàn diện về sở thích của người đánh giá.

•Hướng Phát Triển Tiếp Theo:

- Nghiên cứu khuyến khích phát triển tạo câu chuyện cá nhân hóa, với mục tiêu sử dụng PERSE để cải thiện sự phù hợp giữa mô hình tạo câu chuyện và yêu cầu đặc biệt của người đánh giá.
- Mục tiêu tiếp theo là sử dụng PERSE để cải thiện sự phù hợp giữa mô hình tạo câu chuyện và yêu cầu riêng biệt của người đánh giá



THANK YOU

Xin cảm ơn mọi người đã lắng nghe. Mong rằng những thông tin chia sẻ hôm nay sẽ mở ra những cơ hội mới và mang lại cảm hứng cho chúng ta cùng hợp tác và phát triển. Hãy cùng nhau tạo nên những dấu ấn đặc biệt trong tương lai!