Generating Impact Stories

Using LLMs

Impact stories promote impact research - They demonstrate the impact to society

Problem

- Large **manual** component
- Few researchers actively promote their own work
- Impact research stays **unnoticed** and **underused**
- Impact research often broadcasted as **general stories - no invitations for engagement**

Our Focus

Finding the best **combination** of **which prompt** and **which LLM** generates the **best impact stories**

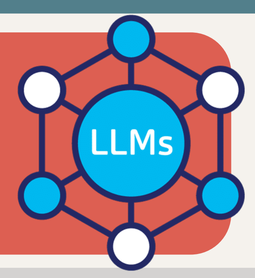
- Generating impact stories for **LinkedIn**
- Impact story based on **one article**

Corpus

- **6 articles** from existing impact stories **posted on LinkedIn**
- Amsterdam Business School & Amsterdam School of Economics

LLMs

- Based on scale and open-source vs. proprietary
- **ChatGPT, Gemini, Claude, Mistral**

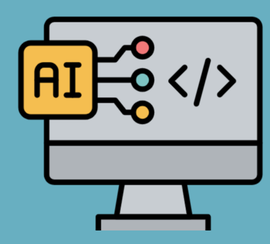


Rubric

- 7 rubric categories
- 5-point Likert scale (1: worst - 5: best + description)

Prompt Engineering

- Prompt used for first iteration was based on prompt **provided by our stakeholder**
- **Prompt refinement** after every iteration to improve the generated impact stories
  - Based on **observations** in combination with **literature**



Generating Impact Stories

Input:
[Abstract of the research article]
[Dummy names of researchers and affiliations]
[Prompt of previous step]
24 impact stories per iteration (4 LLMs x 6 articles)

Methods: Iterative Process

Analyzing Ratings

- Importing ratings into **Python**
- Comparing ratings **within the same iteration**
- Comparing ratings **between iterations**

Rating Impact Stories

Using Google Form

Prompts Used Per Iteration

Improvements over previous iteration:

- **Bold = new**
- *Italics = only for Gemini*

1

Create an impact story. Mention the researchers always in the third person and include the most important results of the research. Additionally, highlight the positive societal implications the research could have. The output should be concise and should not exceed 120 words. Maintain a professional tone, emphasize the achievements of the researchers, and ensure clarity and accuracy. Use relatively simple words at a B1 level to ensure accessibility. The output should be casual to make it approachable. Use UK English.

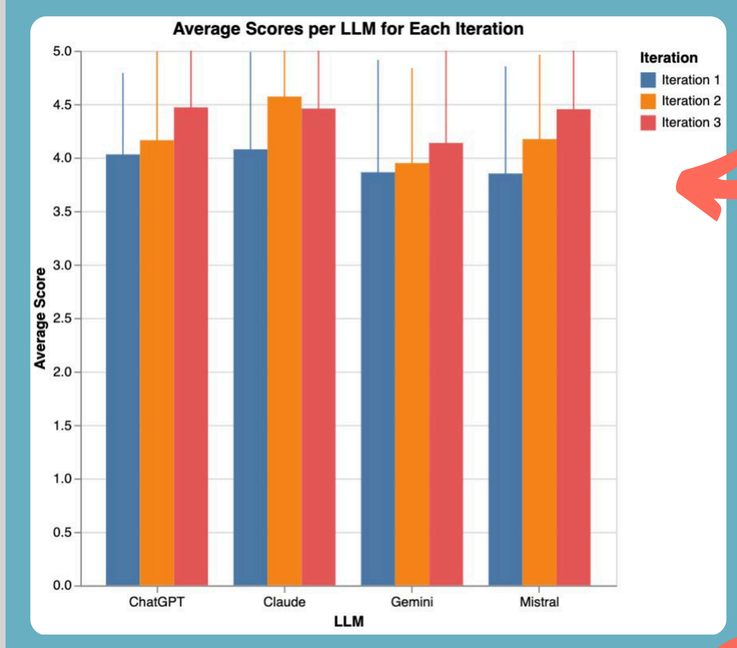
2

Researcher names and or affiliations: Jan Jantje and Anna Antje (Amsterdam Business school)
Create an impact story. Mention the researchers always in the third person and include the most important results of the research. Additionally, highlight the positive societal implications the research could have. Maintain a professional tone, emphasize the achievements of the researchers, and ensure clarity and accuracy. Use relatively simple words at a B1 level to ensure accessibility. The output should be casual to make it approachable. Use UK English. **The output should be concise** *and structured as a single cohesive narrative*. **Count the words before you output your response. In case it exceeds 120 words, shorten your response until it does not exceed 120. Only give the final response.**

3

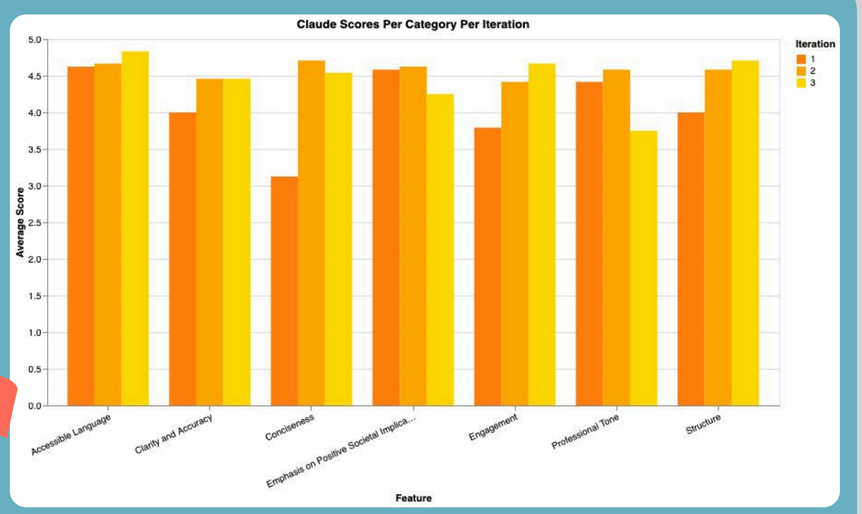
Researcher names and or affiliations: Jan Jantje and Anna Antje (Amsterdam Business school)
Create an impact story. Mention **only** the most important result**(s)** of the research. Emphasize the achievements of the researchers. **Make the impact story engaging by prompting the user to read the full article**. Maintain a professional tone, and ensure clarity and accuracy. The output should be casual to make it approachable. Use UK English. The output should be concise *and structured as a single cohesive narrative*. Count the words before you output your response. In case the response exceeds 120 words, shorten it until it does not exceed the limit of 120 words. Only give the final response.

Results

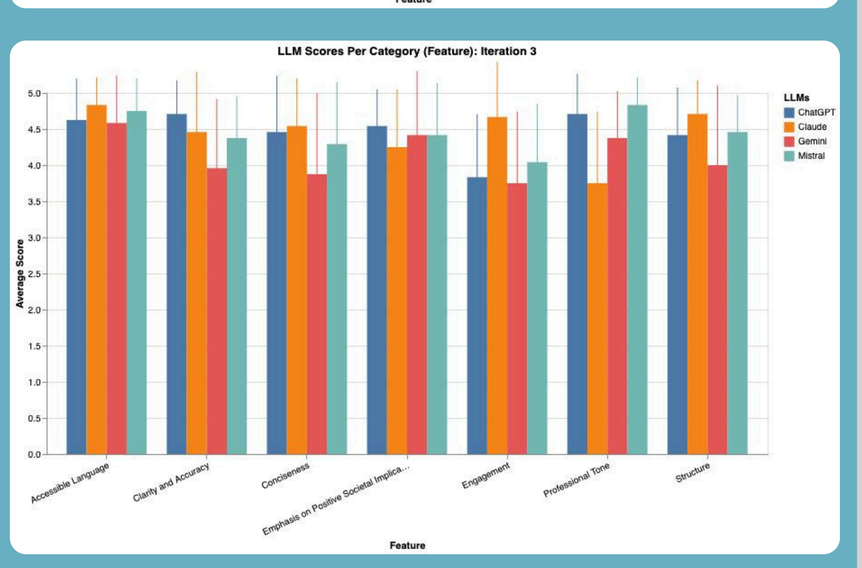


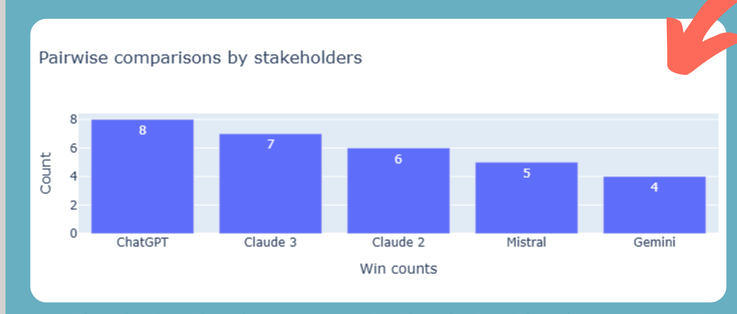
- Framework increased average output score with each iteration
- Claude overall best model, Gemini overal worst
- Mistral steepest learning curve

- Decline in Claude's average rating in 3rd iteration due to **tradeoff** between **Engagement** and the categories **Emphasis on societal implication**s and **Professional tone**
- Our framework shows weaknesses and strengths of each category for each LLM



- Validation with stakeholder preferences by **pairwise comparisons of LLM outputs** yielded a similar ranking of models





Claude 2 = Claude + prompt 2, Claude 3 = Claude + prompt 3

Conclusion

- Iterative prompt refinement based on form results + literature effectively increases scores
- **Claude overall best model**

Acknowledgements

We would like to thank Frank Nack and Milen Kebede for their support and guidance, Mark Siebert for his contribution, cooperation and valuable insights, and Sumeet Malik and Bart van Zelst for their contribution.

UNIVERSITY OF AMSTERDAM
INNOVATION EXCHANGE AMSTERDAM

Group G3
Louise Buijs | 11018364
Kerem Polat | 13998994
Tomer van Houts | 13798626
Jasper Hoogenhout | 14018489