JDSE 2025

Analyzing Menstrual Cycle Awareness in LLM-Generated Fitness Advice

Alessa Mayer[1]

[1]Insitut Polytechnique de Paris (Master's student)

Introduction

- AI applications become more prominent in the personal growth and self-improvement industry [2].
- Allocation bias in NLP describes the issue of lower model performance on data associated with the minority gender [4].
- Fitness advice, as well as related research, has primarily focused on the male population [3], leaving out the important influence of menstruation on sport activities [1].

Research Question

Is the influence of menstruation on performance considered in LLM-generated fitness advice?

Methodology

We designed a set of prompts to evaluate whether large language models (LLMs) consider menstruation when generating fitness advice. Three categories of prompts were used:

- Neutral (no mention of gender): e.g., "Create a training to gain strength within 6 weeks."
- Female (explicit mention of gender): e.g., "Create a training for women to gain strength within 6 weeks."
- Menstruating person (explicit mention of menstruation): e.g., "Create a training for menstruating people to gain strength within 6 weeks."

We formulated 50 different prompts and refined them for each category, resulting in a total of 150 different prompts.

For each generated output, we applied two evaluation methods: i) Keyword detection, which consists in searching the output for the terms "menstrual", "luteal", "follicular", "ovulation", "hormonal", "pms", "menstruation", "bleeding", "menstruating"; ii) LLM-as-judge by directly asking the model whether menstruation or hormonal changes were considered in the generated output.

Contact

For questions or collaboration:

`alessa.mayer@ip-paris.fr`

(Preliminary) Results

Table 1: Proportion of LLM-generated fitness advice outputs that mention menstruation or hormonal changes, by prompt type. Values are percentages of 50 prompts.

Llama3.2	**Keyword**	**LLM-as-Judge**
Neutral (no gender)	0%	14%
Female (gendered)	16%	54%
Menstruating person (explicit)	100%	86%
Mistral	**Keyword**	**LLM-as-Judge**
Neutral (no gender)	0%	0%
Female (gendered)	2%	12%
Menstruating person (explicit)	98%	74%

(Preliminary) Conclusions

- Menstruation is consistently underrepresented in LLM-generated fitness advice.
- Findings suggest male-default bias in LLM output, overlooking menstrual cycle considerations.
- Results align with the frequent omission of menstruation in fitness research.

Future Work

- Include human evaluation (by experts)
- Test more models
- Test more prompts
- Study other signs of allocation bias in output

References

[1] Findlay, R.J., Macrae, E.H.R., Whyte, I.Y., Easton, C., Forrest (née Whyte), L.J.: How the menstrual cycle and menstruation affect sporting performance: experiences and perceptions of elite female rugby players. British Journal of Sports Medicine **54**(18), 1108–1113 (2020)

[2] Lab, U., Mattila, V., Rahman, M., Pang, Y.Z.: Ai-driven innovation in the personal growth industry: Transforming self-improvement through personalization (11 2024)

[3] Navalta, J.W., Davis, D.W., Stone, W.J.: Implications for cisgender female underrepresentation, small sample sizes, and misgendering in sport and exercise science research. PLOS ONE **18**(11), 1–15 (11 2023). https://doi.org/10.1371/journal.pone.0291526

[4] Sun, T., Gaut, A., Tang, S., Huang, Y., ElSherief, M., Zhao, J., Mirza, D., Belding, E., Chang, K.W., Wang, W.Y.: Mitigating gender bias in natural language processing: Literature review. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. pp. 1630–1640. Association for Computational Linguistics, Florence, Italy (Jul 2019)

Junior conference on Data Science and Engineering