Generative social choice: the good, the bad, the challenges

Generative social choice framework. This paper [1] introduces the generative social choice framework to inform democratic decisions that satisfy the BRJ guarantee, demonstrating that LLMs can be new building blocks for democracy [4]. Through an interactive process of statement generation, discriminative rating prediction, and satisfied agent removal, a final slate of statements can be selected [1].

Considerations. The first consideration is biases in LLMs stemming from their training data, which disproportionately represent WEIRD (*Western, Educated, Industrialized, Rich, and Democratic*) societies [2], and fine-tuning processes such as RLHF [1,3]. Other considerations include a lack of transparency and interpretability due to the black-box nature and opaque training data. Challenges also arise from safety concerns such as malicious input and scalability issues related to context windows and computational costs. Lastly, LLMs have a tendency to hallucinate, particularly when applied to specialized or niche queries.

Evaluation challenge. While discriminative queries can be assessed against ground truths, the evaluation of the generative queries is more complex. Despite outlining an ensemble approach to evaluate different generation sources, one author notes that the approach's interpretability is suboptimal [4]. A potential approach could involve implementing an interactive process to gather more nuanced responses, including direct feedback on the generated statements.

Adapting the framework for marketing insights. This framework can be adapted to aggregate customer opinions and generate consensus-driven statements, providing actionable insights. An interactive consumer feedback loop could be integrated for iterative refinement and evaluation. In terms of market segmentation and representation, LLMs can help identify and address underrepresented demographics, thereby expanding the reach of marketing efforts. Condensed insights can balance completeness and efficiency by summarizing customer data while maintaining critical details. Lastly, incorporating human validation can ensure the reliability of generated insights.

Last comment. In the pilot, the recruitment source through the online platform Prolific, the incentive of free responses (*we will reward thoughtful answers with a bonus of $2*), and the limited emphasis on underrepresented groups could potentially skew responses toward widely accepted statements, compromising the diversity of perspectives.

[1] Fish, S., Gölz, P., Parkes, D. C., Procaccia, A. D., Rusak, G., Shapira, I., & Wüthrich, M. (2023). Generative social choice. arXiv preprint arXiv:2309.01291.
[2] Atari, M., Xue, M. J., Park, P. S., Blasi, D., & Henrich, J. (2023). Which humans?.
[3] Santurkar, S., Durmus, E., Ladhak, F., Lee, C., Liang, P., & Hashimoto, T. (2023, July). Whose opinions do language models reflect?. In International Conference on Machine Learning (pp. 29971-30004). PMLR.
[4] Generative Social Choice, Youtube