

Analysis of Linguistic Stereotypes in GenerativeAI

Adriano De Cesare Anna Lisa Maddaloni Giovanni Giordano Silvia Mantione
hh Matteo Di Gregorio

Abstract

This document is a supplement to the general instructions for *ACL authors. It contains instructions for using the L^AT_EX style files for ACL conferences. The document itself conforms to its own specifications, and is therefore an example of what your manuscript should look like. These instructions should be used both for papers submitted for review and for final versions of accepted papers.

1 Introduction

Describe here the objectives and the context of application of your experiment. This text can be based on the application description or on the SEMeval task requirements

1.1 Research Questions

- R.Q.1:** What types of linguistic stereotypes do LLMs reproduce?
- R.Q.2:** Does prompt structure (zero-shot, role prompting, chain-of-thought) amplify or reduce bias?
- R.Q.3:** Can multi-agent critique frameworks reduce stereotypical outputs?

2 Background

The ability of Large Language Models (LLMs) to pick up biases encoded in training data can lead to the risk of systematic judgments on social stereotypes, reflecting different categories of biases, that can be based on gender or on race. (?)

In the context of linguistic stereotypes, more recent students have highlighted covert forms of prejudice emerging through language usage itself, rather than through explicit mentioning. More into details, the authors have focused on African American English (AAE) and Standard American English(SAE). The study demonstrates that dialect prejudice has visible consequences: LLMs are more likely to assign AAE speakers to lower-prestige jobs, predict criminal behavior, or recommend harsher legal outcomes. (?)

Beyond English, studies on German dialects show that LLMs match dialect speakers with negative traits and occupational assignment with a lower score. (?)

Analogous patterns emerge in work on Egyptian Arabic, where LLMs exhibit higher bias compared to Modern

Standard Arabic, reflecting biases against low-resource dialects. (?)

However, existing research conducted on Italian linguistic variations, focuses primarily on the model's metalinguistic awareness and understanding of non-standard linguistic structures. (?)

Building on the matched-guise methodology(?), this paper investigates covert stereotypes arising from the usage of Italian dialects.

3 System overview

Describe all the methodological choices that you performed: e.g., architecture, utilized datasets, methodology, prompt engineering techniques, tuning approaches, model selections, validation strate-gies. Use separate subparagraphs for each piece of information. Use diagrams to describe architectural choices.

4 Experimental results

Describe all the results of your experimentation. Provide separate subparagraphs for each of the defined research questions.

5 Conclusions

Provide a critical evaluation of your work. What are the main outcomes and limitations? How can the work be extended?

5.1 Citations

Table ?? shows the syntax supported by the style files. We encourage you to use the natbib styles. You can use the command \citet (cite in text) to get “author (year)” citations, like this citation to a paper by ?. You can use the command \citep (cite in parentheses) to get “(author, year)” citations (?). You can use the command \citealp (alternative cite without parentheses) to get “author, year” citations, which is useful for using citations within parentheses (e.g. ?).

A possessive citation can be made with the command \citepos. This is not a standard natbib command, so it is generally not compatible with other style files.

References

Fatma Elsafoury and David Hartmann. 2025. Out of sight out of mind: Measuring bias in language models

against overlooked marginalized groups in regional contexts. *arXiv:2504.12767*.

Angelina McMillan-Major Emily M. Bender, Timnit Gebru and Shmargaret Shmitchell. 2021. *On the dangers of stochastic parrots: Can language models be too big?* *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*.

Dan Gusfield. 1997. *Algorithms on Strings, Trees and Sequences*. Cambridge University Press, Cambridge, UK.

Kalluri P.R. Jurafsky D. et al. Hofmann, V. 2024. *Ai generates covertly racist decisions about people based on their dialect*. *Nature* 633.

Angelapia Massaro and Giuseppe Samo. 2023. Prompting metalinguistic awareness in large language models: Chatgpt and bias effects on the grammar of italian and italian varieties. *Verbum, vol. 14*.