Contribution ID: **46**   Tür: **Oral Presentation**

Racio-Linguistic Biases in Large Language Models - Exploring Methods and Themes

18 Aralık 2024 Çarşamba 14:00 (30 dakika)

In this talk, I investigates possible forms of racial biases in the output of Large Language Models. The study is based on linguistic anthropological theory, where 'race'is understood as linguistic and discursive construction (Alim, Reyes & Kroskrity 2020). The language generated by AI is influenced by these constructions. The data collection to test this is based on a research-lead MA seminar where students inspected output of ChatGPT for racial and minority discrimination. Six different prompts were developed that were tested on whether the output would create racially biased language or content. Prompts were entered by ten different individuals and in different languages (English, German, Arabic, French, Spanish, Ukrainian, Italian, Portuguese, Tamil, and different minority/ non-standard varieties such as Piedmontese, Tanglish, Andean Spanish, Siculish, Kiezdeutsch or Berlinisch).

The results of the study indicate that there are several aspects that influence the output of ChatGPT. These include topics of previous prompts of the user, choice of language or language variety, sensitivity of the topic and how the prompt has been phrased. The analysis overall gives indirect access to possible limitations of the data-set (especially if the language used is not English), to the relevance of use history and to debiasing techniques that in some cases are likely to have influenced the output. In this sense, the study serves as an explorative approach to methods that can be used to evaluate the functioning, limitations and advantages of the tools.

At the same time, the study is indicative of how an LLM can be used to reflect on social practices of speaking as appearing in data sets, which registers or topics designers of technologies and developers of algorithms deem relevant or appropriate language practices (Markl & McNulty 2022) and how language culture may be reconfigured because of the existence of AI language models.

Presentation language / Sunum Dili

EN (English)

Disciplines / Disiplinler

Linguistics / Dilbilim

E-mail / E-posta

bschneider@europa-uni.de

ORCID ID

0000-0003-0377-5808

Country / Ülke

Germany

Institution / Affiliation / Kurum

European University Viadrina, Faculty of Social and Cultural Sciences

Başlıca yazarlar::   Britta Schneider (European University Viadrina)

Sunu yapanlar: Britta Schneider (European University Viadrina)

Session Classification: Session 1.3 (Day 1)

Track Classification: Congist'24: Case Studies and Applications