A Cross-Lingual Analysis of Bias in Large Language Models Using Romanian History

Adrian Marius Dumitran[1], Răzvan Cosmin Cristia[2], AND Matei-Iulian Cocu[3]

[1]*University of Bucharest, Softbinator*
marius.dumitran@unibuc.ro
[2]*University of Bucharest*
cristiarazvan@gmail.com
[3]*University of Bucharest*
cocu.matei24@yahoo.com

Abstract

In this case study, we select a set of controversial Romanian historical questions and ask multiple Large Language Models to answer them across languages and contexts, in order to assess their biases. Besides being a study mainly performed for educational purposes, the motivation also lies in the recognition that history is often presented through altered perspectives, primarily influenced by the culture and ideals of a state, even through large language models. Since they are often trained on certain data sets that may present certain ambiguities, the lack of neutrality is subsequently instilled in users. The research process was carried out in three stages, to confirm the idea that the type of response expected can influence, to a certain extent, the response itself; after providing an affirmative answer to some given question, an LLM could shift its way of thinking after being asked the same question again, but being told to respond with a numerical value of a scale. Our research brings to light the predisposition of models to such inconsistencies, within a specific contextualization of the language for the question asked.

Keywords: Romanian History, LLM Linguistic Bias, LLM Training and Assessment, Natural Language Processing, Digital Humanities

1 Introduction

Reasoning - the process of drawing conclusions to facilitate problem-solving and decision-making (Leighton, 2003); a significant number of studies indicate the fact that reasoning has become a prominent feature of LLMs (), but along with this quality comes a certain bias towards some ideologies of certain domains. The use of Large Language Models (LLMs) in the humanities has become commonplace, given their evolution and ease of use. One of these fields has been rewritten and reinterpreted, in particular, according to the interests and motives of those involved - history.

2 Methodology

The methodology for this study was structured into several key stages, (each designed to ensure a comprehensive analysis of the biases present in Large Language Models (LLMs) when addressing controversial Romanian historical questions.)

1. In the initial stage, a set of (15 intrebari, menite sa aduca la lumina biasul)
2. In the second stage,
   (a)
3. The third stage

2.1 LLM Selection

For our experiments, we chose

2.2 Questioning Process

2.2.1 Prompt

The following prompt template was used

2.2.2 Question Selection

3 Answer Comparison

4 Conclusions

References

nume complet an-ref *paper* editura, etc.

Defining and describing reasoning: Reasoning as mediator. 2003 *The nature of reasoning*, pages 1-11.

. , .

. , .

American Psychological Association. 1983. *Publications Manual*. American Psychological Association, Washington, DC.

Ashok K. Chandra, Dexter C. Kozen, and Larry J. Stockmeyer. 1981. Alternation. *Journal of the Association for Computing Machinery*, 28(1):114–133.

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

Mohammad Sadegh Rasooli and Joel R. Tetreault. 2015. Yara parser: A fast and accurate dependency parser. *Computing Research Repository*, arXiv:1503.06733. Version 2.

Benjamin Borschinger and Mark Johnson. 2011. A particle filter algorithm for Bayesian wordsegmentation. In *Proceedings of the Australasian Language Technology Association Workshop 2011*, pages 10–18, Canberra, Australia.