

Examining the Efficacy of LLMs in Comprehending Homophones

Clair Ross
Nuoqi Liu

March 6, 2024

1 Introduction

Homophones are words that sounds similar but have different meanings, like 'new' and 'knew'. This phenomenon is rare in English, but extremely frequent in certain languages, like Chinese. Homophones can be manually created easily for almost every Chinese word. As an example, the word 警察 (policeman) sounds exactly the same as 井茶 (tea in the well), 景查 (scenery and investigation) and 颈獐 (neck and badger).

Manually created homophones are commonly used for pranks, and for bypassing censorship in social media. Human can understand such homophone expressions easily. However, it is suprising that state-of-the-art LLMs like ChatGPT has no ability at all in understanding such homophone expressions. In this study, we will build a test set, and use it to measure some state-of-the-art LLMs' ability in comprehending homophone expressions.

2 Objectives

A test set of 100 questions about Chinese homophones will be created manually. One example question goes as:



你

In the sentence '今天有两个井茶来到了学校', the word '井茶' is a homophone word, can you tell its true meaning?

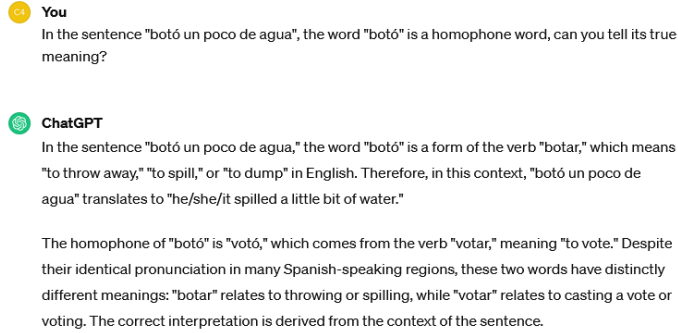


ChatGPT

The word "井茶" does not have a commonly recognized meaning in Mandarin Chinese. It seems like a typographical error or a made-up term. It's possible that you meant to type a different word or phrase. If you provide more context or clarify the intended term, I'd be happy to help you understand its meaning.



Also, a similar test set of 100 questions about Spanish homophones will also be created. One example question goes as:



The test set will be ran on 5 LLMs or translators. The performance for each model will be measured and analyzed. Possible strategies to improve the performance, like chain-of-thought prompting, will be tried and analyzed.

3 Evaluation Metrics

LLM's efficacy of comprehending homophones will be measures by accuracy on test set. For the example question '今天有两个井茶来到了学校', if the correct answer '警察' or 'policeman' is shown in LLM's reply, the answer is marked as correct.

4 Task Division

Firstly, all group members will read the related papers, to get a broader view of this topic.

Nuoqi Liu will create the Chinese homophone test set, run the test set on 5 LLMs.

Ross Clair will create the Spanish homophone test set, run the test set on 5 LLMs.

Nuoqi Liu will undertake the task of data analysis and academic writing.

Ross Clair will undertake the task of literature review.

5 Existing Literature

We found no study focusing on this topic. Most study focus on homophones in voice chat other than text chat. Some papers partly related to this topic includes:

[Stability Analysis of ChatGPT-based Sentiment Analysis in AI Quality Assurance](#)

[Speech Translation with Large Language Models: An Industrial Practice](#)

We plan to read these papers, and hopefully get some enlightenment.