Our model *shocks* the industry with a novel and innovative approach WOW!

Wumbolo, Tegla, Nikola Kraljević

University of Zagreb, Faculty of Electrical Engineering and Computing
Unska 3, 10000 Zagreb, Croatia
{Wumbolo,Tegla,Nikola.Kraljevic2}@fer.hr

Abstract

This paper presents our approach to tackling the Sentence Puzzle task in the SemEval-2024 Task 9 competition: "BRAIN-TEASER: A Novel Task Defying Common Sense." Diverging from the conventional multi-class classification framework, we introduce a novel pairwise comparison methodology. Instead of forcing the model to choose the correct answer from four options, we reformulate the task into multiple binary decisions, evaluating pairs of answers at a time. This approach, inspired by the efficiency of One-vs-One and One-vs-Rest strategies, aims to enhance the model's reasoning capabilities and accuracy . Our results showcase... ako budu dobri rezultati tu cemo flexat, xD

1. Introduction

Question answering tasks have long been a central challenge within NLP. The competition, named "SemEval 2024 BRAINTEASER: A Novel Task Defying Common Sense" described in the paper (**?**) is here to push modern models to their limits, putting them to a quite adversarial setting. The paper makes a distinction between vertical and lateral thinking. Vertical thinking, also known as linear, logical, convergent, is mostly a sequential and analytical process, requiring direct memory recall or a few logical steps to come to a sensible conclusion. Lateral thinking, also known as "thinking outside the box", is more of a divergent and creative process, where the question might not make sense when reading it at first and to come up with an answer you need to explore mutliple angles. While LLMs show good vertical thinking capabiliteis, they are notorious for hallucinating answers. An even more adversarial task for LLMs is answering questions which require lateral thinking. State of the art models such as ChatGPT show an accuracy of 60%, humans show a 90% accuracy, while a random guessing gets close to 25% accuracy as there are four answer candidates for each question [1].

In this paper, we tackle the second subtask from the competition "Sentence puzzles" with an alternative approach which we believe will make it easier for the LLM to reason about all the possible answers.

2. Related Work

Related Work

3. Dataset

Dataset

4. Evaluation Results

Eval Results

5. Conclusion

Conclusion

[1] These accuracies are fore one of the subtasks of the competition - "Sentence puzzles", which we will describe in more detail in the following chapters

6. Acknowledgements

Acknowledgements: (**?**) semeval, (**?**) ails-lab

References

Yifan Jiang, Filip Ilievski, Kaixin Ma, and Zhivar Sourati. 2023. Brainteaser: Lateral thinking puzzles for large language models.

Ioannis Panagiotopoulos, Giorgos Filandrianos, Maria Lymperaiou, and Giorgos Stamou. 2023. Ails-ntua at semeval-2024 task 9: Cracking brain teasers: Transformer models for lateral thinking puzzles.