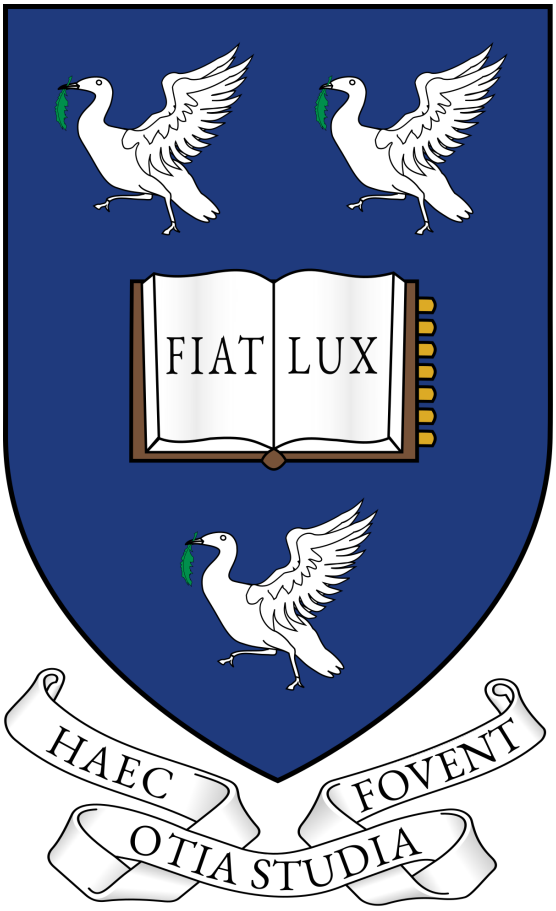

THE UNIVERSITY OF LIVERPOOL

COMP702-MSC FINAL PROJECT

Visual Spatial Reasoning for Large Language Models

Author Jinlong Liu
ID 201678181
Supervisor Frank Wolter

June 21, 2023

I. Project Description

In recent years, Large Language Models (LLMs) have achieved great success in the field of Natural Language Processing (NLP). The LLMs are trained on large-scale text corpus, and can be used to solve many NLP tasks, such as machine translation, question answering, and text generation. The LLMs are usually trained on text corpus, and can be used to solve many NLP tasks, such as machine translation, question answering, and text generation.
However, the LLMs are not good at solving the Spatial Reasoning tasks. In this project, I will design an experiment for most popular LLMs, such as GPT-4, to test their ability of Visual Spatial Reasoning. The experiment will be based on the Visual Question Answering (VQA)–a task that requires the machine to answer questions about the given image. For example, given an image with different objects, the machine needs to answer the question "What the relevant position of the object A to the object B?". With these tests, we can evaluate what level of Spatial Reasoning ability the LLMs have.

II. Aims and Objectives

III. Key Literature and Background Reading

[1]

IV. Development and Implementation Summary

V. User Interface Mockup

VI. Data Sources

VII. Testing

VIII. Evaluation

IX. Ethical Considerations

X. Project Plan

XI. Risks and Contingency Plans

References

[1] A. Borji, "A categorical archive of chatgpt failures," *arXiv preprint arXiv:2302.03494*, 2023.