



THE UNIVERSITY OF LIVERPOOL

COMP702-MSC FINAL PROJECT

Visual Spatial Reasoning for Large Language Models

<i>Author</i>	Jinlong Liu
<i>ID</i>	201678181
<i>Supervisor</i>	Frank Wolter

June 21, 2023

I. PROJECT DESCRIPTION

In recent years, Large Language Models (LLMs) have revolutionized the field of Natural Language Processing (NLP) with their remarkable achievements. These models, trained on vast amounts of text data, have proved highly effective in solving numerous NLP tasks, including machine translation, question answering, and text generation. Their ability to comprehend and generate human-like text has been nothing short of impressive.

However, despite their proficiency in various linguistic tasks, LLMs have encountered challenges when it comes to Spatial Reasoning. Unlike tasks centered purely on language, Spatial Reasoning requires understanding and manipulating visual and spatial information. To shed light on this limitation, I aim to conduct an experiment utilizing popular LLMs like GPT-4. The focus will be on testing their capability in Visual Spatial Reasoning.

The experiment will revolve around a widely used task known as Visual Question Answering (VQA), which demands machines to answer questions based on provided images. For instance, given an image depicting different objects, the machine will be tasked with responding to inquiries such as "What is the spatial relationship between object A and object B?". Through these carefully crafted tests, we can gauge the extent of Spatial Reasoning proficiency exhibited by LLMs.

By assessing their performance in VQA and their ability to comprehend and reason about spatial relationships in visual content, we can obtain valuable insights into the Spatial Reasoning capabilities of LLMs. These findings will not only deepen our understanding of the strengths and limitations of these models but also pave the way for further advancements in the realm of NLP, bringing us closer to more comprehensive and versatile language models

II. AIMS AND OBJECTIVES

III. KEY LITERATURE AND BACKGROUND READING

[?]

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.