LLM Agent Whiteboard Spatial Benchmark

Benjamin Wagley

July 28, 2025

Background

Benjamin Wagley is a PhD Student working with Dr. Mehmet Belviranli and Dr. Bo Wu, studying AI and Systems. His latest projects have focused on understanding how LLMs process spatial information and how that can be leveraged to build visualizations. Building on his latest work, he is looking for an enthusiastic, research-minded group of students to help build an open-source benchmark tool for analysing agentic spatial understanding in LLMs and MLLMs.

Project Outline

The rise in Multimodal Large Language Models, with their ability to process image input, has opened the avenue for new approaches in building LLM backed agentic tools. However, it is important to understand and evaluate their capabilities in solving spatial tasks for these tools to reach their full potentials.

In this project, the team will build and enhance a LLM-agent whiteboard spatial reasoning benchmark to be open-sourced at the end of the project. This benchmark aims to examine and rate the performance of current LLM models and techniques as they perform complex visual tasks. Building upon existing digital whiteboards such as TLDraw and Excalidraw, the team will use Javascript to build a browser-based application that allows LLM agents to see and interact with the whiteboard environment. To build out the benchmark, they will implement a series of tests and evaluations within this environment. The final product will allow for individual testing for close evaluation, as well as automated testing for quickly determining model performance.

Logistics

- We are looking for students who have some experience with Javascript. We will be using Typescript and React frameworks. Experience with LLMs and LLM-Agents is welcome, but not required. Interest in research is welcome, we will foster a research-minded environment.
- Location: Most meetings will be held over Zoom.
- Outcomes: Students will produce an open-source LLM Benchmarking Tool. They will learn to work with the latest LLMs from OpenAI, Anthropic, and Google, as well as open models such as Llama, Mistral, and Qwen.
- Size: 3-4 Students.
- Project Lead / Point of Contact: Benjamin Wagley, bwagley@mines.edu