Project Proposal - Multi Agent Debate

Project Introduction

The aim of the project is to look at self correction in LLMs and how multi-agent debate can drive the LLM to produce the correct solution.

Multi-agent debate is a framework through which multiple LLMs come together to discuss and provide the correct next logical step.

Project Objectives

The objective of the project is to take a task for which 1 LLM has its limitations when it comes to producing the correct output and try to improve its output using the multi-agent debate framework.

Project Scope

The task is specific to only chess. We won’t be exploring any other game/domain.

We strictly use only LLMs and no external tools to provide us feedback. (In our case, we won't be using chess engines to predict the next move.)

Project Description

In general, LLMs are not good at playing chess. Chess involves playing a move based on all the previous moves of both parties. LLMs perform well upto certain moves for well known openings in chess. But once it encounters moves which are not documented as such, it starts producing illegal moves. Our aim is to correct this by the use of Multi-agent debate. Two or more LLMs debate amongst themselves to provide a valid next move.

Design of Experiment

For implementing the framework we require the following -

- 2 debaters - We will use amongst various LLMs and assign them the debater role.
- 1 Judge - There will be an LLM which is assigned the judge role.

We will start with a single LLM to find its limits. Then by using this framework, with different LLM base models we will try to push the limits of a single LLM.

It requires no prerequisite dataset. The starting positions of the chess board will be constant over all experiments.

Expected Outcome

We expect the Multi-agent debate framework to give us at least 2-3 moves ahead of the current best (achieved by only 1 LLM)

Note that we won’t be looking for the best move but any move that is valid. Our hypothesis is that this might lead to achieving more number of next steps.

Metrics of Evaluation

Any move that takes the chess board from a valid state to another valid state is considered as a valid move.

We will also consider the suggestions of a chess engine to see how good the move is. Though getting the best move might not be the primary objective.

Project Cost

The following is the estimation of the time that would be required -

- Planning and refinement - 1 week
- Experimental setup - 2 weeks
- Conducting experiments - 2 weeks
- Documentation of results - 1 week

Total cost 6 weeks