

# Project Proposal

## Introduction

Chess engines such as Stockfish are becoming increasingly more powerful, and are now able to easily outperform even the best humans in much less time. Many players use the evaluations and continuations suggested by these engines to better understand positions and improve their ability. They often achieve this by looking through the various continuations of moves that the engine suggests in order to identify the tactical or strategical advantage that one move has over another.

However, less experienced players will have more difficulty identifying these differences, since they will have encountered the tactical and strategic ideas less frequently in the games they have played. Furthermore, many beginners will not have been explicitly taught about some tactical ideas, and may not be able to understand why a position is favourable for one player even after going deep down a path of suggested moves. Understanding why a certain move was better than another is one of the most important aspects needed to learn from a mistake made in one game, and to make a better move in a similar position in future games.

My project aims to combine approaches to interpretable machine learning models with techniques from more traditional chess engines in order to provide accurate evaluations and move suggestions, as well as giving reasons for them. This would be achieved by training an interpretable model on a dataset of chess positions, with the features being tactical and strategical aspects of the position and the output being the evaluation assigned to the position by a leading chess engine. A chess engine will then be built using the trained model as the evaluation function, and combining the interpretable output of the positions in the search tree to provide reasoning for the evaluation in terms of tactics and strategy.

## Work to be Undertaken

### 1. Preparing the data:

I will need to familiarize myself with different representations of chess positions, and decide on the best one to use based on how quickly I will be able to extract the tactical and strategical features, as well as which is supported by the engine I will use for my training data evaluations. I will also need to find and clean a dataset of chess positions in this representation to use as training and testing data for my interpretable model, by calculating the evaluation of each position using an existing chess engine.

### 2. Devising the model:

I will need to investigate and decide on the model which is most appropriate for the project. This will involve researching and experimenting with multiple interpretable models to find one that provides both accurate evaluations as well as having clear interpretations. During this stage I will also need to decide on which features I use. The speed at which the model can generate evaluations for a new position will also be an important factor, since it will determine the search depth that I will be able to use in the engine.

### 3. Training the model:

Once I have chosen which model to use, I will need to train the model using the dataset that I have prepared. This stage may also involve the tuning of some hyperparameters depending on the model which I choose to use.

#### 4. **Implementing the chess engine:**

Using the trained model as an evaluation function, I will implement a chess engine using the minimax algorithm, using techniques such as alpha-beta pruning to increase the depth which can be computed to in the same amount of time. During this stage I will also need to determine how to combine the interpretations of the children into the interpretation for a position.

#### 5. **Evaluating the chess engine:**

There are two aspects of the chess engine which will need to be evaluated, the accuracy of the evaluations it assigns to positions, and the explanations for these evaluations. The accuracy of the evaluations can be evaluated by comparing them to the evaluations assigned by the original chess engine that was used to train the model. The interpretations will be more difficult to evaluate, as they are often more subjective. One approach could be to find a dataset of puzzles which have a defined explanation for which is the best move.

## Extensions

#### 1. **GUI to highlight reasoning:**

In order to make it easier for users to visualize the interpretations, I could design a graphical user interface which highlights and explains the tactical and strategic aspects of the position that contribute most to the evaluation.

#### 2. **LLM to provide natural language explanations:**

In order to make the explanations for the evaluation easier to understand for beginners, I could use a Large Language Model to combine the most important factors with explanations of how they create an advantage for one player to provide a more natural explanation.

## Success Criteria

The project should be considered a success if the chess engine is able to produce sensible evaluations of chess positions, as well as the tactical and strategical aspects that contributed most to that evaluation. The project should be considered very successful if the evaluations align closely with those from the best chess engines, and if it produces explanations for the evaluations in natural language and/or through a visualisation.

## Starting Point

## Timetable

## Resources