9^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2023, Copenhagen, Denmark

Alternatives to Winning: Exploring the Implications of Human-like Chess AI

Keywords: Deep Learning, Chess, Human-Centered AI, Opponent Modeling, Reinforcement Learning

Extended Abstract

The advent of machine learning systems that surpass human ability in various domains raises the possibility that people could learn from and interact with superhuman AI. However, tasks like developing algorithmic teaching aids are currently made difficult by the fact that AI systems typically behave very differently from humans. To build AI systems that can collaborate with humans we need to have models that can understand concepts that are relevant to humans, not just to maximizing an easily measured objective function. In this work we define three more complex objectives and build reinforcement learning systems that can optimize for them. We then use these systems to explore the intersection of human and AI collaboration to both build systems that are human compatible and antagonist to human cognition.

Data. We use data from the largest open-source online chess platform, Lichess, which has hosted billions of games played by millions of players. Lichess uses a rating system for measuring player skill, with higher values corresponding to higher skill. We used games from 2017 to 2022 for training and validation, and show all our results on held out games from 2023.

Specific Task. We aim to build interactive chess systems (engines) that attempt to play chess with goals beyond winning that present both quantitatively and qualitatively interesting properties. Specifically, we present three axis for optimization: *Humanness*, *Agreeableness*, and *Mistake Frequency*.

Methods. We started with the models from Maia Chess [3], which are versions of *AlphaZero Chess* [4] that predict moves made by human players at a specific skill level. We then trained a series of models using similar methods that achieved higher levels of accuracy and a broader range of skill levels. To verify that they are human-like, we both use the move matching accuracy of the original paper [3] and use the knowledge acquisition probing method from [2] to show that the models contain representations that are homologous to specific human created heuristics.

We then used a variant of Monte-Carlos Tree Search (MCTS) [5, 1] that has the dual objectives of maximizing the probability of winning and maximizing/minimizing another objective value (*Humanness*, *Agreeableness*, or *Mistake Frequency*) that uses the model's predictions as an oracle to derive a probability distribution over their actions on a given board state.

Optimization Targets. Our optimization targets are *Humanness*, *Agreeableness*, and *Mistake Frequency*. For all objectives we use a reinforcement learning search framework to search the space of possible moves (*actions*) for a given chess position (*state*) and select the move that maximizes the objective. The three objective values are:

HUMANNESS. the probability that the model will make a move that a human player would make. We use the move matching accuracy and perplexity to measure it. Minimizing *Humanness* is thus equivalent to maximizing the probability that the model will make a move that a

human player would not make. But purely maximizing this value would result in a model that merely plays the worst possible move, so we also regularize to win probability, allowing an explicit trade-off between *Humanness* and performance.

AGREEABLENESS. the probability that an opponent will experience 'easy' positions. We measure the 'easiness' of a position using the value, this gives the expectation value of the change in win probability of each move, given the moves are sampled from our human-like oracle function ($\sum_{\text{legal moves}} P_{pred} \times winrate$). With this definition we can then conduct a depth limited search to find moves that will lead to opponent encountering positions where they are likely to make errors (*swindle model*) or least likely to make mistakes (*swaddle model*). Again, we need to regularize to win probability to prevent pathological behaviour, *i.e.*, the model immediately losing.

MISTAKE FREQUENCY. the number of mistakes the model makes during a game. We measure this by using a strong chess engine (*Stockfish*) as an oracle to determine if a move was a mistake and how severe it was using the winrate loss value. Using this framework we can have a model that is trying to minimize the number of mistakes it makes, while also trying to win, which is the normal objective of a chess engine. But we can also have a model that is trying to maximize the number of mistakes it makes, while also trying to win, which leads to much more interesting behaviour. This optimization target is also intended to be used with the *humanness* objective. This allows the creation of a *blunder-free human model*, allowing us to model how a human would perform if they never made large errors, i.e. ask what is optimal *human* performance at chess.

Results. We are in the process of running user studies, these are run as chess bots on the Lichess platform, with the model choice being randomized between participants, providing a single blind experimental design. We will both be asking users for feedback on the bots, and using metrics such as the rematch rate to determine if the bots are interesting to play against and have the desired qualitative properties, i.e., we hypothesis that the *swaddle model* model will be more pleasant to play against than the *swindle model*. This feedback is then being used to optimize the models further. Quantitative results will also be gathered, specifically measurements of how well the bots are able to achieve their optimization objectives. Early results on model calibration for humanness are shown in figure 1, notably our current search methods increase accuracy at the cost of calibration.

References

[1] Athul Paul Jacob et al. "Modeling strong and human-like gameplay with KL-regularized search". In: *International Conference on Machine Learning*. PMLR. 2022, pp. 9695–9728.

[2] Thomas McGrath et al. "Acquisition of chess knowledge in alphazero". In: *Proceedings of the National Academy of Sciences* 119.47 (2022), e2206625119.

[3] Reid McIlroy-Young et al. "Aligning Superhuman AI with Human Behavior: Chess as a Model System". In: *Proceedings of the 25th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2020.

[4] David Silver et al. "A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play". In: *Science* (2018).

[5] David Silver et al. "Mastering chess and shogi by self-play with a general reinforcement learning algorithm". In: *arXiv* (2017).

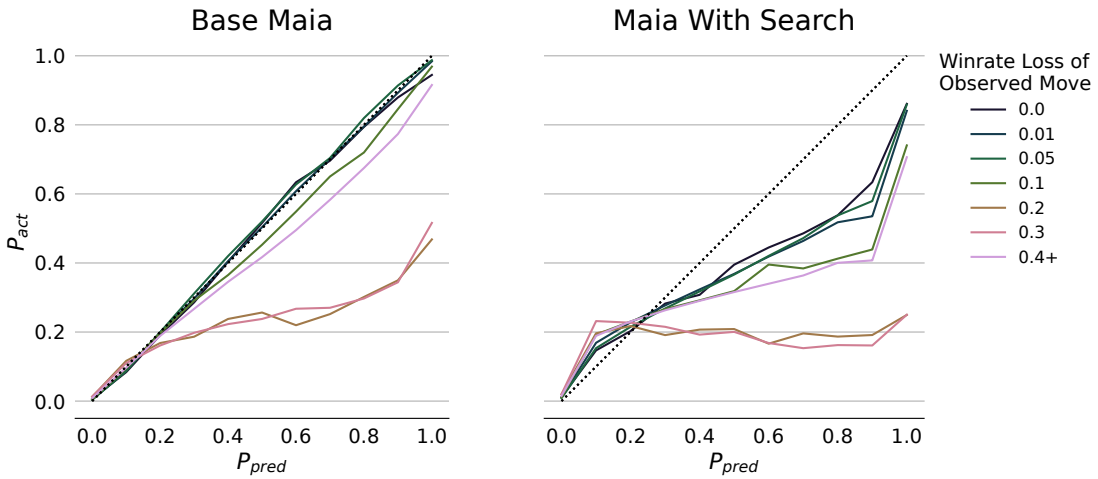

Figure 1: Calibration of the 1900 *Maia* model with no search (pure policy network) and with 50 rollouts of MCTS. Despite being less well calibrated the search model has a slightly higher accuracy (53% to 53.5%). Winrate loss is the difference between the expected winrate of the optimal move and the winrate of the move the chosen move by the player. Our definition of winrate loss also means that there the 0.4+ set is larger than the 0.3 set since it includes any move that does not lead to mate soon if another move does.