Maia Chess
A Human-Like Neural Network Chess Engine

Ashton Anderson
Reid McIlroy-Young
University of Toronto
Chess and AI: A Long History

Alan Turing created the first chess-playing algorithm in 1948, before computers could run it. Claude Shannon described minimax in chess in 1949.

Long-standing AI problem: “Chess is the drosophila of artificial intelligence” — John McCarthy, 1967

“However, computer chess has developed much as genetics might have if the geneticists had concentrated … on breeding racing Drosophila. We would have some science, but mainly we would have very fast fruit flies.” — John McCarthy
Chess and AI: A Long History

Deep Blue defeats Kasparov in 1997

AlphaZero defeats Stockfish in 2017
Chess AI and People

"Here’s what I would do" — 3400-rated calculation beast 🤖

“But what should I do?” — 1100-rated human 😞
Chess AI and People

Stockfish level 4

Ashton
Chess AI and People

Stockfish level 4

Ashton
Chess AI and People

Stockfish level 4

Ashton
Chess AI and People

How do we bridge the gap between artificial and human intelligence?

Chess as a model system:

1) AI reached superhuman performance at least 15 years ago
2) There is voluminous data on human activity in chess
3) Wide range of human skill levels
How can we algorithmically capture human style in chess?
Predict the next move a human, at a specific skill level, will make in a real game.
Data

lichess.org

game database

The largest openly available dataset of human games
1.7B games and counting
Players, moves, move times, etc.
Background: AlphaZero and Leela

Background: Attenuated Engines

Depth-limited Stockfish

Limit search depth to simulate fallible thinking

Early Leela models

Take Leela versions early in their self-play training to simulate not-fully-evolved understanding
Do Attenuated Engines Perform Well?

Attenuated engines match aggregate human performance (rating)

Do they match granular human decision-making (moves)?
Evaluation Set

From December 2019:

1. Create bins for each range of 100 rating points
2. Divide games into bins by rating of both players
3. Select 10,000 games from each bin between 1100 and 1900
4. Metric: move-matching accuracy: % of positions for which model’s move matches the human move played in the game.
What Does It Mean to Perform Well?

Move-matching Accuracy

Rating

1100 1300 1500 1700 1900
Attenuated Stockfish: limited to various depths

Move-matching accuracy increases with rating of players being predicted

Relatively low accuracy, no skill-level targeting

Interesting non-monotonicity: d1 and d15 both more accurate than d5
Leela

Attenuated Leela: versions at different points in training

Move-matching accuracy flat across ratings of players being predicted

Somewhat higher accuracy, no skill-level targeting
Key idea:
Learn from human play instead of self-play

Goal: predict the next move a human will play
# Maia–Leela Comparison

<table>
<thead>
<tr>
<th></th>
<th>Leela</th>
<th>Maia</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data</strong></td>
<td>Self-play games</td>
<td>Human games</td>
</tr>
<tr>
<td><strong>Goal</strong></td>
<td>Optimal play</td>
<td>Human-like play</td>
</tr>
<tr>
<td><strong>Method</strong></td>
<td>Reinforcement learning</td>
<td>Classification</td>
</tr>
<tr>
<td><strong>Tree search</strong></td>
<td>Monte-Carlo tree search</td>
<td>No tree search</td>
</tr>
<tr>
<td><strong>Policy</strong></td>
<td>Optimal moves</td>
<td>Human moves</td>
</tr>
<tr>
<td><strong>Value</strong></td>
<td>Probability of winning given optimal play</td>
<td>Probability of human winning</td>
</tr>
</tbody>
</table>
Train a Maia for each rating level between 1100 and 1900

Maia 1100  12M games between 1100s
Maia 1200  12M games between 1200s
Maia 1300  12M games between 1300s
Maia 1400  12M games between 1400s
Maia 1500  12M games between 1500s
Maia 1600  12M games between 1600s
Maia 1700  12M games between 1700s
Maia 1800  12M games between 1800s
Maia 1900  12M games between 1900s
Maia Move-Matching Performance

Maia: versions trained on different rating levels

High accuracy: best performance >50%

Move-matching accuracy maximised at target rating level

Maia’s lowest accuracy, Maia 1900 tested on 1100-rated players (46%), is still higher than any Stockfish or Leela model we tested

High accuracy and skill-level targeting
Maia Move-Matching Performance

Shaded region: rating levels not explicitly targeted (2000+)

Graph showing accuracy over rating for Maia, SF, and Leela at different depths.
Maia: Architecture Decisions

Base Maia vs. a version with no history and a version with Leela-style rollouts (calculation)

Including calculation hurts move-matching performance by ~5–8% (moves become too strong)

Dropping history hurts performance by 2–3%
Playing Maia 1100 is like playing a committee of 1100-rated players, etc.
7,000+ unique human opponents in 4 days

“I’ve been playing engines for over 15 years. This is the first time I’m playing an engine that feels like a person. When it makes mistakes, they are natural mistakes, like missing something in the midst of tricky tactics.”

“this is amazing to practice against…Definitely a 9.9/10 from me”

Max # of games started in an hour: 1,491

Top opponent by # of games: 149
Maia: A Framework to Understand Human Play

Since we can predict human play at different levels, there is a reliable, predictable, and maybe even algorithmically teachable difference between one human skill level and the next.
In this position, Maia levels 1100–1400 correctly predict White will play the tempting but wrong move b6 (the move played in the game).

Maia levels 1500–1900 predict that, on average, players rated 1500 and above will play the correct bxa6, forcing the Queenside open to decisive effect.
Maia: A Framework to Understand Human Play

Maia predicts mistakes surprisingly well.

When players make even the most glaring blunders, Maia predicts the exact move >25% of the time.

Performance increase relative to other models is larger for blunders.
Maia is a human-like neural network chess engine

Maia captures human style in chess at targeted skill levels

You can play Maia online now!

Next: algorithmic and data-driven improvement tools
Thanks!

“Aligning Superhuman AI with Human Behavior: Chess as a Model System”
Knowledge Discovery and Data Mining (KDD), 2020.

maiachess.com
@maia1 @maia5 @maia9 on Lichess

Reid McIlroy-Young
University of Toronto

Siddhartha Sen
Microsoft Research

Jon Kleinberg
Cornell University

Ashton Anderson
University of Toronto