école normale supérieure paris—saclay

Learning to Play Poker with Transformers

Exploring LLMs for Strategic Decision-Making

Anatole Vakili
Soël Megdoud
Julien Delavande

MVA Master Students

Paris-Saclay University

May 22, 2025

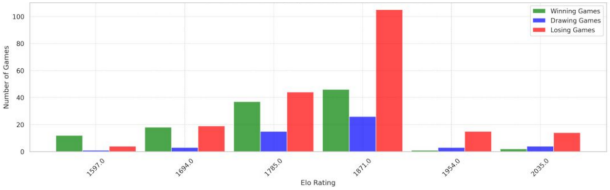
Chess as Inspiration for Poker

- GPT-3.5 Turbo fine-tuned on PGN chess games reaches **1800 ELO**
- Learns to play without explicit rules
- Predicts next move adapts style dynamically
- Outperforms ChatGPT in real chess matches

```
[Event "F/S Return Match"]
[Site "Belgrade, Serbia JUG"]
[Date "1992.11.04"]
[Round "29"]
[White "Fischer, Robert J."]
[Black "Spassky, Boris V."]
[Result "1/2-1/2"]

1. e4 e5 2. Nf3 Nc6 3. Bb5 a6 4. Ba4 Nf6 5. O-O Be7 6. Re1 b5 7. Bb3 d6 8. c3
O-O 9. h3 Nb8 10. d4 Nbd7 11. c4 c6 12. cxb5 axb5 13. Nc3 Bb7 14. Bg5 b4 15.
```

(a) PGN format of a chess game



(b) Win/Draw/LooseS of the model by ELO

Figure: Formats and rating system in chess

Poker as a Sequential Decision Problem

Similarities with Chess

- Hands can be represented as a sequence of actions (raise, fold, call, etc.) like a **natural sequence**
- There is no optimal strategy but good strategies depend on other profiles at the table

Differences

- Multiplayer (+2)
- Masked information
- Randomness

Our Data: Raw Poker Hand Logs

- 45,000 hands from online logs
- Extracted from professional player sessions

```
Game started at: 2016/11/29 15:25:37
Game ID: 787027613 0.50/1 (PRR) Karkadann (Hold'em)
Seat 7 is the button
Seat 1: StephCurry (105.78).
Seat 2: PANDAisEVIL (101).
Seat 3: AironVega (103.20).
Seat 4: IlxxxIT (43).
Seat 5: pineapplesand (40).
Seat 6: dankmann (206.52).
Seat 7: ElvenEyes (115).
Seat 8: gust (46).
Seat 9: VegetablesArentYummy (273.59).
Player gust has small blind (0.50)
Player VegetablesArentYummy has big blind (1)
Player gust received a card.
Player gust received a card.
Player VegetablesArentYummy received a card.
Player VegetablesArentYummy received a card.
Player StephCurry received a card.
Player StephCurry received a card.
Player PANDAisEVIL received a card.
Player PANDAisEVIL received a card.
Player AironVega received a card.
Player AironVega received a card.
Player IlxxxIT received card: [9d]
Player IlxxxIT received card: [7d]
Player pineapplesand received a card.
Player pineapplesand received a card.
Player dankmann received a card.
Player dankmann received a card.
Player ElvenEyes received a card.
Player ElvenEyes received a card.
```

```
Player StephCurry folds
Player PANDAisEVIL folds
Player AironVega folds
Player IlxxxIT folds
Player pineapplesand folds
Player dankmann raises (3)
Player ElvenEyes folds
Player gust folds
Player VegetablesArentYummy folds
Uncalled bet (2) returned to dankmann
Player dankmann mucks cards
------ Summary ------
Pot: 2.50. Rake 0
Player StephCurry does not show cards.Bets: 0. Collects: 0. Wins: 0.
Player PANDAisEVIL does not show cards.Bets: 0. Collects: 0. Wins: 0.
Player AironVega does not show cards.Bets: 0. Collects: 0. Wins: 0.
Player IlxxxIT does not show cards.Bets: 0. Collects: 0. Wins: 0.
Player pineapplesand does not show cards.Bets: 0. Collects: 0. Wins: 0.
*Player dankmann mucks (does not show cards). Bets: 1. Collects: 2.50. Wins: 1.50.
Player ElvenEyes does not show cards.Bets: 0. Collects: 0. Wins: 0.
Player gust does not show cards.Bets: 0.50. Collects: 0. Loses: 0.50.
Player VegetablesArentYummy does not show cards.Bets: 1. Collects: 0. Loses: 1.
Game ended at: 2016/11/29 15:26:36
```

Raw log of a poker hand

Conversion to usable format

JSON Part 1


```
{"date": "2016/9/4 1:55:32",
  "game_id": "718933960",
  "variant": "PRR",
  "table_name": "Monopod",
  "type_game": "Short",
  "button_seat": 2,
  "players": [
    "BIGRAISE", "tcoll16", "cracypoker", "vege",
    "bjv1105", "IlxxxlI", "WalterBlack",
   ↪  "TheFront7", "NoSugarJoe"],
  "players_seats": [1, 2, 3, 4, 5, 6, 7, 8, 9],
  "starting_stacks": [120.0, 216.0, 227.55, 213.5,
 ↪  522.98, 80.0, 179.55, 263.95, 100.0],
  "player_small_blind": "cracypoker",
  "small_blind": 2.0,
  "player_big_blind": "vege",
  "big_blind": 4.0,
  "player": "IlxxxlI",
  "cards_player": ["4c", "5h"],
  "dealed_cards": {
    "flop": ["3d", "Qs", "2s"],
    "turn": ["Ks"],
    "river": ["6s"]},
```


JSON Part 2


```
  "actions": {
    "pre-flop": {
      "players": ["bjv1105", "IlxxxlI",
     ↪  "WalterBlack", "TheFront7",
                  "NoSugarJoe", "BIGRAISE",
                 ↪  "tcoll16",
                  "cracypoker", "vege"],
      "actions": ["RAISE", "FOLD", "CALL", "FOLD",
     ↪  "CALL", "FOLD",
                  "FOLD", "FOLD", "FOLD"],
      "values": [12.0, null, 12.0, null, 8.0, null,
     ↪  null, null, null]
    },
    "post-flop": {},
    "post-turn": {},
    "post-river": {}
  },
  "card_shown_by_players": [null, null, null, null,
 ↪  null, null, "Ac 3s", null, null],
  "finishing_stack": [120.0, 216.0, 225.55, 209.5,
 ↪  489.18, 80.0, 254.1, 263.95, 64.2]
}
```

Creating Suitable Sequence Format

Advantages of this format

- No peudos
- Stacks converted in BB
- Stacks precomputed
- Small vocabulary (90 tokens) so smaller models

```
['[TABLE_CONFIGURATION]', '[PREFLOP]', '[STACKS]', '[RIVER]', '[FLOP]',
 '[TURN]', 'CHECK', 'RAISE', 'ALLIN', '[PAD]', '[EOS]', 'FOLD', 'CALL',
 'POT', 'BTN', 'BET', 'BB', 'SB', '2s', '3s', '4s', '5s', '6s', '7s',
 '8s', '9s', 'Ts', 'Js', 'Qs', 'Ks', 'As', '2h', '3h', '4h', '5h', '6h',
 '7h', '8h', '9h', 'Th', 'Jh', 'Qh', 'Kh', 'Ah', '2d', '3d', '4d', '5d',
 '6d', '7d', '8d', '9d', 'Td', 'Jd', 'Qd', 'Kd', 'Ad', '2c', '3c', '4c',
 '5c', '6c', '7c', '8c', '9c', 'Tc', 'Jc', 'Qc', 'Kc', 'Ac', '0', '1',
 '2', '3', '4', '5', '6', '7', '8', '9', '.', ':', '=', '\n', ' ', '[', ']']
```

Token vocabulary

```
[TABLE_CONFIGURATION]
BTN=P7
SB=P1 0.5BB
BB=P2 1BB

[STACKS]
P1: 174.3BB
P2: 126.2BB
P3: 195.6BB
P4: 62.1BB [Ad Kd]
P5: 98.4BB
P6: 190.2BB
P7: 40.0BB
POT=1.5BB

[PREFLOP]
P3: FOLD
P4: RAISE 3BB
P5: CALL 3BB
P6: CALL 3BB
P7: FOLD
P1: FOLD
P2: FOLD

[STACKS]
P4: 58.6BB [Ad Kd]
P5: 94.9BB
P6: 186.7BB
POT=12.0BB
```

```
[FLOP][10d 2h 5s]
P4: CHECK
P5: CHECK
P6: CHECK

[STACKS]
P4: 58.6BB [Ad Kd]
P5: 94.9BB
P6: 186.7BB
POT=12.0BB

[TURN][10d 2h 5s 2c]
P4: CHECK
P5: CHECK
P6: CHECK

[STACKS]
P4: 58.6BB [Ad Kd]
P5: 94.9BB
P6: 186.7BB
POT=12.0BB

[RIVER][10d 2h 5s 2c 4d]
P4:
CHECK
```

Input format

Training Pytorch Models from Scratch

Training configuration

- Architecture suited to small vocabulary
- Trained several Transformers model (1M to 100M params)
- Used 52K hands for training

Difficulties

- Always predicts ”FOLD”
- Only learnt the format but not the game
- Class balancing in CrossEntropy did not improve this

```
Token: 'CHECK' | Correct: 0 / 1103 occurrences | Accuracy: 0.00%
Token: 'RAISE' | Correct: 0 / 1226 occurrences | Accuracy: 0.00%
Token: 'FOLD' | Correct: 6603 / 6603 occurrences | Accuracy: 100.00%
Token: 'CALL' | Correct: 0 / 773 occurrences | Accuracy: 0.00%
Token: 'ALLIN' | Correct: 0 / 156 occurrences | Accuracy: 0.00%
Token: 'BET' | Correct: 0 / 624 occurrences | Accuracy: 0.00%
Token: 'POT' | Correct: 0 / 0 occurrences | Accuracy: 0.00%
Token: 'BB' | Correct: 2778 / 2779 occurrences | Accuracy: 99.96%
Token: 'SB' | Correct: 0 / 0 occurrences | Accuracy: 0.00%
Token: 'BTN' | Correct: 0 / 0 occurrences | Accuracy: 0.00%
Token: '[PAD]' | Correct: 0 / 31713 occurrences | Accuracy: 0.00%
Token: '[EOS]' | Correct: 10485 / 10485 occurrences | Accuracy: 100.00%
Token: '1' | Correct: 0 / 595 occurrences | Accuracy: 0.00%
Token: '2' | Correct: 960 / 1015 occurrences | Accuracy: 94.58%
Token: '3' | Correct: 0 / 533 occurrences | Accuracy: 0.00%
Token: '4' | Correct: 0 / 168 occurrences | Accuracy: 0.00%
Token: '5' | Correct: 0 / 139 occurrences | Accuracy: 0.00%
Token: '6' | Correct: 0 / 161 occurrences | Accuracy: 0.00%
Token: '7' | Correct: 0 / 130 occurrences | Accuracy: 0.00%
Token: '8' | Correct: 0 / 76 occurrences | Accuracy: 0.00%
Token: '9' | Correct: 0 / 101 occurrences | Accuracy: 0.00%
```

Performance analysis of Poker models

Fine-tuning a Pre-trained LLM

Model: SmolLM-135M (Non-Instruct)

- smallest model: 135M parameters
- no prior poker knowledge

Training

- Specific Dataset format
- Hugging Face `Trainer` API.

Does the Model Actually Play Poker?

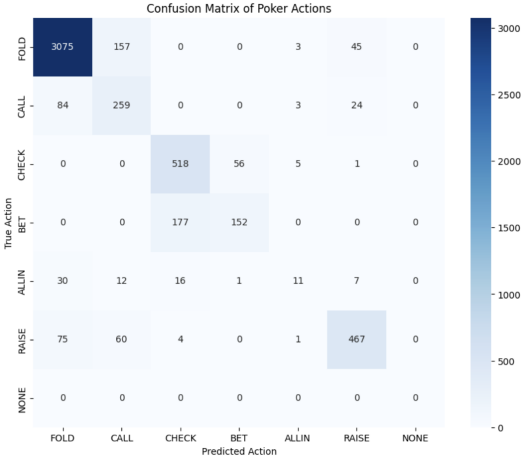


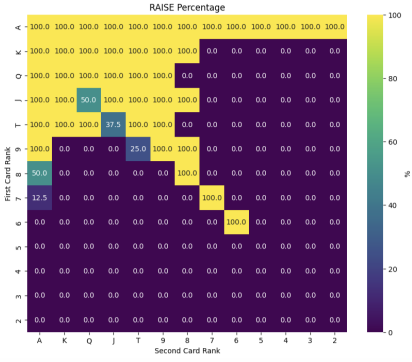
Figure: Fold Percentage Heatmap (Model vs Test Set)

Key Findings:

- The model makes statistically good decisions on unseen hands.
- It successfully generalizes beyond memorization.
- A pre-trained LLM might be necessary for stronger strategic understanding.

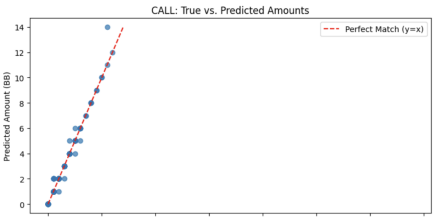
Does the Model Understand Hand Ranges?

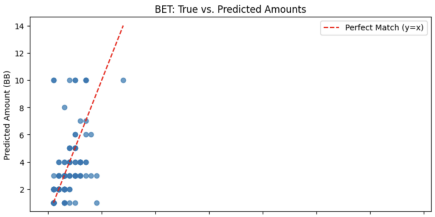
AA	AKs	AQs	AJs	ATs	A9s	A8s	A7s	A6s	A5s	A4s	A3s	A2s
AKo	KK	KQs	KJs	KTs	K9s	K8s	K7s	K6s	K5s	K4s	K3s	K2s
AQo	KQo	QQ	QJs	QTs	Q9s	Q8s	Q7s	Q6s	Q5s	Q4s	Q3s	Q2s
AJo	KJo	QJo	JJ	JTs	J9s	J8s	J7s	J6s	J5s	J4s	J3s	J2s
ATo	KTo	QTo	JTo	TT	T9s	T8s	T7s	T6s	T5s	T4s	T3s	T2s
A9o	K9o	Q9o	J9o	T9o	99	98s	97s	96s	95s	94s	93s	92s
A8o	K8o	Q8o	J8o	T8o	98o	88	87s	86s	85s	84s	83s	82s
A7o	K7o	Q7o	J7o	T7o	97o	87o	77	76s	75s	74s	73s	72s
A6o	K6o	Q6o	J6o	T6o	96o	86o	76o	66	65s	64s	63s	62s
A5o	K5o	Q5o	J5o	T5o	95o	85o	75o	65o	55	54s	53s	52s
A4o	K4o	Q4o	J4o	T4o	94o	84o	74o	64o	54o	44	43s	42s
A3o	K3o	Q3o	J3o	T3o	93o	83o	73o	63o	53o	43o	33	32s
A2o	K2o	Q2o	J2o	T2o	92o	82o	72o	62o	52o	42o	32o	22

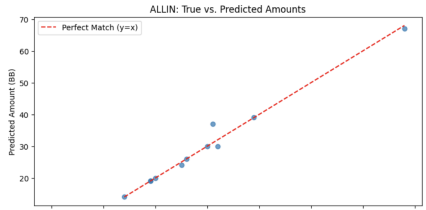


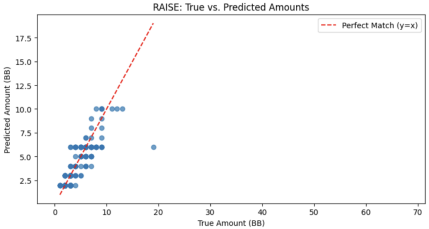
- The model mirrors pro-like tendencies. Overall range structure aligns well with optimal play.
- Premium pairs are preferred for aggression. Suited hands are more valued than offsuit hands.

How Well Does the Model Size Its Bets?









Surprisingly, the model exhibits coherent bet sizing patterns, aligning well with human decision-making.

Conclusion

Key Takeaways

- Training a language model from scratch for poker was unsuccessful—likely due to the complexity of the task and limited data.
- However, fine-tuning a pre-trained LLM led to impressive results, demonstrating a solid understanding of poker strategy.

Next Steps

- Test the model in a real poker environment against naive strategies, Monte Carlo agents, or even human players.
- Expand the dataset to improve generalization.
- Integrate reinforcement learning (RL): train the model through self-play with a reward based on its earnings per hand.
- Enhance inference with more compute, such as using $<$ think $>$ tokens for better decision-making.