

Can Mistral LLM win at Poker?

Mistral Fine-tuning Hackathon, June 2024

Julien Delavande
Anatole Vakili
Soël Megdoud



Plan

1. Inspiration and context
2. The case of Poker
3. Our methodology
4. Results and performance
5. Key improvements areas



Recent experiences show that GPT 3.5 Turbo trained on standardized chess games achieves 1800 ELO

Inspiration & context

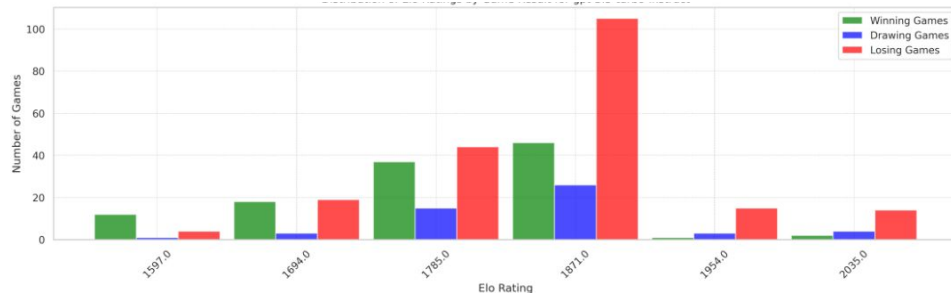
LLM fine-tuned on PGN chess games:

- Returns the most likely next move
- Adapts its style to the previous sequence of moves.
- Is able to play chess without knowing the rules, with few forbidden moves.
- Outperforms ChatGPT and reaches impressive 1800 ELO

```
[Event "F/S Return Match"]  
[Site "Belgrade, Serbia JUG"]  
[Date "1992.11.04"]  
[Round "29"]  
[White "Fischer, Robert J."]  
[Black "Spasky, 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.  
Nh1 b6 16. Bb4 c5 17. dxc5 Nxc5 18. Bxc5 Qxc5 19. exd6 Qf6 20. Nhd2 Nxd6 21.
```

PGN chess game format



Distribution of ELO ratings

Finetuning an LLM the same way could be interesting for Poker specificities (missing information, bluffing, etc.)

The case of Poker

Poker specificities



Missing information:

Not all knowledge of the environment is available (hidden cards).



Bluffing and psychology:

Poker involves bluffing and not playing like a robot to remain unpredictable.



Non-deterministic environment:

Poker involves chance with card shuffling and dealing.



Multi player:

Poker games involve a dynamic number of players

Opportunities for LLMs

Games format:

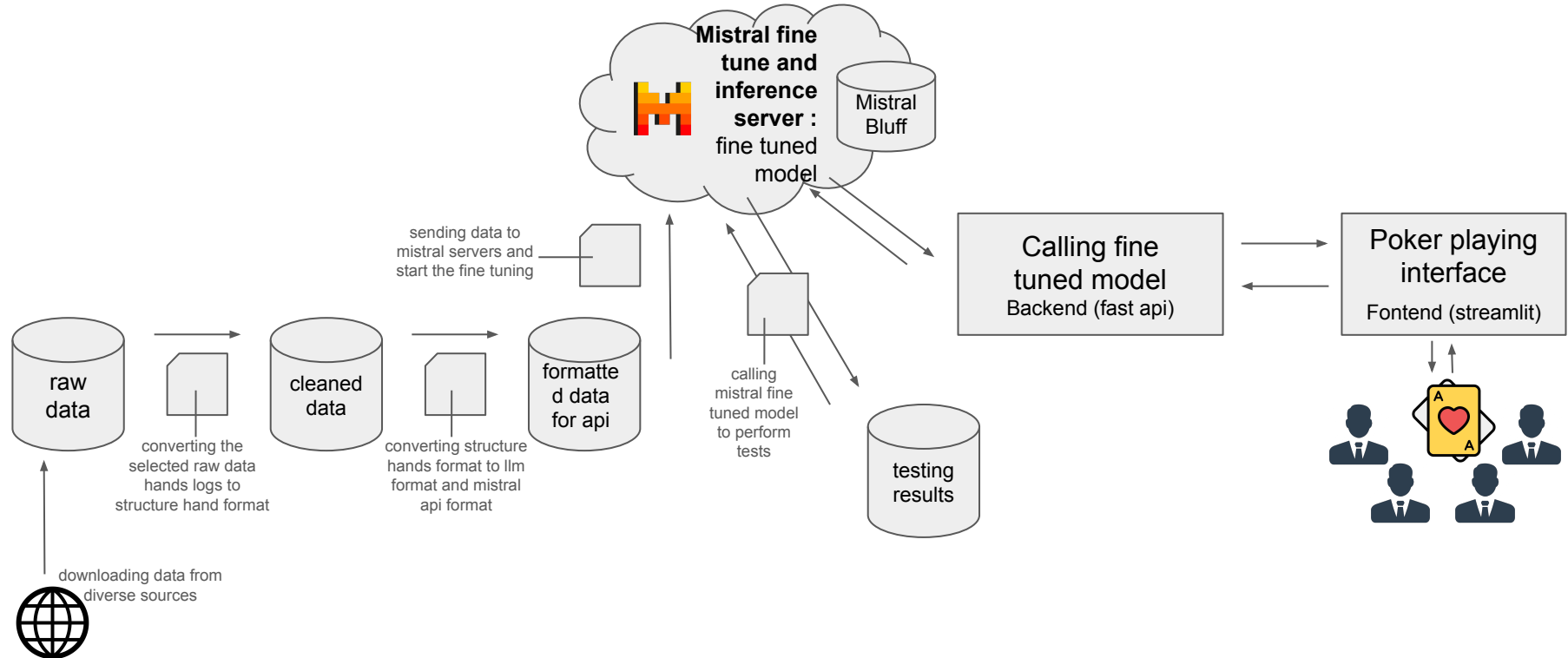
Poker games can be structured in a PGN-like format, with contextual information and a sequence of actions.

Adaptable style of play:

LLMs may understand opponents' style of play and exploit advantageous situations.

A 3-steps pipeline for finetuning has been implemented with a backend and a frontend for online testing

Our methodology



Finetuning led to MistralBluff, an LLM playing only legit moves, similar to the pros

Results and performance: open range UTG position

Professionals vs MistralBluff moves

	bets	calls	raises	allin	checks	folds	caps
bets	16	0	0	0	20	0	0
calls	0	38	5	0	0	59	0
raises	0	4	45	0	0	46	0
allin	0	0	0	2	1	6	0
checks	8	0	1	0	151	0	0
folds	0	19	15	0	0	3017	0
caps	0	0	0	0	0	1	0
nombre de coup légal: 3454 3454							

Professionals vs Mistral non fine tuned moves

	bets	calls	raises	allin	checks	folds	caps
bets	14	0	0	0	8	0	0
calls	0	16	1	0	0	0	0
raises	0	3	5	0	0	0	0
allin	0	1	0	0	1	0	0
checks	36	0	0	0	44	0	0
folds	7	87	172	0	48	1	0
caps	0	1	0	0	0	0	0
nombre de coup légal: 429 3454							

- **MistralBluff did not play a single forbidden move** (3454/3454) against 3025/3459 forbidden for Mistral AI non fine-tuned
- **MistralBluff plays as professional for most of the moves** not as Mistral AI non fine-tuned

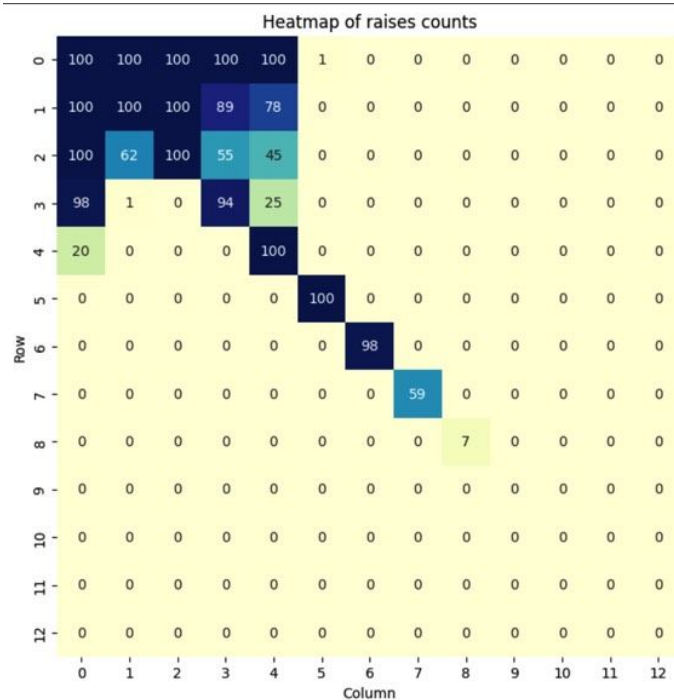
MistralBluff has almost the same opening ranges as professional players

Results and performance: opening range UTG position

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

Professional player UTG openraise range

- MistralBluff opening range at UTG position is very close to professional poker players
- It understood that suited hands are better than off-suited ones



MistralBluff UTG openraise range

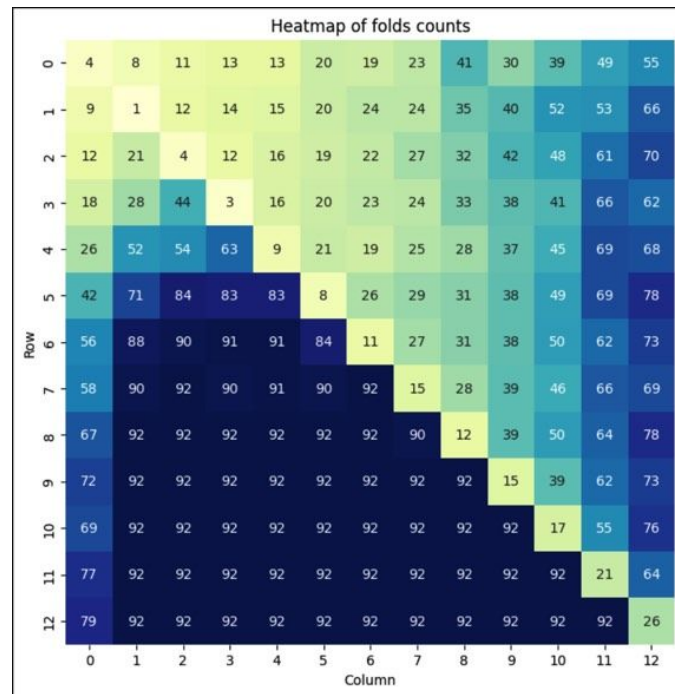
MistralBluff changes its opening range according to its position at the table, as professional players

Results and performance: open range BB position

	A	K	Q	J	T	9	8	7	6	5	4	3	2	
A	6.1	3.3	2.7	2.3	1.6	1.2	0.8	0.5	0.2	0.2	0.1	0EV	-0.1	
K	3.0	5.0	0.7	0.3	-0.1	-0.6	-0.9	-0.9	-0.9	-1.0	-1.1	-1.2	-1.2	
Q	2.4	0.2	4.3	-0.1	-0.3	-0.6	-0.8	-1.0	-1.0	-1.1	-1.2	-1.3	-1.3	
J	2.0	-0.2	-0.6	3.8	-0.2	-0.5	-0.7	-0.9	-1.1	-1.2	-1.2	-1.4	-1.4	3-bet/Call
T	1.5	-0.6	-0.8	-0.7	3.2	-0.4	-0.6	-0.8	-1.0	-1.3	-1.3	-1.4	-1.5	
9	0.8	-1.1	-1.2	-1.0	-0.9	2.7	-0.6	-0.8	-1.0	-1.2	-1.5	x	x	3-bet shove
8	0.4	-1.4	-1.4	-1.3	-1.2	-1.1	2.3	-0.7	-0.9	-1.1	x	x	x	
7	0.1	-1.4	-1.6	-1.5	-1.4	-1.3	-1.2	1.9	-0.8	-1.0	x	x	x	Call
6	-0.2	-1.5	-1.6	-1.7	-1.7	-1.5	-1.4	-1.3	1.5	-0.9	-1.1	x	x	
5	-0.3	-1.6	-1.7	-1.8	x	x	x	x	x	1.0	-1.0	x	x	Fold
4	-0.4	-1.6	-1.8	-1.8	x	x	x	x	x	x	0.7	x	x	
3	-0.5	-1.7	-1.9	-2.0	x	x	x	x	x	x	x	0.4	x	
2	-0.6	-1.8	-2.0	-2.0	x	x	x	x	x	x	x	x	0EV	

Professional player BB opening range

- MistralBluff learned to play “loose” when at Big Blind position
- MistralBluff opening range is consistent with professional players



MistralBluff BB opening range

Despite these promising results, there are still many areas for improvement

Key improvements areas

- **Sizing:** MistralBluff is bad at raise sizing, sizing could be improved in several ways: using bet sizes relative to the pot or blinds, using an agent, using external functions with simple rules.
- **Data:** One of the main difficulties was finding quality data. MistralBluff was trained on data from single professional player and representing 8M tokens. By having more quality hands, MistralBluff could see more game configurations.
- **Evaluation:** It is difficult to evaluate the level of a poker player, as there is no easily calculated ELO score. There are few metrics to evaluate post-flop play and we would have needed more time to test on dozens of online games.