Learning Chess Blindfolded: Evaluating Language Models for State Tracking

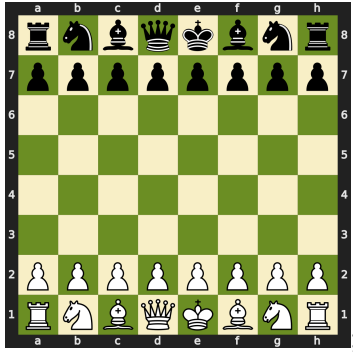
Shubham Toshniwal

Research Scientist, FAIR NYC

Entity Tracking in Chess

Test out ideas for entity tracking via language models in chess

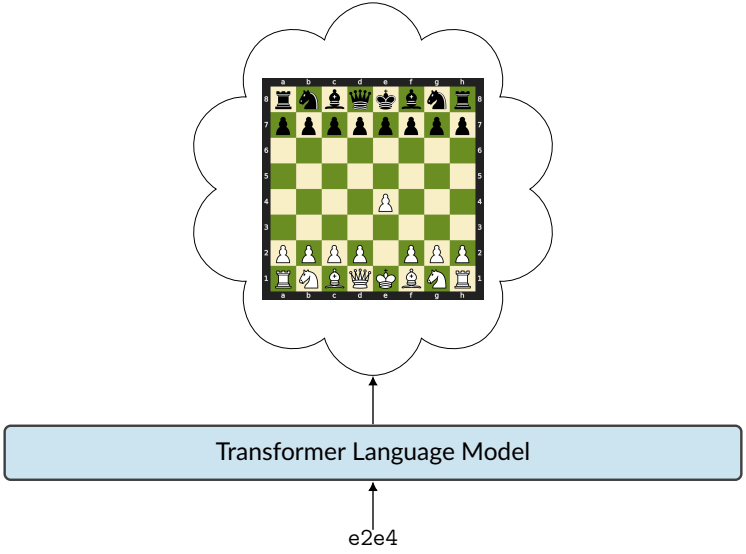
Why Chess? *Simple, closed domain*



;

Entities: Chess pieces
Entity State: Piece Location

Learning Chess Blindfolded



Learning Chess Blindfolded

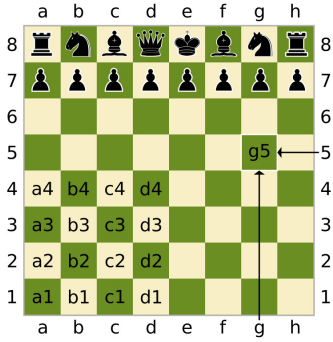
g1f3 d7d5 g2g3 ....

d2d4 d7d5 g1f3 ....

e2e4 e7e5 g1f3 ....

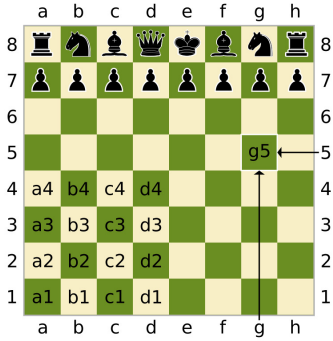
Algebraic Notation

Position Naming



Algebraic Notation

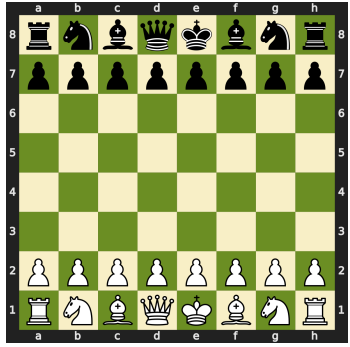
Position Naming



Piece Types

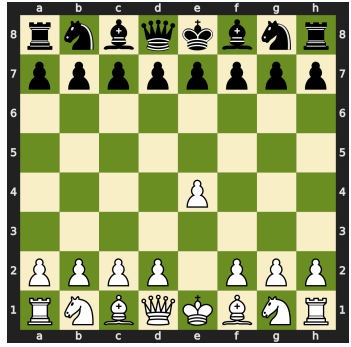
Rook	Knight	Bishop	Queen	King	Pawn
R	N	B	Q	K	P

Chess Notation



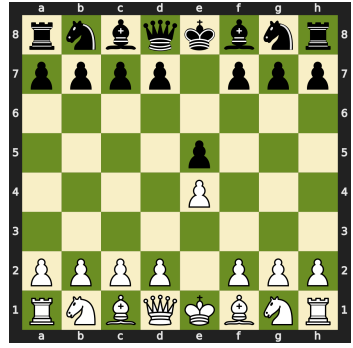
Translation of moves

Chess Notation



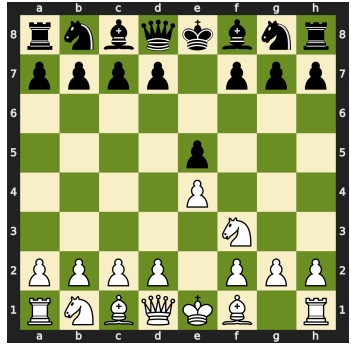
Translation of moves	
`e2e4`	(Pawn) moved from `e2` to `e4`

Chess Notation



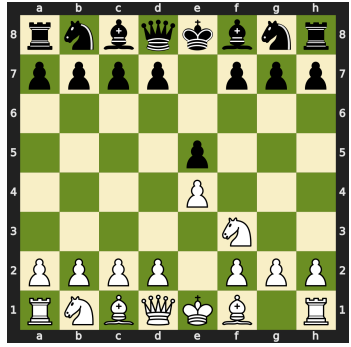
Translation of moves	
`e2e4`	(Pawn) moved from `e2` to `e4`
`e7e5`	(Pawn) moved from `e7` to `e5`

Chess Notation



Translation of moves	
`e2e4`	(Pawn) moved from `e2` to `e4`
`e7e5`	(Pawn) moved from `e7` to `e5`
`g1f3`	(Knight) moved from `g1` to `f3`

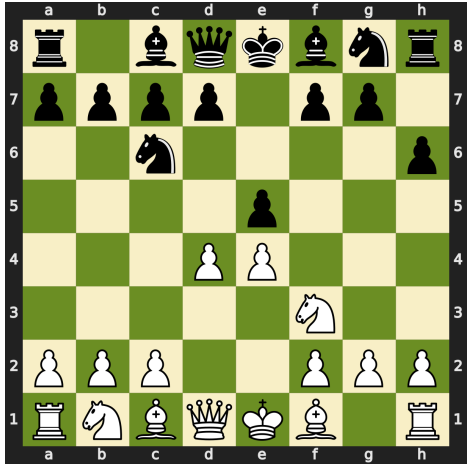
Chess Notation



Translation of moves	
`e2e4`	(Pawn) moved from `e2` to `e4`
`e7e5`	(Pawn) moved from `e7` to `e5`
`g1f3`	(Knight) moved from `g1` to `f3`
⋮	⋮

Entity Tracking Task: Ending Square

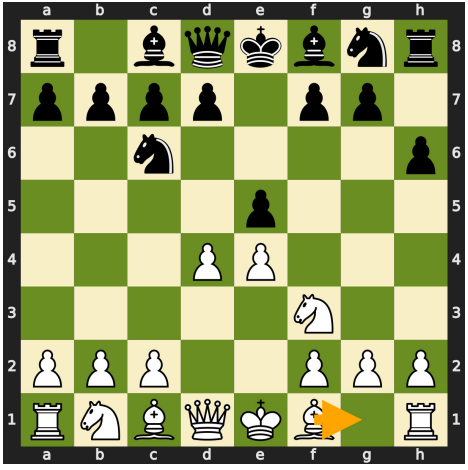
Chess Notation allows for probing for entity state via prompting!



e2e4 e7e5 g1f3 b8c6 d2d4 h7h6 f1??

Entity Tracking Task: Ending Square

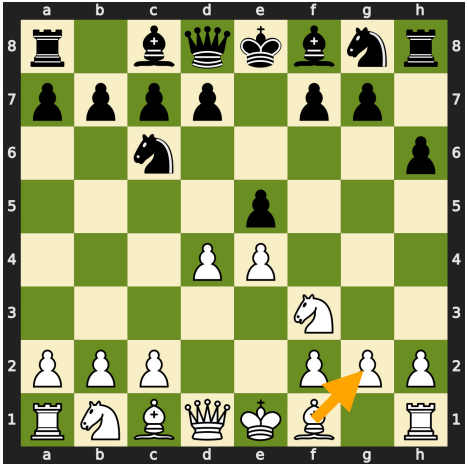
Chess Notation allows for probing for entity state via prompting!



e2e4 e7e5 g1f3 b8c6 d2d4 h7h6 f1g1

Entity Tracking Task: Ending Square

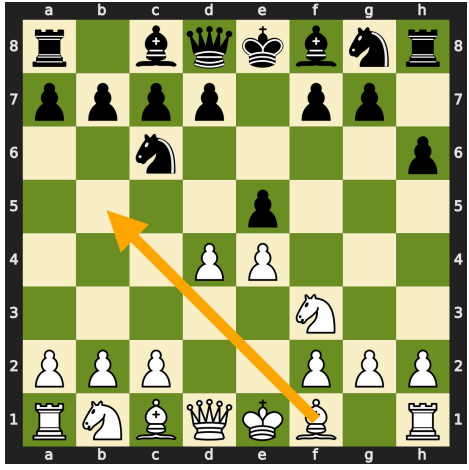
Chess Notation allows for probing for entity state via prompting!



e2e4 e7e5 g1f3 b8c6 d2d4 h7h6 f1g2

Entity Tracking Task: Ending Square

Chess Notation allows for probing for entity state via prompting!



e2e4 e7e5 g1f3 b8c6 d2d4 h7h6 f1b5

Randomly Annotated Piece Type (RAP)

Can a language model benefit from the knowledge of piece types?

Randomly Annotated Piece Type (RAP)

Can a language model benefit from the knowledge of piece types?

Randomly introduce piece types in text sequences during training

Vanilla Training	`e2e4 e7e5 g1f3 b8c6 d2d4 h7h6`

Randomly Annotated Piece Type (RAP)

Can a language model benefit from the knowledge of piece types?

Randomly introduce piece types in text sequences during training

Vanilla Training	`e2e4 e7e5 g1f3 b8c6 d2d4 h7h6`
+ RAP (p=15)	`e2e4 e7e5 Ng1f3 b8c6 d2d4 h7h6`

Piece Types

Knight
N

Pawn
P

Randomly Annotated Piece Type (RAP)

Can a language model benefit from the knowledge of piece types?

Randomly introduce piece types in text sequences during training

Vanilla Training	`e2e4 e7e5 g1f3 b8c6 d2d4 h7h6`
+ RAP (p=15)	`e2e4 e7e5 Ng1f3 b8c6 d2d4 h7h6`
+ RAP (p=50)	`Pe2e4 e7e5 Ng1f3 b8c6 d2d4 Ph7h6`
+ RAP (p=100)	`Pe2e4 Pe7e5 Ng1f3 Nb8c6 Pd2d4 Ph7h6`

Randomly Annotated Piece Type (RAP)

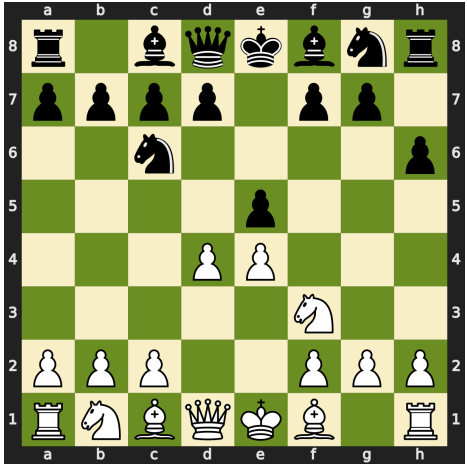
Can a language model benefit from the knowledge of piece types?

Randomly introduce piece types in text sequences during training

Vanilla Training	`e2e4 e7e5 g1f3 b8c6 d2d4 h7h6`
+ RAP (p=15)	`e2e4 e7e5 Ng1f3 b8c6 d2d4 h7h6`
+ RAP (p=50)	`Pe2e4 e7e5 Ng1f3 b8c6 d2d4 Ph7h6`
+ RAP (p=100)	`Pe2e4 Pe7e5 Ng1f3 Nb8c6 Pd2d4 Ph7h6`
Inference	`e2e4 e7e5 g1f3 b8c6 d2d4 h7h6`

Entity Tracking Task: Starting Square

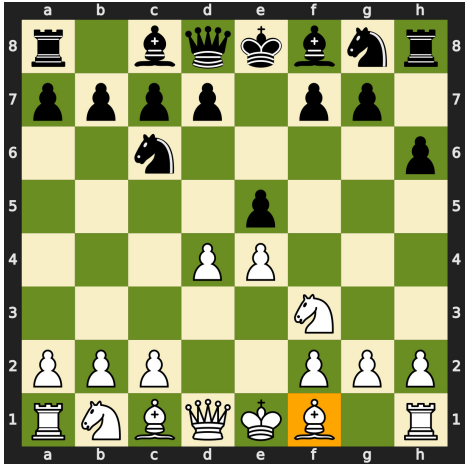
Training with RAP also allows for directly probing for piece location



e2e4 e7e5 g1f3 b8c6 d2d4 h7h6 B??

Entity Tracking Task: Starting Square

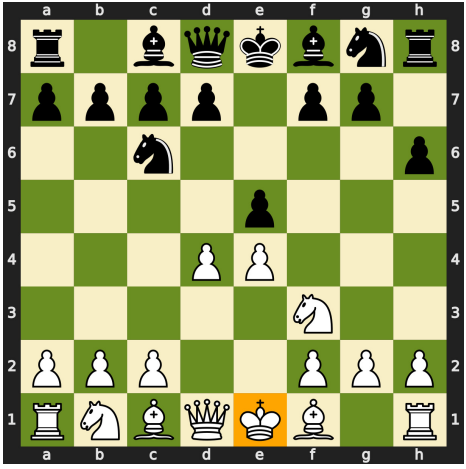
Training with RAP also allows for directly probing for piece location



e2e4 e7e5 g1f3 b8c6 d2d4 h7h6 B**f1**

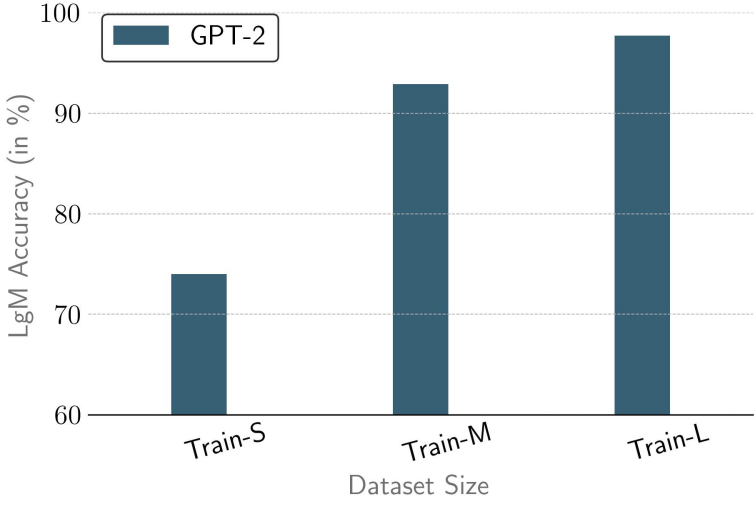
Entity Tracking Task: Starting Square

Training with RAP also allows for directly probing for piece location

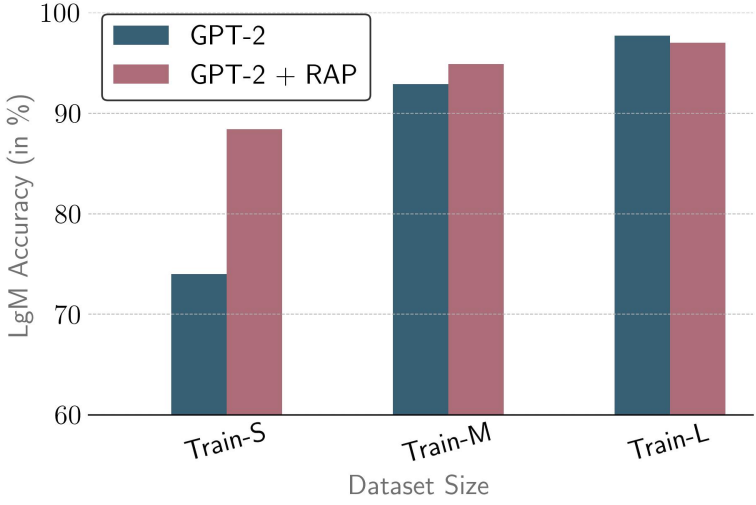


e2e4 e7e5 g1f3 b8c6 d2d4 h7h6 Be1

Entity Tracking Results



Entity Tracking Results



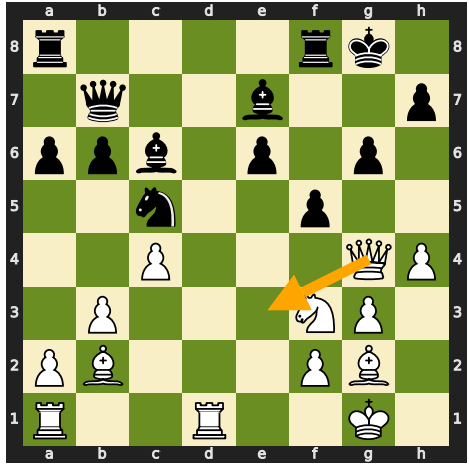
Error Categories

Automated error analysis possible for domains such as chess

Error categories:

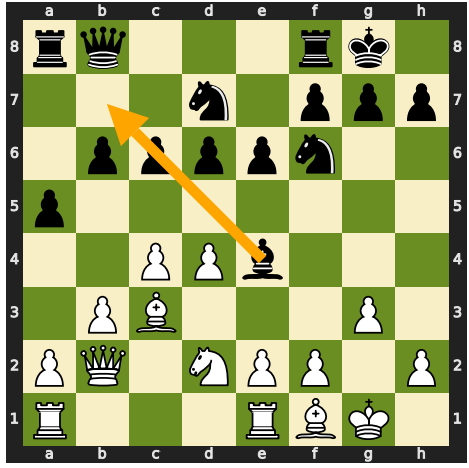
- Syntax
- Path Obstruction
- Pseudo Legal

Error Category: Syntax



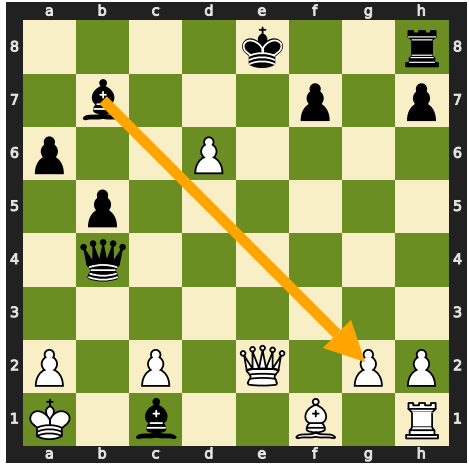
Queen trying to move like a knight

Error Category: Path Obstruction



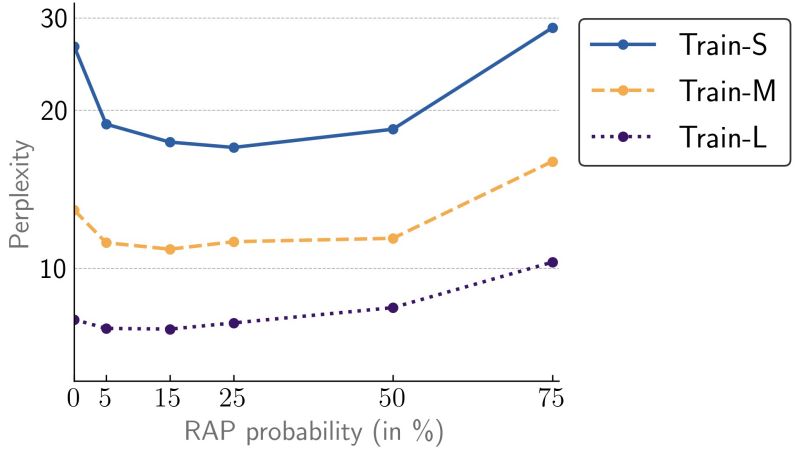
Bishop eager to retreat

Error Category: Pseudo Legal



Protect the king first

Language Modeling Results



Takeaways

Proposed chess as a testbed for entity tracking in language models

Data augmentation using RAP improves both entity tracking and language modeling results for low data settings