

Learning to Generate Move-by-Move Commentary for Chess Games from Large-Scale Social Forum Data: APPENDIX

Appendix A: Additional Data Examples

Text	Categories
Unpins and defends the knight , but it does n't matter , as the time is ripe .	Desc
He gets fed up and exchanges Queen for Rook .	Desc
Rxc3 , I just retake with my queen , whilst if he attempts defense with the bishop , then after 17.Bd2 , Ne4 , 18.Rxc3 , Nxc3 , 19.Rxc6 , Nxc1 , I 've won a rook outright .	Desc,Rationale
Preparing to castle , and threatening now white 's e pawn for real.	Desc
Simply getting my rook off that dangerous diagonal , and protecting the b pawn .	Desc
I throw in a check	Desc
Threatening mate with Qxh2	Desc,Quality
A punch drunk move !	Quality
This is not the best move.	Quality
The most logical move.	Quality
This move is dubious.	Quality
The check gains time to support the advance of the a-paw maybe Ke1 was better	Desc,Quality
I did n't want to retreat the N and I rejected 11 .	Rationale
I wish to both defend the pawn , and threaten indirectly the black queen , gaining a tempo	Rationale
it would suite me better if my opponent made a queenside castling , since then my advanced pawn on the d-file would assist in a future attack on the king 's position .	Comparative
but better would be nd2 to get the knight in the game , the queen rook , too .	Comparitive
i think it would have been better to play nxe5 and maintain a material advantage .	Comparitive
although not as effective as the bishop move , even 10.0-0-0 is better than the text , though 10 ... bg4 would have been very nasty .	Comparitive
fianchettoing , so that when black does complete his development , his b will be on a better diagnol .	Comparitive
He doesn't notice that his Knight is hanging ...	GameInfo
Now of course my forces are anchored around the pawns on e3 and h5 , and the black rook loses his hope of penetrating the white position on the e-file	GameInfo
Well, now the game will get interesting soon	GeneralInfo
He tries his trick , which of course is noticed	GeneralInfo
This is often what I will do , when I 'm playing white.	GeneralInfo

Table 1: Some commentary texts from each of the six categories. The **Categories** column lists those into which the example falls. As pointed out earlier, the category labels are not exclusive i.e. a text can belong to multiple categories, though texts with more than one category are few in our dataset. ('Desc' is shor for 'Move Description')

Appendix B: Additional details for methods

Templates

- *Move Description*: For the Move Description category, we consider following templates:
 1. **Capture** moves : [PLAYERMOVED] captures the [CAPTUREDPIECE] at [FINALSQUARE] using the [PIECEMOVED] at [INITIALSQUARE].
 2. **Non-Capture** moves: [PLAYERMOVED] moves the [PIECEMOVED] from [INITIALSQUARE] to [FINALSQUARE].
 3. **Castling** moves: [PLAYERMOVED] does a castling.

For moves which lead to a CHECK in the resultant board state, an additional *putting the king in check* is added to the template. [PLAYERMOVED] (Black/White), [INITIALSQUARE], [FINALSQUARE], [PIECEMOVED] are filled in based on the move description on the input side.

- *Move Quality*: Based on the move score (as calculated by the chess engine *Stockfish*) $> \theta$ or $< \theta$, one of the following two is generated:
 1. A good move.
 2. A bad move. The threshold θ is found by tuning it on the validation set to maximize BLEU. We start from $\theta = 0$.

Appendix C: Qualitative examples

Some qualitative examples.

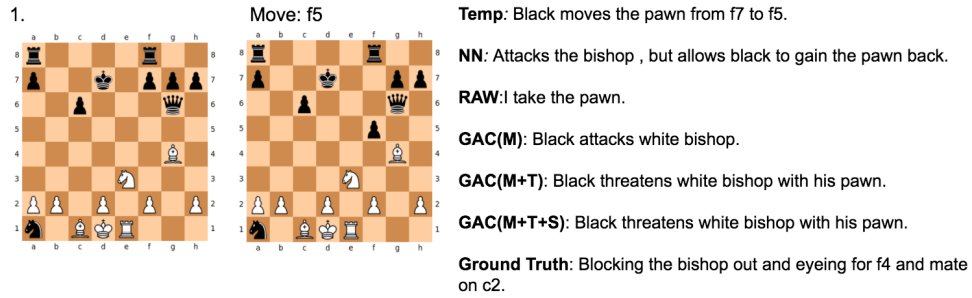


Figure 1: Example output 1: Move description subset of data.

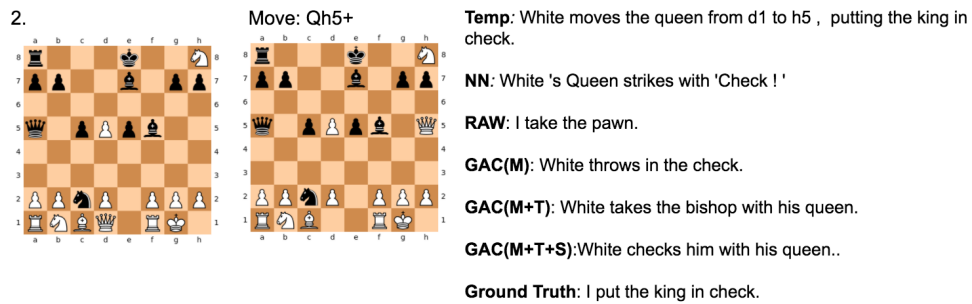


Figure 2: Example output 2: Move description subset of data.

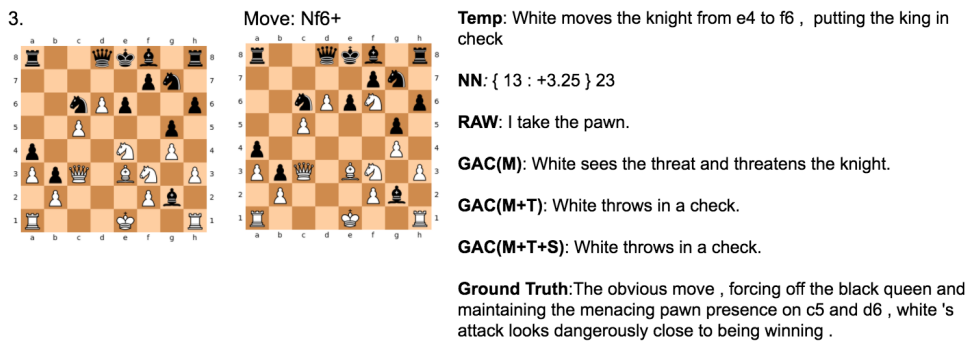


Figure 3: Example output 3: Move description subset of data.

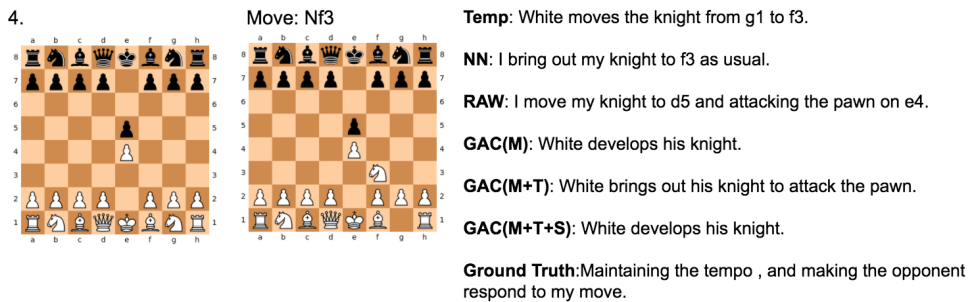




Figure 4: Example output 4: Move description subset of data.

5.  **Move: Qd6** 

Temp: Black moves the queen from d8 to d6

NN: Black brings out his queen , which breaks the queen and provides protection the the e-pawn

RAW: I take the pawn.



GAC(M): Black attacks white queen.

GAC(M+T): Black brings out his queen to the attack .

GAC(M+T+S): Black brings his queen out

Ground Truth: Letting the Queen join the attack.

Figure 5: Example output 5: Move description subset of data.

6.  **Move: a5** 

Temp: Black moves the pawn from a6 to a5

NN: The pawn advance begins.

RAW: I take the pawn.

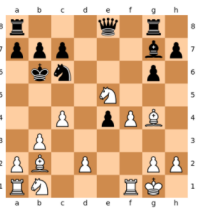

GAC(M): Black brings up his bishop to cover his bishop.

GAC(M+T): Black pushes up his pawn to free white pawn.

GAC(M+T+S): Black develops his bishop and attack the pawn.

Ground Truth: Bg4 links the queen and the rock , which would free up black to take the e5 pawn and even out material .

Figure 6: Example output 6: Move description subset of data.

7.  **Move:: Nd7+** 

Temp: White moves the knight from e5 to d7 , putting the king in check

NN: ... with the only piece near my king and no protection .

RAW: I take the pawn.

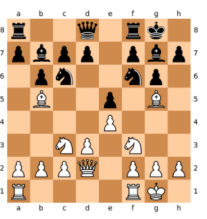
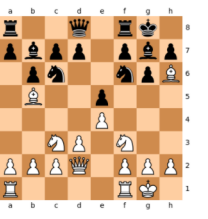
GAC(M): White sees the check and threatens the knight.

GAC(M+T): White throws in a check.

GAC(M+T+S): White checks him with his knight.

Ground Truth: I want to rid blacks DSB 's reduce the outposts for his knight and hopefully overwork his superior piece with my minor pieces..

Figure 7: Example output 7: Move description subset of data.

8.  **Move notation: Bh6** 

Temp: White moves the bishop from g5 to h6

NN: Bxd8 Qxd8

RAW: I take the pawn.

GAC(M): White develops his bishop .

GAC(M+T): White decides to exchange his bishop .

GAC(M+T+S): White develops his bishop to the pin and the bishop .

Ground Truth: ... Bxh6 10 .

Figure 8: Example output 8: Move description subset of data.

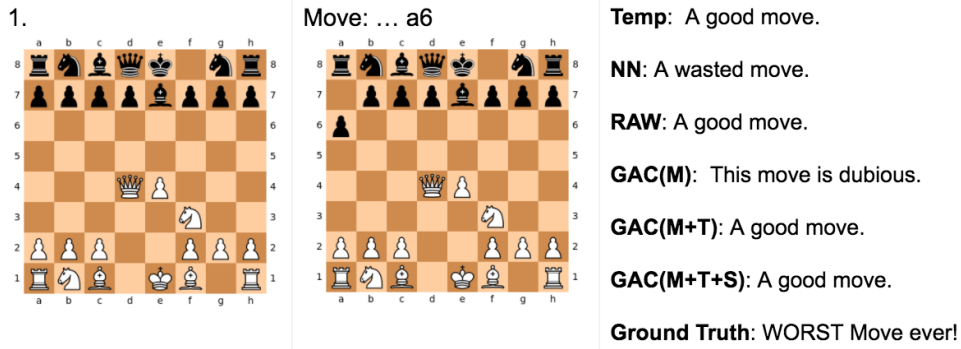


Figure 9: Example output 1: Move quality subset of data.

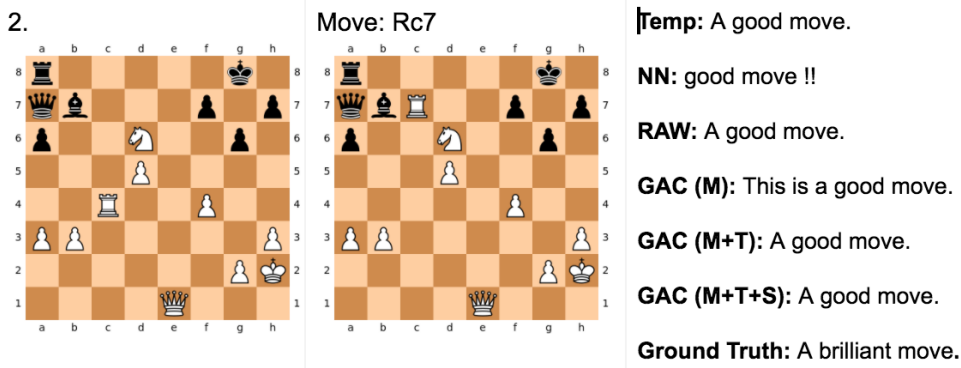


Figure 10: Example output 2: Move quality subset of data.

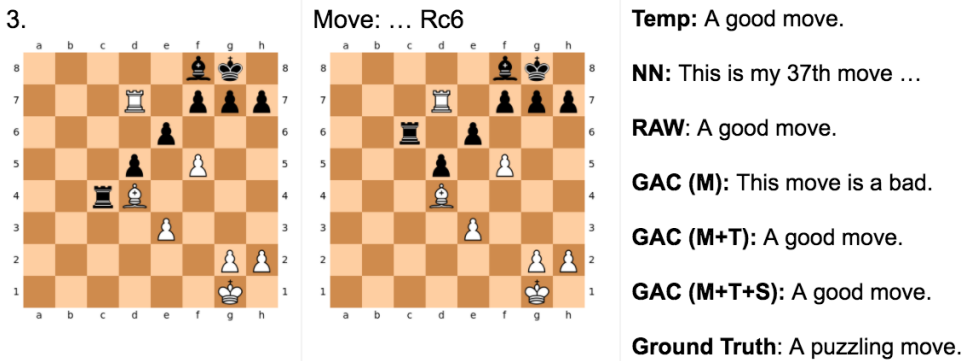


Figure 11: Example output 3: Move quality subset of data.

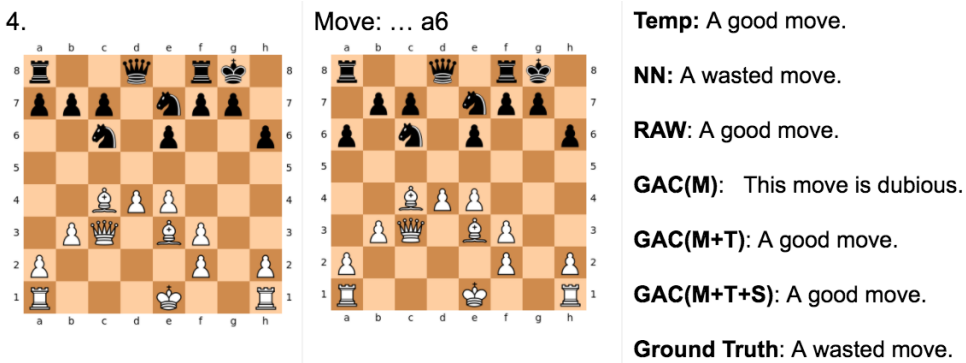


Figure 12: Example output 4: Move quality subset of data.

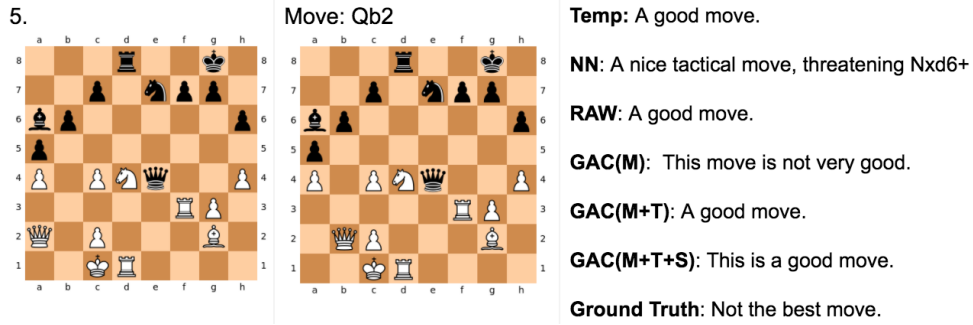


Figure 13: Example output 5: Move quality subset of data.

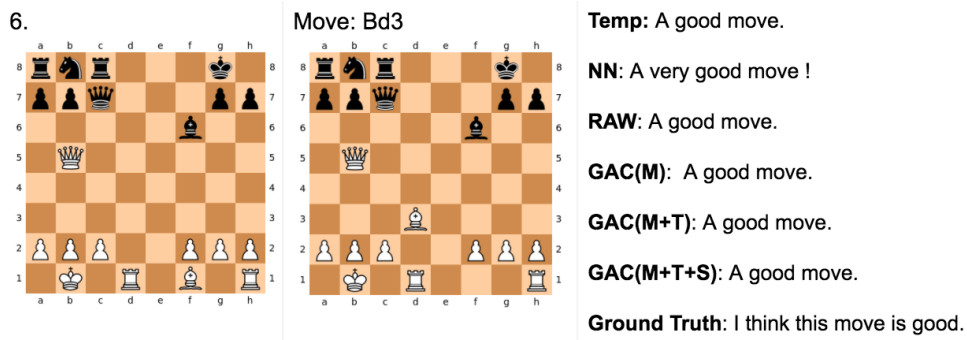


Figure 14: Example output 6: Move quality subset of data.

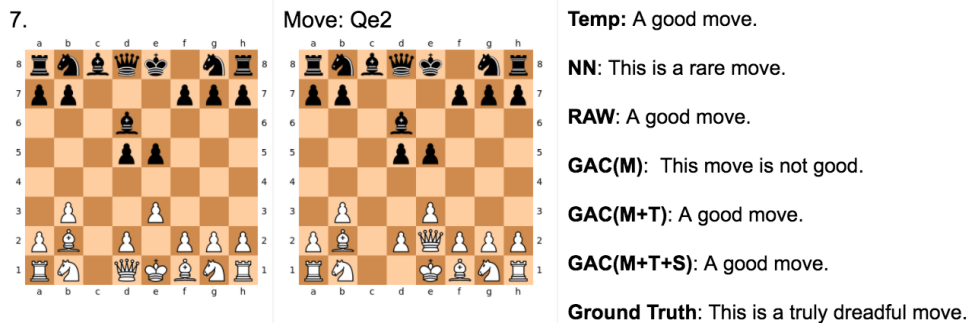
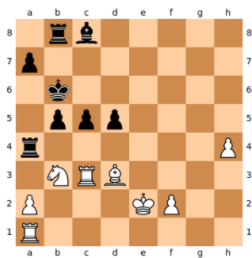


Figure 15: Example output 7: Move quality subset of data.

1.



Move: Bg4+



NN: knight c6 would have been much better for obvious reasons .

RAW: It would have been better to play the pawn.

GAC(M): better was to develop the knight.

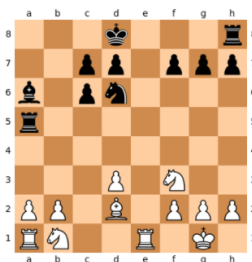
GAC(M+T): it would have been better to play f5.

GAC(M+T+S): it would be better to play the pawn , but i think it would be a little better .

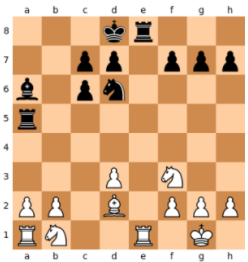
Ground Truth: black would have been much better off doing rxh4 , because eventually , that pawn does become a queen ..

Figure 16: Example output 1: Comparative subset of data.

2.



Move: Re8



NN: what better way to defend than to attack.

RAW: better would have been to retreat the knight.

GAC(M): that would have been better.

GAC(M+T): it would have been better to play the rook.

GAC(M+T+S): it would be better to play the knight to move to the king , but i think it would be better to play the knight to move to the king side.

Ground Truth: as better would 've been to retreat her ra5.

Figure 17: Example output 2: Comparative subset of data.

3.



Move: Rc8



NN: i think c3 was better.

RAW: better would have been to retreat the bishop to retreat.

GAC(M): that would have been better.

GAC(M+T): it would have been better to do the bishop to the king side.

GAC(M+T+S): it would be better to play the knight to move to the knight , but i wanted to play the knight to move to the king side of the board.

Ground Truth: understandable , but 5 bc4 might be better.

Figure 18: Example output 3: Comparative subset of data.

**Appendix D: Additional information on
AMT experiment**

Instructions (Click to collapse)

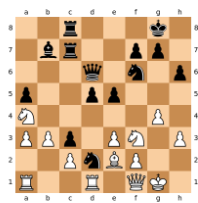
This task requires basic knowledge of the game of chess. Please participate only if you have decent knowledge about chess.

Our first two questions check some basic knowledge of chess game.

Thereafter, for the remaining questions, you will be shown a chess move through the previous board and the resulting board, along with information on which piece moved. With this context, you will be shown a text commentary on the game. You have judge the commentary on :

- 1) Correctness: Is the text commentary a valid commentary for the chess board
- 2) Completeness: Does the commentary correctly describe the chess move which occurred. In other words, given the commentary and the *previous* board, would you be able to figure out the move which was taken?
- 3) English language Fluency: Is the commentary in fluent English?

Proficiency question 1.

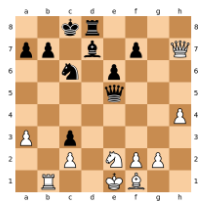


Current board state

Q. Is the game over with a checkmate?

Yes No

Proficiency question 2.



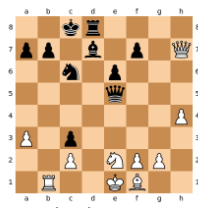
Current board state

Q. Does black knight attacks d4?

Yes No

Figure 19: AMT (Amazon Mechanical Turk) sample HIT (Human Intelligence Task): Part 1 of 2 : Two chess proficiency questions are asked at beginning of a HIT

Proficiency question 2.



Current board state

Q. Does black knight attacks d4?

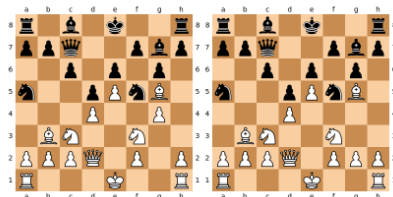
Yes No

You have judge the commentary on :

- 1) Correctness: Is the text commentary a valid commentary for the chess board
- 2) Completeness: Does the commentary correctly describe the chess move which occurred. In other words, given the commentary and the *previous* board, would you be able to figure out the move which was taken?
- 3) English language Fluency: Is the commentary in fluent English?

1. Commentary text: Back to standing in front of the king !

Which piece was moved: white pawn g2



Previous board state

Current board state

1.1 Is commentary correct for the shown chess move?

Yes No

1.2 Can you infer what the move is from commentary given only the previous board state?

Yes No

1.3 On a scale of 1-5, with 5 being most fluent, rate the English fluency of the commentary text

1 2 3 4 5

2. Commentary text: \${c2}

Which piece was moved: \${m2}

Figure 20: AMT (Amazon Mechanical Turk) sample HIT (Human Intelligence Task): Part 2 of 2: 7 sets of questions are asked to judge quality of generated text. Each of the seven texts is output from a different method.

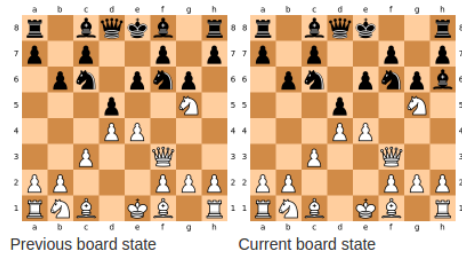


Figure 21: Commentary text: *I develop my bishop to the queen .*

An example instance where output commentary from our method was marked as not valid for the given chess move

Checking chess proficiency of annotators

Our proficiency test questions are chosen from a subset of questions by (Cirik et al., 2015). Each question consists of a chess board and a question about the board configuration or game situation. The paper proposes a range of question types such as enumerating pieces of a type, enumerating pieces of a player, whether one piece threatens another, and whether the configuration corresponds to a checkmate or stalemate. For simplicity we stick to only those question types that have binary answer response.

We classify the question types into **Easy** and **Hard** question types. Each annotator is presented with one **Easy** and one **Hard** question at the start of a HIT.

References

Volkan Cirik, Louis-Philippe Morency, and Eduard Hovy. 2015. Chess q&a: Question Answering on Chess Games. In *Reasoning, Attention, Memory (RAM) Workshop, Neural Information Processing Systems*.