Appendix A: Additional Data Examples
Unpins and defends the knight, but it does n’t matter, as the time is ripe .

He gets fed up and exchanges Queen for Rook .

Rxc3 , I just retake with my queen , whilst if he attempts defense with the bishop , then after 17.Bd2 , Ne4 , 18.Rxc3 , Nxg3 , 19.Rxc6 , Nxh1 , I ’ve won a rook outright .

Preparing to castle , and threatening now white’s e pawn for real .

Simply getting my rook off that dangerous diagonal, and protecting the b pawn .

I throw in a check

Threatening mate with Qxh2

A punch drunk move !

This is not the best move .

The most logical move .

This move is dubious .

The check gains time to support the advance of the a-pawn maybe Ke1 was better

I did n’t want to retreat the N and I rejected 11.

I wish to both defend the pawn, and threaten indirectly the black queen , gaining a tempo maybe Nd2 was better

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 .

but better would be Nd2 to get the knight in the game, the queen rook, too .

i think it would have been better to play nxe5 and maintain a material advantage .

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 .

fianchettoing , so that when black does complete his development, his b will be on a better diagnol .

He doesn’t notice that his Knight is hanging ...

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

Well, now the game will get interesting soon

He tries his trick , which of course is noticed

This is often what I will do , when I ’m playing white.

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 the following templates:

  1. **Capture** moves: \([\text{PLAYERMOVED}]\) captures the \([\text{CAPTUREDPiece}]\) at \([\text{FINALSquare}]\) using the \([\text{PIECEMOVED}]\) at \([\text{INITIALSquare}]\).

  2. **Non-Capture** moves: \([\text{PLAYERMOVED}]\) moves the \([\text{PIECEMOVED}]\) from \([\text{INITIALSquare}]\) to \([\text{FINALSquare}]\).

  3. **Castling** moves: \([\text{PLAYERMOVED}]\) 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. \([\text{PLAYERMOVED}]\) (Black/White), \([\text{INITIALSquare}]\), \([\text{FINALSquare}]\), \([\text{PIECEMOVED}]\) 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 \(\theta\) 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.
Figure 5: Example output 5: 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 to 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 6: Example output 6: 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 rook, which would free up black to take the e5 pawn and even out material.

Figure 7: Example output 7: 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 8: Example output 8: Move description subset of data.

8. Move notation: Bh6

- Temp: White moves the bishop from g5 to h6
- NN: Bxg8 Qxg8
- 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 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.
Figure 16: Example output 1: Comparative subset of data.

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

Figure 18: Example output 3: Comparative subset of data.
Appendix D: Additional information on AMT experiment
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.
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**