Knowledge-aware ELECTRA: (1) conduct an entity matching for each question, concatenate the domain label to the end of history. (2) add a bit-knowledge flag to the end of the final hidden vector of the token `<bos>` to classify.

Knowledge-seeking Turn Detection

(1) Retrieve and Rank Model: 
Retrieve candidate knowledge base entities using text-matching based heuristics.
(2) Three-step Model: 
Predict domain, entity and document of the desired knowledge snippets with separate Roberta models.
(3) Ensemble Model: 
Ensemble the Retrieve & Rank model with Three-step model together.

Model architecture. The trapezoidal mask is to prevent response information leakage from our bi-directional encoder.

(1) Latent variable: Capture the response information, work like shortcut connection, optimize under KLD
(2) Knowledge Copy: Combine auto-regressive generation with copy mechanism by weight adding.
(3) Segmented Response Generation(SRG): Generate knowledge part and greedy part separately
(4) Modified Beam Search(FFBS): Fix different first word of response then do normal beam search for each group
(5) Post-processing Strategies: Plus semantic similarity score and minus word similarity score between response candidates from FFBS and knowledge to NLL score.

Knowledge-Teacher: 
We use pre-trained language models, ELECTRA and Roberta, as our base encoder for different subtasks. 
Subtask 1 and 2: coarse-grained information like domain and entity are used. 
Subtask 3: latent variable at encoder and generate responses combined with copy mechanism.

Our proposed system ranks second under objective metrics and ranks fourth under human metrics.