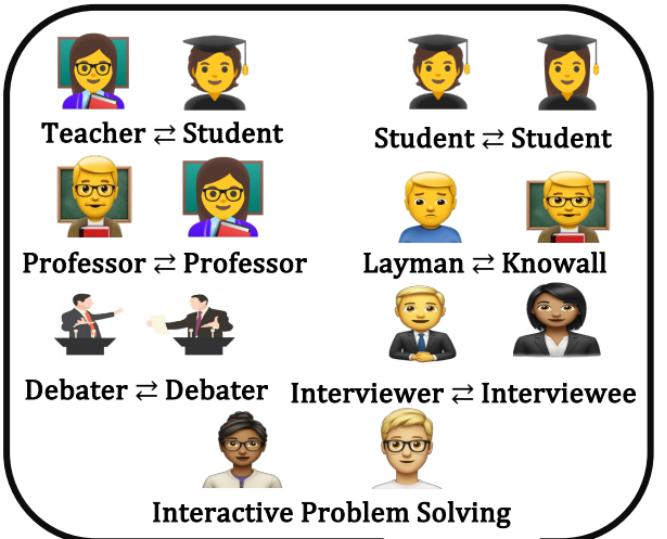
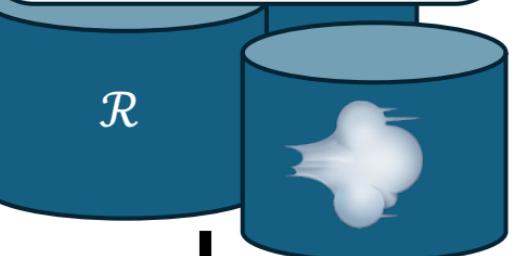


Prompts, $\mathcal{P} \in [1, 7]$

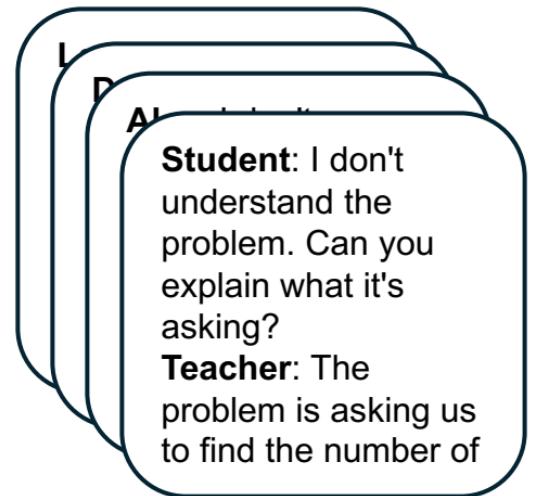


Raw Data, $\mathbf{r}_j \in \mathcal{R}$

Browse Questions #
How many numbers are there between 20000 and 30000 in which the digits are 2,3,5,6,7 and each digit can be repeated any number of times.



Conversations, $\mathbf{s}_{\{i,j\}} \in \mathcal{S}$



Conversation Generator, \mathcal{M}

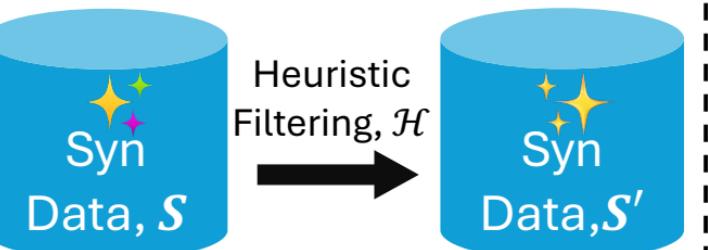
Generate Synthetic Dialogue Corpora using **MIND**
 $\mathcal{R} \rightarrow \mathbf{MIND} \rightarrow \mathcal{S}'$

Language Model, \mathcal{E}



Continuous Pretraining

Language Model, C



Heuristic Filtering, \mathcal{H}

Evaluate **MIND** generated data
 $\mathcal{S}' + \mathcal{D} \rightarrow \mathcal{E}; \mathcal{D} = \{\mathcal{S}' \cup \mathcal{R}_{pt}\}$