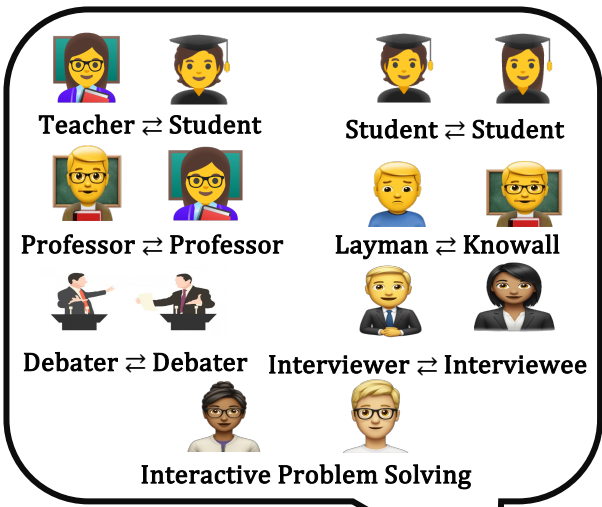
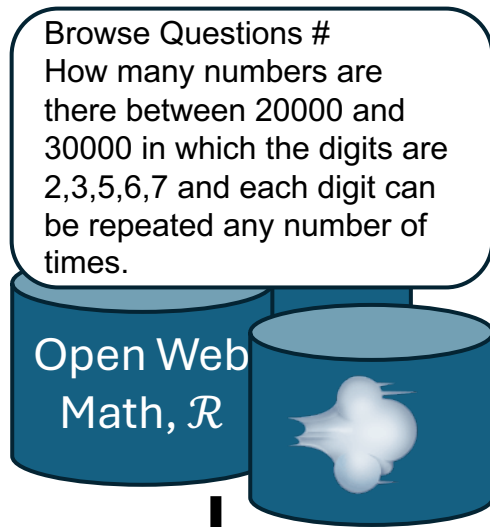


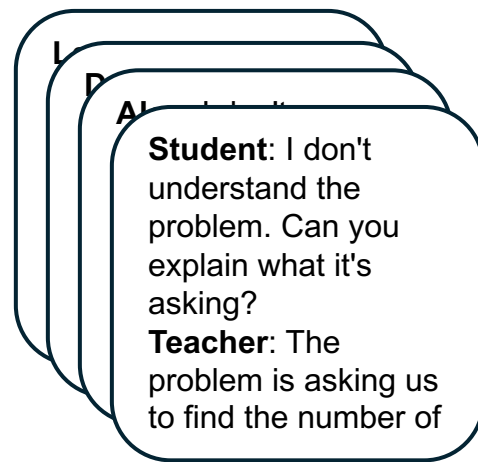
Prompts, $\mathcal{P} \in [1, 7]$



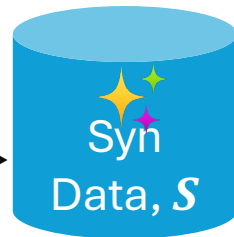
Raw Data, $\mathbf{r}_i \in \mathcal{R}$



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



Conversation Generator, \mathcal{M}



Heuristic Filtering, \mathcal{H}



Continuous Pretraining

Language Model, ε



Language Model, \mathcal{C}

+

