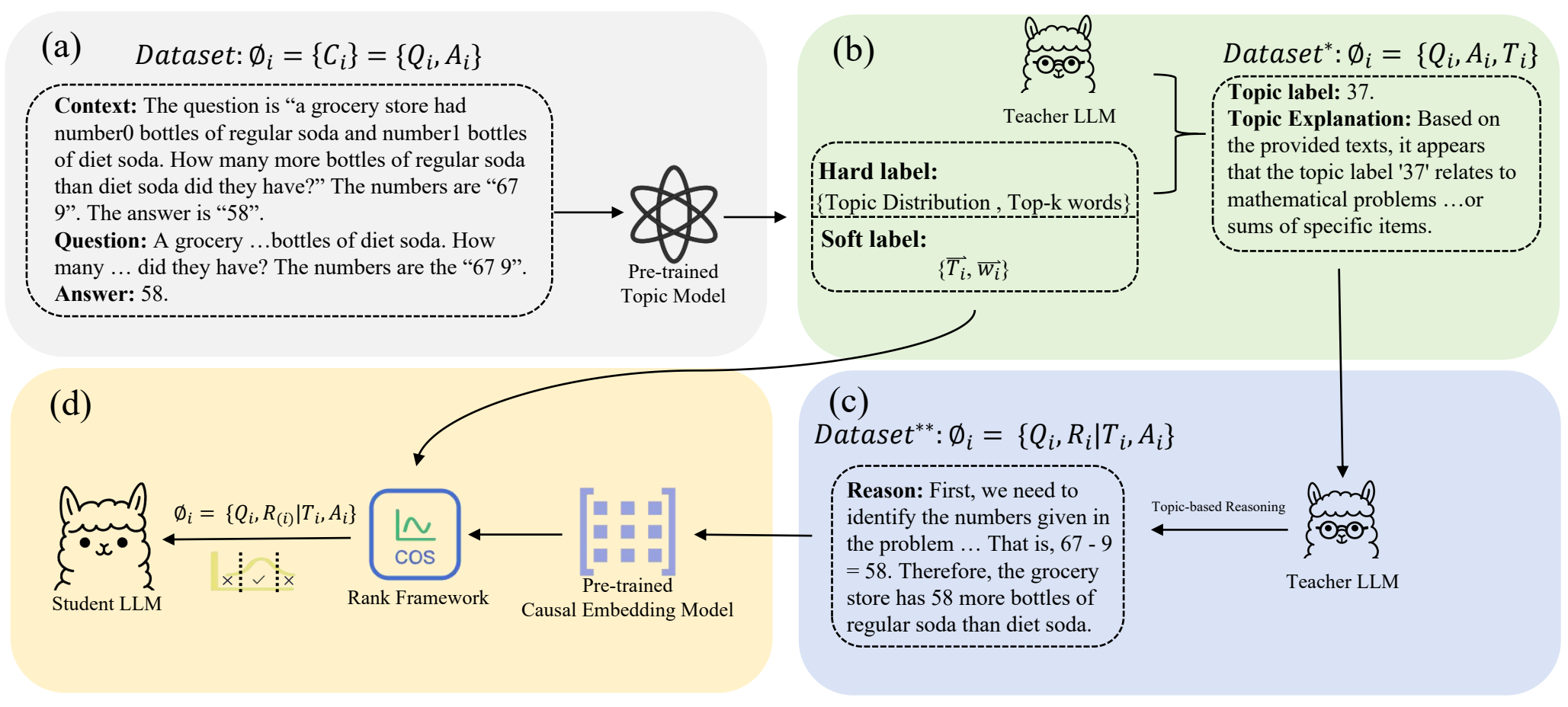
(a)
$Dataset: \emptyset_i = \{C_i\} = \{Q_i, A_i\}$
Context: The question is "a grocery store had number0 bottles of regular soda and number1 bottles of diet soda. How many more bottles of regular soda than diet soda did they have?" The numbers are "67 9". The answer is "58".
Question: A grocery …bottles of diet soda. How many … did they have? The numbers are the "67 9".
Answer: 58.
Pre-trained Topic Model
(b)
Teacher LLM
Hard label:
{Topic Distribution , Top-k words}
Soft label:
$\{\overrightarrow{T_i}, \overrightarrow{w_i}\}$
$Dataset^*: \emptyset_i = \{Q_i, A_i, T_i\}$
Topic label: 37.
Topic Explanation: Based on the provided texts, it appears that the topic label '37' relates to mathematical problems …or sums of specific items.
(c)
$Dataset^{**}: \emptyset_i = \{Q_i, R_i | T_i, A_i\}$
Reason: First, we need to identify the numbers given in the problem … That is, 67 - 9 = 58. Therefore, the grocery store has 58 more bottles of regular soda than diet soda.
Topic-based Reasoning
Teacher LLM
(d)
Student LLM
$\emptyset_i = \{Q_i, R_{(i)} | T_i, A_i\}$
COS
Rank Framework
Pre-trained Causal Embedding Model