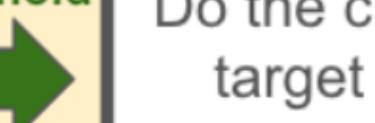**Stage 1: Non-Redundancy**
Is the requested information already present in the input?
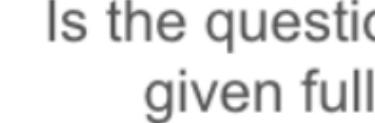$$r_{\text{non-redundancy}} = 1 - \frac{\text{redundant}}{\text{total}}$$

> Threshold

**Stage 2: Diversity**
Do the clarification questions target different pieces of missing information?
$$r_{\text{diversity}} = 1 - \frac{\text{similar}}{\text{total}}$$

> Threshold

**Stage 3: Answerability**
Is the question answerable given full privileged information?
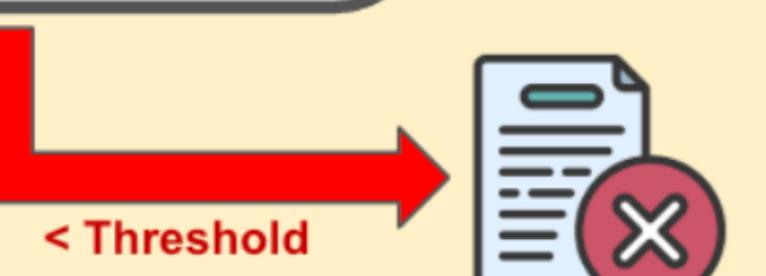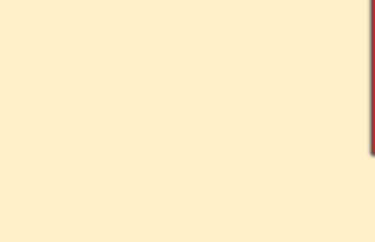$$r_{\text{answerability}} = \frac{\text{answerable}}{\text{total}}$$

> Threshold

**Stage 4: Task Relevance**
Do the questions target information most critical for task success?
$$r_{\text{relevance}} = \frac{1}{N} \sum w(\text{category}_i)$$

< Threshold

< Threshold

< Threshold

**Filtered generation**
Final = Score1 + 0 + 0 + 0

**Filtered generation**
Final = Score1 + Score2 + 0 + 0

**Filtered generation**
Final = Score1 + Score2 + Score3 + 0