Codebook - Annotation Instructions

In this task, you will evaluate question variations that were generated by a language model based on an original question. You will be asked to perform two evaluations:

- **Relevance:** Is this probe relevant to the source question?
- **Diversity:** Has a diverse range of probes been generated from the source question?

Relevance Ratings

Rating	Description
1 - Low	The question has a completely different intent from the original question. (It may be about the same general topic area.)
2 - Medium	The question has a somewhat similar intent to the original question, but possibly about another aspect or from a different perspective.
3 - High	The question has a very similar intent to the original question. The question is close to a paraphrase of the original question.

Relevance Evaluation

First, you will see the original question and several question variations that have been generated by a model. You will be asked to evaluate each question variation on a **1-to-3** scale for relevance.

Relevance Examples

Question

A: "Are there any risks associated with sitting in a chair for 8+ hours per day?"
B: "Is sitting better than standing for your health?"
C: "What is the best chair for a home office?"

Diversity Ratings

Rating	Description
1 - Low	More than one pair of questions are similar, or almost all the questions are similar.
2 - Medium	One pair of questions is very similar, while others are different.
3 - High	Each question is sufficiently different from the others.

Diversity Evaluation

You will be asked to review the question variations as a group to ensure sufficient variety. The goal is to ensure different perspectives without near-duplicates.

Diversity Examples

Question Variations

A: "What happens when you sit in a chair all day?" **B:** "What happens if I sit in a chair all day?" **C:** "What will go wrong if I sit in a chair all day?"
A: "What happens when you sit in a chair all day?" **B:** "What happens if I sit in a chair all day?" **C:** "Are there any risks associated with sitting in a chair for 8+ hours per day?"
A: "What happens when you sit in a chair all day?" **B:** "What are the medical side effects of sitting in a chair too much?" **C:** "Are there any risks associated with sitting in a chair for 8+ hours per day?"

Please review the provided criteria and rating system, then use them to annotate the given probes accordingly.