**Participant Guide: Generating One Positive and One Negative Sample**

**Overview:**

You will be asked to complete two tasks. In Task 1, you will receive one abstract question randomly selected from a predefined question bank. Your goal is to write a **specific version** of this question. In Task 2, you will receive the full list of abstract questions, and your goal is to write a question that is **unrelated to any of them**. Each participant will submit **one positive sample** and **one negative sample**.

### ✅ Task 1: Generate One Positive Sample (Matching Question)

You will be randomly assigned **one abstract question** from the question bank. Based on this abstract question, you should design **a more concrete and specific version** of the same question.

You are encouraged to use, but are not limited to, the following strategies to make your question more realistic or varied while keeping it aligned with the original meaning:

* **Incorporate dataset-specific details**  
  e.g., replace general terms like “independent variable” with specific features like “house size” or “income.”
* **Use synonym substitutions**  
  e.g., replace “impact” with “influence” or “effect.”
* **Rephrase or paraphrase the sentence**  
  e.g., “What happens to the weight of the fish as its length increases” instead of“How does the dependent variable change with an increase in the independent variable(s)?”
* **Include minor typographical errors**  
  e.g., “affct” instead of “affect,” “incraesing” instead of “increasing.”
* **Introduce small grammatical deviations**  
  e.g., “How age affects price?” instead of “How does age affect price?”

Your final question should retain the original **intent and structure**, but appear more natural or varied as a real-world user question.

### ❌ Task 2: Generate One Negative Sample (Non-Matching Question)

You will also be shown a **abstract question bank**. Your task is to design a question that is **not related to any** of the abstract questions in meaning, even if it sounds similar on the surface.

You may use, but are not limited to, the following strategies to create an unrelated question that still appears plausible:

* **Ask about data processing or formatting**, not analysis  
  e.g., “What columns are missing from the dataset?”
* **Focus on visualization or summary statistics**  
  e.g., “Can we plot a histogram of ages?”
* **Ask about model configuration or setup**  
  e.g., “Which loss function is used in this model?”
* **Avoid asking about model accuracy, variable influence, or overfitting**  
  e.g., “Is dropout used in the model architecture?”
* **Use surface-level similarities to make it challenging**  
  e.g., “Which features are used in the model?” (sounds close, but not asking about impact or importance)

Your negative sample should **not be answerable using any abstract question from the list**, even though it may look similar at first glance.