Task-2 Guidelines

▼ Instruction:

To assess a peer review and check how accurately it reflects the paper's content.

Evaluation Process:

1. **Read the entire paper first**.

   a. It is recommended to download the paper from the portal **only**, for the purpose of highlighting important points or for easier reading.

2. Read the review. Identify every **discrete claim**.

   a. In case of "Summary" section of a review, the **entire content** of the "Summary" section should be identified as a **discrete claim.**

   b. In case of "Strength", "Weaknesses" and "Questions" sections of the review, **each of the bulleted points** should be identified as **discrete claims**.

3. For each claim, categorize it using the labels below.

4. If you are unfamiliar with a particular concept in the paper content/review content, you can search it up or consult GPT for further explanations to it if required.

Claim Categorization:

- **Valid (V):**

  - The claim is factually correct and aligns with the paper's content.

  - The review accurately identifies a strength, weakness, or limitation that is consistent with what the paper presents.

  - **Example:** *The review mentions that a particular method is not explained well, and indeed it is not explained through the paper content.*

- **Factual Error (FE):**

- The claim is factually incorrect based on the paper's content.

- This occurs when the reviewer states something that contradicts what is clearly presented in the paper.

- **Example:** *The reviewer claims "The method is not explained," but the method is fully described in Section 2.1.*

- **Redundant Critique (RC):**

  - If the paper presents a fair limitation (e.g., *"no human evaluation was done due to lack of resources"*), and the review still criticizes it as *"The authors failed to do human evaluation"*, this should be labeled RC only if the annotator deems the explanation provided by the authors fit. Otherwise, this should be labelled as V.

- **Irrelevant (IR):**

  - The claim is an opinion or suggestion that is not relevant to the paper's content (Hallucinations)

  - **Example**: The paper is about POS Tagging and review talks about NER.