Questions for Authors

- How did you convert numerical quantities into textual representations during Data Preparation stage? An example would be helpful.

We have used rule-based functions along with LLM as ChatGPT, Mistral 7b to convert and explain numerical quantities.
The process goes as follows: first numbers, operations and abbreviations are converted into textual representation. For example *“Arm/Group Description: Phase 1 Cyclophosphamide Dose Escalation: Fludarabine 5x25mg/m2 + Cyclophosphamide 2x60mg/kg + Cells intravenous (IV) "* is converted to *“Arm/Group Description: Phase one Cyclophosphamide Dose Escalation: Fludarabine five times twenty-five milligram/ meters two plus Cyclophosphamide two times sixty milligram/ kilograms plus Cells intravenous (IV)."* The next steps involve using LLM as chatGPT or Mistral 7b to give an explanation for the converted text. The following prompt is feed to the LLM *“ Explain the following sentences in five sentences: text".* The following explanation is provided by ChatGPT: *“ The given sentence describes a specific group or category within a clinical trial referred to as "Arm/Group Description" during the phase one stage. This particular arm involves a treatment protocol called "Cyclophosphamide Dose Escalation," which aims to determine the optimal and safe dosage of Cyclophosphamide, a chemotherapy drug. The treatment regimen includes the intravenous administration (IV) of Fludarabine at a dose of twenty-five milligrams per square meter of body surface area, repeated five times over a specific time period or schedule. Additionally, Cyclophosphamide is administered intravenously (IV) at a dose of sixty milligrams per kilogram of body weight, repeated twice over the same time period or schedule. The treatment also involves the administration of cells intravenously (IV), although the specific type of cells is not mentioned in the provided sentence."*. Finally, Both Converted text and explanation are concatenated.

- Does the first and second last entry in the Table 1 means that both model achieved exact same results or is this an ensemble?

They are ensembles of different models, however, they could achieve the same results.

- What is the final combination of model architectures, loss functions and prompts that achieved the highest score ?

The final model was an ensemble of different methodologies mainly an ensemble of:

- DeBERTa
    - Architecture: Mean Pooling
    - Loss function: Cross Entropy
    - Prompt: Comparison type [SEP] token_special premise [SEP] premise”
- DeBERTa
    - Architecture: Mean Pooling
    - Loss function: Cross-Entropy and Contrastive Learning
    - Prompt: “premise [SEP] hypothesis”.
- DeBERTa
    - Architecture: GeM Pooling
    - Loss function: Cross-Entropy
    - Data preparation: Converted numeric values and abbreviation
    - Two separate models for each comparison type
    - Prompt: “ premise [SEP] Is this statement correct based on previous CTR reports: hypothesis? ”.
- DeBERTa
    - Architecture: GeM Pooling
    - Loss function: Cross-Entropy
    - Data preparation: Converted numeric values and abbreviation
    - Prompt: “ premise [SEP] hypothesis”

The following weights were used : [0.25 0.25 0.25 0.25]

- What is GeM pooling in "Model Architecture"?

Generalized mean (GeM) pooling: pooling techniques that compute the generalized mean of each channel in a tensor (mainly the mean of the CLS vector of the LM model).
Uses the following equation:

$$\mathbf{e} = \left[\left(\frac{1}{|\Omega|} \sum_{u \in \Omega} x_{cu}^{p} \right)^{\frac{1}{p}} \right]_{c=1,\cdots,C}$$

From: https://paperswithcode.com/method/generalized-mean-pooling

- Did you try all prompts with all models? And if not, why not? How did you choose the final prompts?

No
We have tried all of the prompts on DeBERTa, T5, and Sci-five models.

We have chosen prompt which improves the model performance on f1-score of the dev-set

---