**Motivation & Objectives**

As far as we humans are concerned, in order to achieve good performance on reading comprehension tasks, reading through a given passage and understanding what it says are just the preliminary requirements. We can see that the texts in the reading comprehension tests from the elementary school level to the postgraduate examination level gradually changed from simple to complex, and the question types also changed from simple text extraction to complex logical reasoning. And in this process, we humans have developed some general reading comprehension abilities through continuous learning and training. As a result, it is natural to ask that whether a computer, or specifically an AI system, can also acquire this combination of natural language understanding and logical reasoning.

With recent advances in deep learning techniques, it seems possible to achieve human-level performance in certain language understanding tasks, and a surge of effort has been devoted to the machine comprehension task where people aim to construct a system with the ability to answer questions related to a document that it has to comprehend. Just like the human language test, the most common way to test whether an intelligent agent (a person or an AI system) can fully understand a piece of text is to require her/him/it to answer questions about the text. This is called Question Answering (QA).

For the simple QA in which the candidate answers are directly extracted from the context, many models can achieve excellent performance. And for complex QA which requires logic reasoning ability, many dataset have been developed in recent years and many researchers claimed that they have developed models with the logic reasoning ability.

However, I suspect that these AI models, while performing well on relevant complex datasets, do not actually understand the texts and do not acquire general logical reasoning capabilities at all. They just learn different implicit biases and features contained in different datasets, and use this information to obtain answers on test sets with the same data distribution. More specifically, suppose we train and test a model on a dataset based on graduate reading tests which require the ability to do complex reasoning and to understand more complex texts. If the model performs well and we assume that it has real reasoning ability, then the model should perform equally or better on lower-level reading tests. In other words, a person who can achieve good grades in reading comprehension at the graduate level is unlikely to perform worse in reading comprehension at the junior high school level.

So in this project, I want to first (1) conduct experiments to see whether the current state-of-the-art pre-trained models which are trained and behave well on complex QA dataset have learned general reading and reasoning skills and if possible (2) to seek that whether I can improve the model performance by adding some new variant.

**Methodology**

**NLP tasks**

In this project, I

**Dataset**

ReClor.

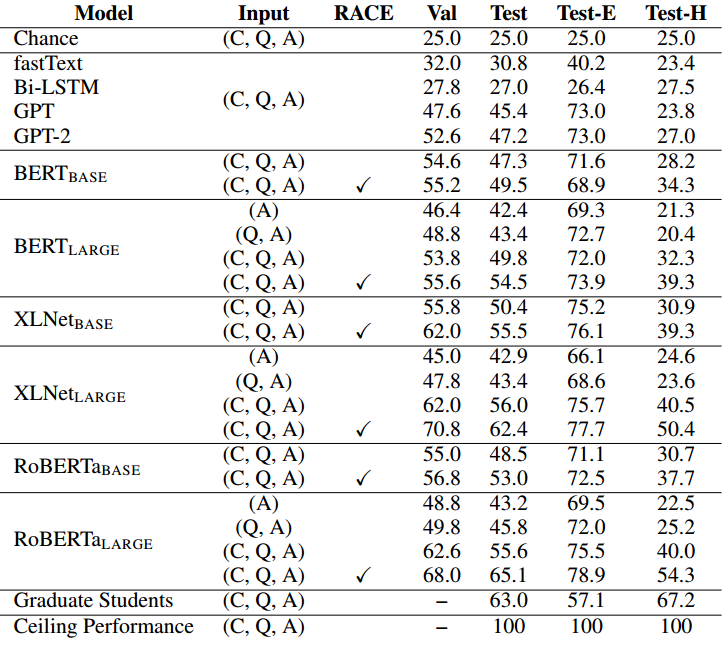
**Neural Models**

**Evaluation**

**Baselines**

**quantitative analysis**

**qualitative analysis**



**Evaluation**

Use accuracy to evaluate implemented model. Comparing our model with baseline model.

Expectation: increase acc, especiallly on Hard Set.

Qualitative evaluation: depend on results of our model. It can not be discussed now.