

## Analysis - Social Interaction Question Answering (Social IQA)

- SocialIQA Dataset consists of positive and negative answers. Negative answers are the options which are generated from the same context but they are correct answers for a different question. These cannot be handled by the regular models as there is a lot of similarity between all the answer choices and regular models cannot differentiate between them. So, we need additional knowledge to handle such problems.
- BERT, RoBERTa and ELMo are different pre-trained models. They are trained using a large set of knowledge bases. Only using pre-trained models for the Social IQA dataset may result in many misclassification errors. The main misclassifications that may occur are when
  - The questions are drawn out of the context.

**Context:** Bailey sorrowfully confessed to cheating on Jan.

**Question:** Why did Jan do this?

**Options:** (a) wanted to humiliate Bailey (b) wanted to make amends (c) wanted to apologize to Jan

**Answer:** wanted to humiliate Bailey

**Analysis:** The context only tells about the action done by Bailey. But the question asked about the action done by Jan.

- The incorrect choices are more similar to the context than the correct choice.

**Context:** Carson liked Cameron enough to ask them to play video games with them.

**Question:** How would Carson feel afterwards?

**Options:** (a) Like they want to play games (b) a good person (c) a friendly person

**Answer:** a friendly person

**Analysis:** Context tends to favor option a according to the similarity than option c which is the correct answer.

The following may also cause the misclassification error in BERT:

- The choices are not similar with the context.
- The context has a multiple occurrence of a word with multiple meanings (problem due to masking in BERT).

### Setbacks Of BERT:

- BERT's bidirectional approach (MLM) converges slower than left-to-right approaches (because only 15% of words are predicted in each batch) but bidirectional training still outperforms left-to-right training after a small number of pre-training steps.

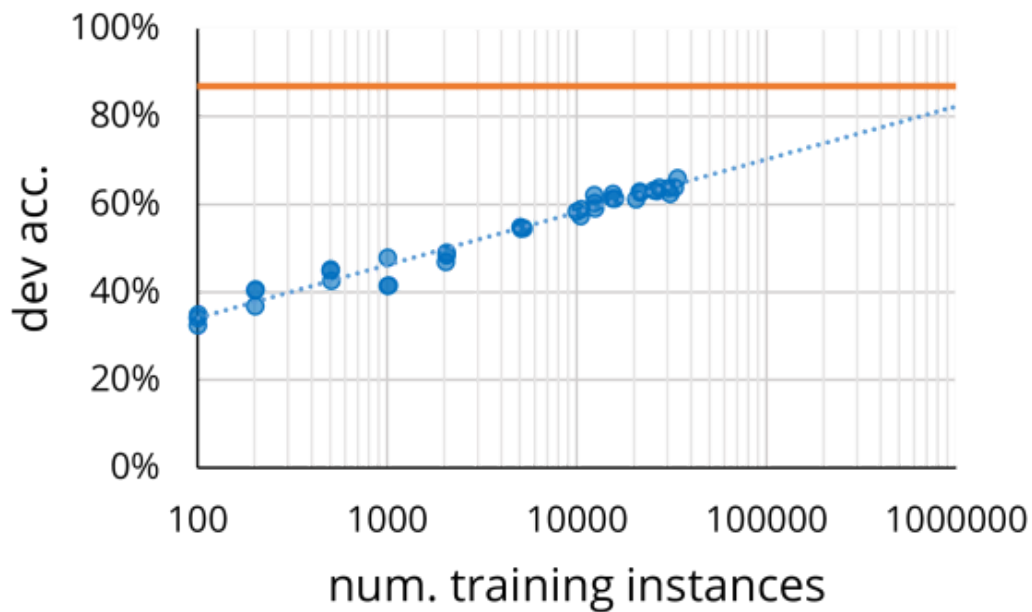
### Advantage Of BERT:

- BERT does a faster Fine-tuning with the dataset.

### Bert analysis curve:

We can improve the performance of the BERT by increasing the number of training examples

Orange line is human intelligence



To decrease the misclassification error caused by BERT, RoBERTa can be used. As RoBERTa iterates on BERT's pretraining procedure, including training the model longer, with bigger batches over more data; removing the next sentence prediction objective; training on longer sequences; and dynamically changing the masking pattern applied to the training data.