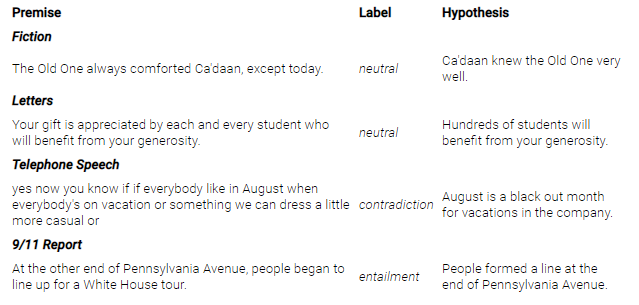
BERT (Bidirectional Encoder Representations from Transformers) is a self-supervised method release by Google in 2018. It relies on unannotated text obtained from the web, as opposed to a language corpus that’s been labeled specifically for a given task.

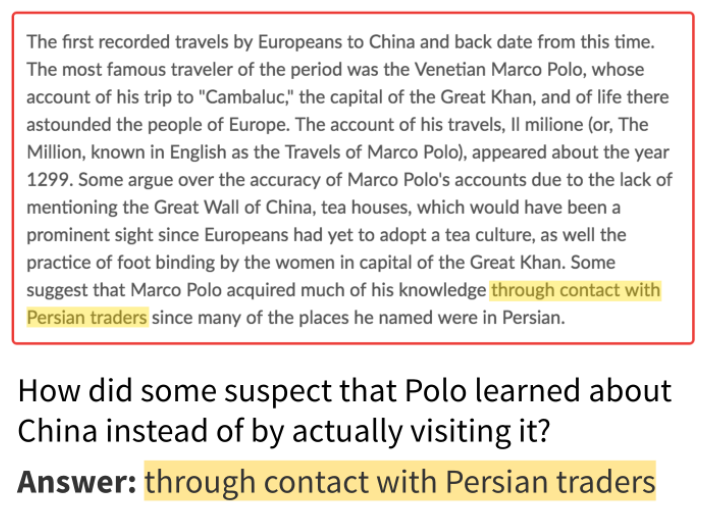
Roberta:

* follows the same strategy as BERT’s language masking, where the system learns to **predict intentionally hidden sections of text** within unlabeled language examples.
* Modifies key hyperparameters in BERT as:
  + Removing BERT’s next-sentence pretraining objective
  + Training with much larger mini-batches and learning rates
* Is trained on a larger dataset than BERT and for a larger amount of time. The dataset is conformed by existing unannotated NLP datasets as well as CC-News, a novel set drawn from public news articles.
* Roberta delivered state-of-the-art performance on the
  + MNLI (Multi-Genre Natural Language Inference corpus is a crowd-sourced collection of 433k sentence pairs annotated with textual entailment information). This data is from different sources: Face-to-face, Telephone, 9/11, Travel, Letters, Oxford University Press, Slate, Verbatim, Government and Fiction



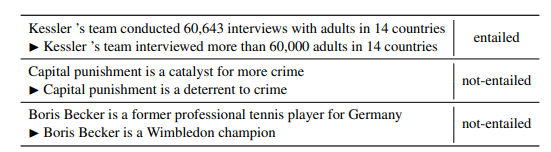


* + QNLI (Question-answering NLI), which is a dataset with question-paragraphs pairs, The task consists in determining if the paragraph contains the answer to the question, with the text being from Wikipedia and the questions being written by an annotator.



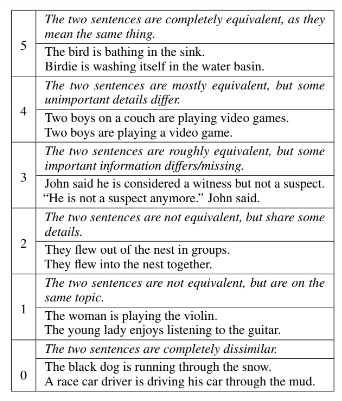


* RTE (Recognizing Textual Entailment) dataset contains pairs of text fragments, with the task requiring go recognize whether the meaning of one text is entailed from the other. Data is from news and Wikipedia text.



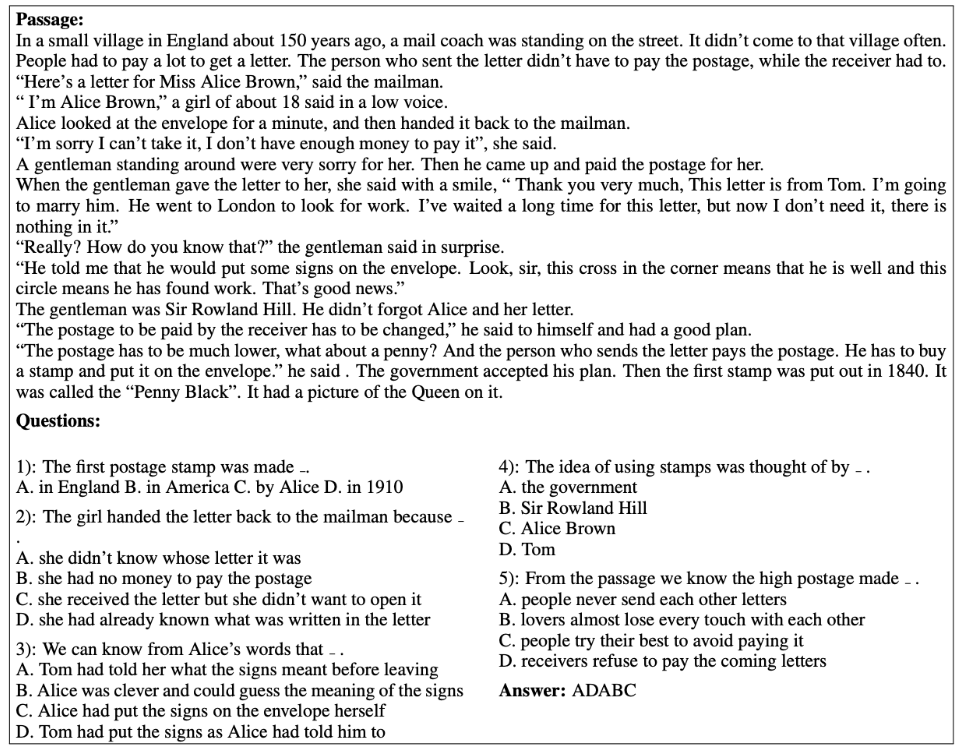


* STS Benchmark (Semantic Textual Similarity), which comprises a selection of datasets formed by image captions, news headlines and user forums.





* RACE (ReAding Comprehension dataset from Examinations) consists of 27,933 passages and 97,867 questions from English exams, targeting Chinese students aged 12-18. Each question has only one correct answer. RACE differs from most machine reading comprehension datasets – instead of generating questions and answers by heuristics or crowd-sourcing, questions are specifically designed for testing human reading skills.





At the time RoBERTa was released in 2019, it reached the top position of the GLUE leaderboard.

Glue (General Language Understanding Evaluation) is a collection of nine natural language understanding tasks that aim to evaluate natural language understanding models.

At the actual moment Roberta is on the 19th position, even though some models that use it combined with other systems occupy higher positions.

