

NorTruthfulQA: Multiple Choice and Generation

TruthfulQA aims to measure how models mimic human falsehoods, i.e. it contains questions that humans might typically answer untruthfully due to beliefs or misconceptions. This task is very similar to previous annotation tasks that you have been given; so please follow our previous setup.

The dataset contains two subsets: TruthfulQA Multiple Choice and TruthfulQA Generation. The same labels apply as for earlier translations, and you are free to make your own inspired (I) sentences this time as well, adding them to the end of the document like last time. Some of the questions can be very specific to the US, including for example discussion of US laws. If you cannot simply localize, but have to change the entire question, you can skip it (U) or write an inspired sentence.

TruthfulQA Multiple Choice

In this dataset setup, there is no “correct answer”-column, as the correct answer is always A. There can potentially be many options, and so the label column is moved towards the left.

TruthfulQA Generation

The setup here is a bit different. Compared to the TruthfulQA Multiple Choice questions, these files have three additional columns: category, best answer and source. You should not translate the categories. The source column indicates correct and incorrect options. These are used to compare model outputs against. These questions also have certain categories which do NOT need to be translated, and they also have a “best answer” column, which is always one of the “correct answer” options. You should first translate the correct answers, and then copy and paste the corresponding correct one into the “best answer” column. When writing your own questions of this type, it can be difficult to decide on what is the best answer, but just select the one you feel is the most appropriate.

You can read more about the original dataset here:

<https://arxiv.org/abs/2109.07958>