**QUESTION 1**

It can be challenging to write guidelines for annotators so that they get sufficient guidance on what they are supposed to do, without the instructions being too long.  Was there a text where you felt uncertain about exactly how you were supposed to label it based on the guidelines?  Explain what was challenging about labeling the text, and suggest how the instructions could have been improved in order to make the task clearer.

There were many texts where I was uncertain about how exactly to label it based on guidelines. Some of the scenarios that caused confusion:

1. Foreign languages. The guidelines did not specify that only English texts suppose to be classified. I assumed it was English only because I read a paper about Amazon Mechanical Turkey. But if I did not, I would not know what to do if the headline was not in English but in other language that I understand (Russian for example). I would suggest to make a clear statement that the English texts only are being classified.

2. There were multiple texts where they sounded like a pro-mitigation but in reality they were not addressing COVID risks. There were rather promoting strategies for saving struggling businesses. They were not anti-mitigation but the measures offered were not safe or making covid to go away any faster. For example promoting re-opening salons with the use of special guidelines or re-opening certain entertainments involving large groups of people under restrictions, etc. I would suggest to make an example and recommend to classify them as "unclear".

3. Opportunistic statements. This is when some benefits (otherwise unjustifiable) are being promoted as a mitigation measure while they do not help in mitigating actual covid risks in any way. For example, petitions to switch from letter grades to pass/fail grades in universities due to covid or making online tests shorter then normal. I would suggest to give an example of such category and classify it as "unclear".

4. Texts that use some names or geographical locations but unknown to the reader. My assumption was to classify texts based on the knowledge that I have on the spot. If I can’t answer because I don’t know who the person is or where the location is, then I would answer “unclear”. But I could have started using google or Wikipedia to get a better understanding. I would suggest to make a clarification in the guidelines that the classification should be done using current knowledge only.

### QUESTION 2

Very often, annotators used to label data for machine learning models (in NLP and other fields) are paid at a very low rate, based on the total number of judgements they provide.  For example, in a task like the one you are working on, annotators might receive $0.10 - $0.20 per task.

A. What risks does this sort of compensation model pose in terms of ensuring that we have a reliable data set to work with?  Are there strategies that annotators could use in order to increase their pay, at the expense of high-quality judgements?

B. Suggest one way in which this compensation structure could be changed so that annotators are paid fairly for their time, but researchers have greater confidence in the quality of data provided.

1. The risk is that annotators would not read and would not employ thinking judgement. The “strategies” that could be employed by these high risk annotators are: 1) clicking through randomly or 2) use automation scripts to get through as many tasks as possible without consideration for quality.
2. Embed items in the tasks with known labels (i.e. texts that are very difficult to annotate incorrectly, texts with foreign language that must not be labeled, etc.). Calculate the accuracy metric based on the responses for these embedded test items. Give a higher pay rate to the annotators with higher accuracy score, reject the work and withhold the payout for those whose accuracy score is below certain threshold.