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.

**Stop treatment of covid-19 patients in Wolfson Cystic Fibrosis unit - This could either be pro or anti mitigation.**

**Reinstate NCLEX testing – don’t know what NCLEX is, so should I say unclear or should I be searching what it is. For all I know NCLEX would be related to covid 19. What if I don’t know a term or name and it is related to covid 19, I would be getting the label wrong unless I search it. But is searching it the right thing?**

**Extend the governments paid parental leave to pregnant women during the covid 19 pandemic – is this pro mitigation or unclear? This statement could support social distancing because if the parental leave is extended, then pregnant women wont have to go back to work and risk getting/spreading covid 19. This could also be unclear  
- Also similar statement: Suspend ALL rent, mortgage, and utility payments during the COVID-19 pandemic – could be pro mitigation or unclear. By suspending rent and such, there would be more social distancing and thus pro mitigation, because less reason to go out and “work” and be helpful to those with no jobs due to covid.**

**Doula support in Nebraska hospitals during the covid 19 outbreaks – could be pro mitigation or unclear. Supporting hospitals could indirectly be in support of covid vaccinations and covid overall**

**Conduct final examination for students in online mode to prevent risk of covid vaccinations– this is clearly pro mitigation, but if you take out “to prevent risk of covid vaccinations. Then it could be pro mitigation or unclear since online final is beneficial for students, it might be purely cuz its easier for them, or it could be that they want to social distancing.**

**Prisoners sentence to be reduced due to covid 19, full lockdown and no visits – pro mitigation since full lockdown and no visits?**

**Grant work from home to IT employees in india – pro mitigation since support social distancing?**

**Make Colorado public schools remote this fall**

**I try to stick to the characteristics given in the guidelines for pro/anti mitigation and unclear. However, this leads to many unclears that could also be pro/anti mitigation if I just thought about how it could indirectly lead to support and not supporting covid shit.**

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?

**Pay means people might go fast without caring about accuracy since more things they annotate more they earn… Lower the pay, the less they may care about the accuracy and only care about the throughput. To increase their pay, annotators could reassure that they are high accuracy and decrease their pay if they annotated it wrong. Also, they for each annotation they could note their reasons for choosing it.**

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.

**The same group of annotations could be given to multiple people and for each annotation, the true judgement would be based on which judgement was chosen the highest. For example, if 3 annotators chose “positive” and 1 chose “negative”, the label for that specific task would be “positive”. In this case, the payment for that task would only be paid to the 3 annotators that chose “positive”. This will make the compensation be calculated after all the annotations are done. The annotators will then be paid for how many annotations they did that corresponded with the majority of the other annotators.**

A. Do you think the judgements provided by students in this class will be more reliable and consistent than those provided by paid contractors (for example, Amazon Mechanical Turk)?  Why or why not?

B. Do you think your judgements will be more or less reliable than the class average?  Why?

**Answers:**

Yes, there were multiple texts that I was uncertain on how to label it based on the guidelines. Some of examples were "Extend the government paid parental leave to pregnant women during the covid 19 pandemic", "Stop treatment of covid-19 patients in Wolfson Cystic Fibrosis unit", and "Suspend ALL rent, mortgage, and utility payments during the COVID-19 pandemic", etc. What was challenging was determining the context and whether the statement indirectly/implicitly supports covid-19 mitigation efforts. For example, the first example I gave, the result could be that the extension of the paid parental leave would allow for more social distancing since pregnant woman and their husband wouldn't have to possible work. This indirectly is pro mitigation since the result would be most likely more social distancing. However, if you simply look at the statement itself, it doesn't directly imply that its pro or anti mitigation. The context and the indirect result of the statement makes it difficult to 100% pick a label. Most of these similar statements, I chose unclear, but it took me a while to get to that judgement. From what I remember, the instructions only had a few examples and each one of them were simple and directly says they support masks, hate business restrictions, or a statement that had nothing to do with covid-19. While the examples had explanations on why the label was chosen, I feel like the instructions should have included more complex examples as well similar to the statements I provided above. In fact, most of the current examples could have been removed and just replaced with more complex examples so the guidelines would be of similar length while being more complex and helpful.

A. With this compensation, some annotators may value speed rather than accuracy due to the pay being at a very low rate. To increase their pay, they would label data as fast as they can without caring too much about how accurate their labeling is. This would result in unreliable data because the labels might not be as accurate as the researchers want them to be.

B. One way the compensation could be changed is increasing the pay rate a little bit while also taking note of the "accuracy" or "reliability" of the annotator, and possible requiring them to include a short explanation for their labeling. Similar to this homework, one group of task was given to multiple people. Researchers should do the same and after all the annotators are done, their reliability is calculated for each task. If an annotator labeled something as "negative" but a majority of other annotators labeled the same thing as "positive", then the resulting label should be "positive" and the annotator that didn't go with the majority shouldn't be compensated for that specific task. This would make the compensation be dependent on how reliable and how much their labels go along with the majority of other annotators. Now, annotators would focus more on the accuracy of their labeling so that they still get compensated for most if not all of their tasks. This would allow the researchers to have greater confidence in the quality of data provided.

A. I think the judgements provided by students in this class will be a little less reliable and consistent than those provided by paid contractors due to our inexperience. I think a majority of the class has little to no experience labeling, so this would result in less consistency for each student and overall the class. While this labeling task is for our class project, I still think a significant amount of people still don't value this specific task as important and not care as much about the accuracy. Maybe if we were paid for each label we did correctly, then our class would have been more reliable and consistent, probably not that much though. Nevertheless, I feel like the inexperience in labeling is the most significant in making the class judgements being less reliable and consistent than paid contractors, since I assume they have lots of experience with these types of tasks.

B. I think my judgements will be similar or maybe a little less reliable than the class average. At first, I wasn't sure whether to also take note of the indirect impact of the statement or whether it has any implicit connections and label it based on those as well. However, as I labeled more and more, I feel like I became more consistent with how I am judging these statements, and probably more reliable. With regards to how accurate my labeling are, I am not so sure, but I feel like it should be at least close to the class average.