Complete List of Questions:

Geiger’s Questions

|  |
| --- |
| Q1:was the work an original task? |
| Q2: did they use human annotations as labels? |
| Q3:did they use original human annotations? (i.e. annotations they collected themselves?) |
| Q4:did they use external human annotations? (i.e. annotations from an existing dataset?) note: it might that they used both their original annotations as well as external annotations |
| Q5:who were the annotators? (e.g. paper's authors, mturk, experts etc. - i.e. what population were they drawn from?) |
| Q6:did they specify the number of annotators? |
| Q6a: did they estimate how many annotators they would need beforehand? |
| Q7: were there formal instructions for the annotators? |
| Q8: was there training for the annotators? |
| Q9: was there prescreening on the crowdwork platforms? |
| Q10: was there multiple annotator overlap? (i.e. did multiple annotators label the same item?) |
| Q11:reported inter-annotator agreement? |
| Q11a: (or any other metric of label quality?) |
| Q12: did they link to the dataset? note: if they're using all external data, i.e. annotations and raw data, we'd like to track down the original reference and see if information on these questions is available |

My questions