

# Project Proposal



Stephen D. Gardner

---

## Data Labeling Approach

<b>Project Overview and Goal</b>  What is the industry problem you are trying to solve? Why use ML in solving this task?	<p>This classification job will help medical doctors quickly identify cases of pneumonia in children. The dataset contains chest x-ray images of healthy and pneumonia cases that will be used by ML engineers later on down the line to help build a classification product that can help doctors detect pneumonia programmatically across larger datasets.</p>
<b>Choice of Data Labels</b>  What labels did you decide to add to your data? And why did you decide on these labels vs any other option?	<p>The original dataset contains 101 images and only contained two columns (<b>img_url</b>) and (<b>label</b>). The data cleaning was required to create an third image title column (<b>title</b>) to assist further labeling for future jobs, and avoid selecting hyperlinked labels thus degrading the job.</p>

## Test Questions & Quality Assurance

## Number of Test Questions

Considering the size of this dataset, how many test questions did you develop to prepare for launching a data annotation job?

This classification job has 8 test questions given the size of the current data set. During the job each annotator has to complete a verification test after every 5th image classifications in a row. (Ex: 5% test).

## Improving a Test Question

Given the following test question which almost 100% of annotators missed, statistics, what steps might you take to improve or redesign this question?

ID	% CONTESTED	% MISSED	JUDGMENTS	LAST UPDATED	ENABLED
1881190030	<div><div></div></div>	<div><div></div></div>	2	2 days ago	<input checked="" type="checkbox"/>

The following accuracy improvement questions are required during each labeling case where the annotator selects “Pneumonia” to help support dataset labeling accuracy.

- 1.) Is this a chest healthy x-ray image of a patient?
- 2.) Are you 50% confident that the image depicts signs of pneumonia (i.e., at least as likely than not or >50% sure)? (if not select Normal/Unknown)
- 3.) Can you describe your final decision to a medical professional when asked for feedback?

The last question provides an narrative response that can help create future test question iterations classification jobs be modified as additional data is acquired, labeled, and prepared for ML engineers downstream.

## Contributor Satisfaction

Say you've run a test launch and gotten back results from your annotators; the instructions and test questions are rated below 3.5, what areas of your Instruction document would you try to improve (Examples, Test Questions, etc.)

### Contributor Satisfaction ⓘ

Number of participants: 20

**3.2** / 5

Overall

**3.3** / 5

Instructions Clear

**2.9** / 5

Test Questions Fair

**2.8** / 5

Ease Of Job

**3.7** / 5

Pay

Following a test launch where the rating is below 3.5, I would review all the annotator's notes and use them to modify the instruction document for future annotators. I would put more focus on reinforcing the importance of accurately annotating each image and strive to provide future annotators with clear steps to improve image classification.

## Limitations & Improvements

<p><b>Data Source</b></p> <p>Consider the size and source of your data; what biases are built into the data and how might the data be improved?</p>	<p>The annotator may notice that some images are hard to classify between normal and pneumonia symptoms, thus the bias assumes that all images within a given dataset will show healthy or pneumonia x-ray detection regardless if the data has been properly cleaned. If this occurs, the annotator will be asked to select unknown or coordinate with a medical professional to help determine proper classification. This will ensure the most accurate ratings of images for downstream ML experts. Additionally, feedback notes will be available for each image that allow annotators to provide improve future iterations of the x-ray image classification job.</p> <p>During this process the annotator can identify if additional image classifications will be needed in order to accurately determine normal or pneumonia symptoms within images.</p>
<p><b>Designing for Longevity</b></p> <p>How might you improve your data labeling job, test questions, or product in the long-term?</p>	<p>To design for longevity the initial assumption is that all images will be taken from an child. Since, the medical field is always changing it will beneficial to periodical access all annotators feedback notes and to retrain the labeling job on new/dynamic datasets to adjust for bad data (i.e., x-ray images of arms/legs), or unique chest x-ray images (i.e., Unknown; leg/arm x-rays; non-child or low bone density imaging found in older adults).</p> <p>These assumptions will allow ML models downstream within the engineering to better identify outliers in the long-term.</p>