

[◀ Return to Classroom](#)

Create a Medical Image Data Annotation Job

[REVIEW](#)[CODE REVIEW](#)[HISTORY](#)

Meets Specifications

Dear Nourah,

Congratulations on finishing the project 🎉

This was a brilliant submission. The work was exceptional! You did a great job and should be proud of yourself. After reviewing this submission, I am impressed and satisfied with the effort and understanding put in to make this project a success. All the requirements have been met successfully 100 %

Keep doing the great work and all the best for future project.

Files Submitted

The submission zip file includes a complete `Instructions_Preview.html` and `Proposal` file (a pdf).

✅ Yes, the zipped folder submitted has both an HTML file and the PDF proposal. 🙌🕶️

Instructions: Annotator Instructions & Examples

The Instructions html file includes the following sections:

1. Overview
2. Steps
3. Rules & Tips

4. Examples

5. (Untitled) Visible Test Questions, which show the general layout of what an annotator will see during this job; no answers need to be selected.

The Instructions html file includes the following sections:

- ✓ Overview
- ✓ Steps
- ✓ Rules & Tips
- ✓ Examples
- ✓ Visible Test Questions, which show the general layout of what an annotator will see during this job; no answers need to be selected.

Your Overview should briefly describe why you are creating the data annotation job, and the Steps should clearly explain what is expected of an annotator.

A clear explanation of your product goal and annotation steps has been provided in the overview and steps section. Good job !!

Overview

Help us determine whether the images indicate the presence of pneumonia in the patient x-ray or not.

Steps

1. Examine the image.
2. If there are visible symptoms of pneumonia, answer with "Yes"
3. If there weren't visible symptoms of pneumonia, answer with "No"
4. In the case you're unsure of the answer, you should answer with " Not sure".

All possible labels/answers should be clearly defined in the Rules & Tips section. If selecting a certain label is not obvious, it is best practice to add clarifying criteria, here.

The rules and tips are well detailed, explaining how to identify a case with and without pneumonia.

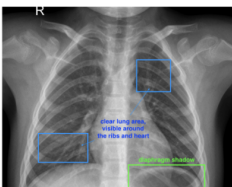
Rules Tips

Rules:

- Only look for Pneumonia symptoms.
- A **normal**, healthy image will depict clear lungs without any areas of abnormal cloudiness/opacity; there may be structured, web-like vasculature in the lungs but otherwise that area should be clear. In healthy images, you are also more likely to see a diaphragm shadow.
- A **pneumonia** image may include a few things: areas of cloudiness/opacity in several concentrated areas or one large area. You may also see a general pattern of opacity that obscures the structure of the lungs, heart, and diaphragm.
- Please in the case of uncertainty, don't take the chance and choose "Not Sure".
-

Tips:

-Some characteristics of a healthy image: a clear lung area.



-Examples of pneumonia symptoms: (Left) a concentrated, opaque area in the lungs, (Right) multiple, smaller opaque areas throughout the lung area, and any diaphragm shadow is obscured.

The Examples section should include at least one example for each possible data label.

Done. The Examples section includes at least one example for each possible data label.

Proposal: Design & Quality Assurance

Include a short overview of this project and the product goal.

A valid project overview and goal has been provided.

Project Overview and Goal

What is the industry problem you are trying to solve? Why use ML in solving this task?

Help medical professionals in identifying possible cases for pneumonia disease in x-ray images of patient lungs faster and eliminate cases that do not have any apparent symptoms. Using ML will help doctors to consider their decisions more carefully if the model show different results than what they assumed.

Explain why you chose the labeling scheme that you did. What are the strengths and weaknesses of such a labeling scheme?

Appropriate labels have been provided for the data annotation job.

Choice of Data Labels

What labels did you decide to add to your data? And why did you decide on these labels vs any other option?

I decided to use three labels "Yes", "No", and "Not Sure". "Yes" & "No" is to determine if the x-rays image is of pneumonia symptoms. "Not sure" is there to leave room for uncertainty because it's a sensitive case. I chose those three labels because they're simpler than say "Healthy lungs", "Pneumonia symptoms", and "Unknown".

You should plan to include at least 5% test questions to mix into your training set or about 1 test question for every 19 data points you want to label.

A relevant estimate of the number of test question has been used.

Your test questions may not be perfect; you should have a plan for revisiting their efficacy. You should answer the provided questions about how you might handle scenarios in which multiple annotators contest or fail a test question.

Great work!! Meets the specification

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"/>

- 1- Remove any ambiguities that might cause confusion.
- 2- Explain the steps more carefully
- 3- Add tips to help annotators identify the correct answers
- 4- Make sure the rules are clear and easy to understand

You should include a description of what might be missing in your data annotation job and what might be improved. You should provide answers to the following questions:

- Could the data source be improved? If so, how?
- Do you think your test questions or labeling scheme could be improved? And how do you think you could ensure the quality of data over time?

You might also consider involving stakeholders (engineers, medical professionals, etc.) in discussions about product improvements.

Impressive work done here.

Data Source

Consider the size and source of your data; what biases are built into the data and how might the data be improved?

Improvement:

- 1- Have a larger set of data large enough for a ML model to learn patterns.
- 2- We need the ML to deal with all possible scenarios
- 3- There's no bias as of now only 24 labels are given, 8 for each label. So in the case that we have some bias we will need to throw away data from the class that have more data
- 4- The Data source could use some improvement such as: different lightening conditions, illumination, angles and other considerations
- 5- Keep the data regularly updates

You should note which areas of your instructions and test questions you might improve according to annotator feedback.

Perfecto!! Good design for longitivity.

[!\[\]\(a8f9309f944226d1420f5fed22e2b6e6_img.jpg\) DOWNLOAD PROJECT](#)

Rate this review

START
