

Thank you so much for participating!

\*If hasn't filled out recruitment survey\* To start, would you mind going to this link to go through the consent form and answer a couple questions? [LINK HERE](#)

I want to start off by giving you a brief summary of this study. We are interested in understanding how medical professionals may interact with an AI assistant in evaluating chest x-rays. I will ask you to first complete a short survey, and then react to mockups of a clinical decision support system, which you will use to evaluate 24 patients' chest x-rays for pneumonia. At the end, I will ask you follow-up questions and have you fill out a final short survey about your experience with the mockups.

Please feel free to ask questions at any moment. Do you have any questions so far?

I would like to ask that you please be honest in your comments and answers throughout the study. There are no right or wrong answers, and your honest feedback will help us think about how AI can best support medical professionals.

Please fill out the survey at this link that I will send in the chat: [LINK HERE](#), and please let me know when you are done filling it out.

**Are you okay with me recording and transcribing from here on? The recording and transcription will be seen only by people working on this study, and if your comments are ever shared or published beyond the study group, they will be anonymized.**

Great! I will now share my screen to present you with the mockups.

I will show you 24 patients' chest x-rays and click through the mockups for you. For each patient, you will be asked to decide if the patient has pneumonia or not. When you are finished deciding, please ask me to click "yes" or "no."

I know there are several factors about a patient that you can take into account when diagnosing patients. If you would want to look into certain additional factors about a given patient before making a decision on whether or not they have pneumonia, please let me know.

**Please note that the AI model will be presenting anomalous input and output more frequently than in the real world, and I will sometimes change how the AI model presents information from patient to patient.**

Please also note that your decisions will not directly impact any AI model. This is only a simulation.

Lastly, I know very little about healthcare and have no background in it, so please let me know if anything we discuss would not make sense in your context.

For each x-ray, an AI suggestion will be shown. The AI provides its prediction of whether or not the patient has pneumonia. The AI also provides its confidence in its prediction and an explanation of its prediction. The confidence represents how confident the AI is that the prediction should be what it said and not the opposite, so a confidence of 50% means it thinks the verdict is equally likely to be pneumonia or no pneumonia, whereas a confidence of 100% means it thinks the verdict is definitely pneumonia or no pneumonia. The explanation is a heatmap highlighting the parts of the image on which

the AI focused in making its prediction. The redder an area is, the more focused the AI is on that area. The heatmap explanations are only for pneumonia predictions.

We will now go through two examples for you to see how this will work. You can see the chest x-ray on the left, and on the right, the AI suggestion with its prediction, confidence, and heatmap explanation. Would you say this patient has pneumonia or no pneumonia? Yes, the patient has \_\_\_\_ OR The patient actually has \_\_\_\_\_. And here is the second example. The same information is presented except there is no heatmap explanation because the AI predicts no pneumonia. Would you say this patient has pneumonia or no pneumonia? Yes, the patient has \_\_\_\_ OR The patient actually has \_\_\_\_\_.

Any questions?

We will now move forward with evaluating each chest x-ray. Please think aloud as you interact with the mockups, and I may also chime in with questions about your interaction as you go along.

Questions to keep them talking as they think aloud while engaging with the mockups:

- What are your thoughts on what you're seeing?
- What are you thinking?
- How did you decide to click yes/no? \*Do this to clarify ANY TIME they don't explain their choice, NOT for specific instances. Don't want to bias.\*
- \*when there's a varied presentation of same x-ray\*
  - o What if the same AI suggestion for the same patient was presented with this more detailed alert - what is your reaction to this additional information?

\*stop sharing screen \* I would now like to ask you some follow-up questions regarding your interaction with the mockups. Please let me know if you would like me to go back to any particular mockups as we discuss them.

Permanent + evolving questions at end of interaction:

- General I
  - o **How would you describe how you used the AI, if at all?**
  - o **How did you utilize the confidence, if at all?**
  - o **How did you utilize the heatmap explanation, if at all?**
- Challenging vs. Trusting the AI as an Assistant
  - o **Under what circumstances, if any, were you inclined to challenge the AI's decision?**
  - o **Under what circumstances, if any, were you inclined to go along with the AI's decision?**
  - o **Under what circumstances, if any, did the AI affect your confidence in your own decisions?**
- Time Efficiency
  - o **When, if ever, did the system slow you down in evaluating the chest x-rays?**
  - o **When, if ever, did the system help you move faster in evaluating the chest x-rays?**
- **PAUSE FOR FOLLOW-UP SURVEY:** I'd like to pause the follow-up questions at this point and have you fill out the final short survey at this link that I'll send in the chat: *LINK HERE*. Please let me know when you're finished filling it out, or if you have any questions.
- Great, I'll now continue with the remaining follow-up questions.
- High Conf **\*pull up slide if needed\***
  - o **How did the high confidence alerts affect your decisions, if at all?**

- How did the high confidence alerts otherwise impact your experience with the system, if at all?
- Low Conf **\*pull up slide if needed\***
  - **How did the low confidence alerts affect your decisions, if at all?**
  - How did the low confidence alerts otherwise impact your experience with the system, if at all?
- Unusual Expls **\*pull up slide if needed\***
  - **How did the unusual explanation alerts affect your decisions, if at all?**
  - How did the unusual explanation alerts otherwise *\*pull up slide\** impact your experience with the system, if at all?
- Model didn't see similar images in training **\*pull up slide if needed\***
  - **How did the alerts saying that the model didn't see similar images in training affect your decisions, if at all?**
  - How did the image-unusual-to-model alerts otherwise impact your experience with the system, if at all?
- More on anomaly alerts **\*close slides\***
  - **If an AI anomaly (low conf, image-unusual-to-model, unusual expl) is associated with reduced AI accuracy, would you prefer that the AI suggestion not be shown to you at all or still be shown to you?**
    - *\*if yes\* Would you like the AI to specify why it has decreased reliability (e.g., low conf, unusual expl, image unlike those seen in training) or would you prefer the AI just give a general alert like "CAUTION: AI has decreased reliability here"?*
    - *\*if no\* Would you prefer that all AI anomalies associated with reduced AI accuracy, such as low conf, image unlike those seen in training, and unusual expl, have one general alert like "CAUTION: AI has decreased reliability here"?*
  - Are there any kinds of suggestions you would like the AI to offer regarding how to deal with anomalous model input/output? For example, the AI might tell you that a second opinion is recommended.
- General II
  - Do you have any other thoughts on your interaction with the system?
- Other questions
  - How may you utilize the AI suggestion in writing a patient's report, if at all?
  - Would you prefer that the AI suggestions be shown to you right away, only on-demand, or right away in certain circumstances?
    - Would you prefer that the AI suggestions be on-demand rather than shown to you right away?
    - What do you think about the AI suggestion only appearing after you've made an initial decision?
    - What about appearing only if it disagrees with you?
  - Is there any form of feedback you would like to be able to give the AI as you interact with the system?
    - Under what circumstances would you give such feedback?
  - *Out of all the AI anomaly alerts you saw (high conf, low conf, unusual expl, and model didn't see similar images in training), were any more helpful to you than others? How so?*
  - *Are there any other kinds of anomalies in the AI's input or output that you think would be helpful for the AI to acknowledge?*

That's everything! Thank you for your time in completing this study.

Before you go, I would like to give you a short debrief on the study. There is one piece of information we did not provide to you at the start. In the introductory slide, we told you that the AI model will be presenting anomalous input and output more frequently than in the real world. The purpose of this study, in full, was to observe how you would react to the *highlighting* of anomalous model input and output. Our hypothesis is that highlighting such anomalies may improve users' ability to utilize the AI to make better and more efficient decisions. Some of the AI output that you saw was therefore manipulated in order for me to present you with diverse anomaly scenarios of interest.

Given this new information, if you would like to have your study responses omitted from our analysis, please let me know.

We thank you very much for your time and contributions to our team's work to better understand how AI can improve decision-making in healthcare.

Do you have any questions?