## Demo Structure

1. Introduction: (Cailin)

* Tuberculosis infects 10 million people a year and kills 1.5 million. If TB is diagnosed and treated early, most people are cured within six months of diagnosis.
* At the present moment, there is a large range of certified CAD products that may be used for TB detection and used to make professional recommendations.
* However, an issue arises: what product is most fit for the purpose? Ideally, you want a product that is the most accurate at detection and offers the best user experience.
* However, due to the amount of research and the difficulty of keeping up with the fast-paced medical technology field, finding all the needed information for analysis is difficult.
* qXR is a highly reputable technology, with the ability to detect multiple abnormal findings on a chest X-Ray in less than one minute. However, to see where this product stands compared to other AI TB detection products, our team will use qXR as a benchmark against other technologies and compare accuracy and usability to determine if qXR is a technology that should be suggested or would an alternative be more advisable.
* The project’s aim is to evaluate the product qXR using the ECLAIR guidelines.
* Primarily, this involves identifying the best computer-aided detection (CAD) commercial products that will be able to diagnose TB that is on the market for our clients to use with high confidence and research to support the decision.
* In addition, the team also plans to have an evaluation on user experience by identifying factors affecting user acceptance and the product’s impact on the clinical workflow.
* A final report will analyze the various AI technologies using the qXR product from the company Qure.ai as a benchmark. At this stage we will evaluate if qXR is the best choice, or suggest an alternative product instead. Factors we will consider are its accuracy and an identification of factors affecting user acceptance. All of these must be taken into consideration by the client when deciding which AI technology to implement.
* To ensure the quality of the content, we are using industry-accepted protocols and guidelines including PRISMA (for our source finding) and ÉCLAIR (for the structure of our evaluation) to ensure high quality systematic reviews and evaluation.

1. Project Pipeline (Benjamin)

Previously, our team had successfully set up and adjusted the project scope, allocated roles and distributed tasks. More importantly, the concept of systematic review and meta-analysis had been fully comprehended and accepted by each of the team members. We divided the research into two separate parts in general, and each of us had an individual area to research. This is to make sure that we do not have an overlap when researching, and the evaluation is done as thoroughly as possible. At this stage, we have already completed our individual part of research following the industrial standard strictly. The result of our individual research along with other project related matters will be presented today. For the next step of the research the most important thing is to adjust our research result and enhance the result according to the clients’ feedback. Moreover, we will try our best to combine our individual results to form a formal integrated conclusion and present it clearly and precisely.

1. User stories: (Kiran)

In order to define the features of our project that would need to be delivered to you in our project we wrote a series of eight user stories. We have included a few on this slide that I will go through to give you a better understanding of what we hope to have achieved by the end of this project as well as giving an indication of the formatting of our user stories. Our top priority user story as indicated by the one is more general and outlines the need for the report to be comprehensive, thus the acceptance criteria is that we cover all requirements you have given to us both in the brief and in meetings, further, we have labelled this as a non-functional requirement. Another user story is from a radiologist that will be using the selected product and shows their need for a more accurate product with an acceptance criteria that we have shown the accuracy of the various AI products in the report and we view this as a functional requirement. The final user story I will show you is again from the view of a radiologist and indicates the need to prove that qXR is the superior product and we believe this can be achieved by benchmarking it against other products to show its value, this is again a functional requirement for our report.

1. Technical Article Gathering (Kiran)

For the technical team we conducted a systematic review following the PRISMA process [show process with removals], we assigned two products to each member of the team so we could focus on gathering data on each of them in comparison to qXR, our search queries were therefore quite simple and involved “qXR”, the name of the other product, “comparison” and “TB” or “tuberculosis”. At this stage many products had to be excluded due to a lack of reliable information on their performance capabilities. On the slide you can see an example of how we stored the found tables with the name of the paper, the database it was retrieved from, a link to the article and the status of whether it was a duplicate, kept or excluded as well as the reasons for exclusion. Dividing the work in the way that we did we had many more duplicates than expected as many members found the same article on the same database due to overlapping analyses.

1. Technical Articles spreadsheet (Kiran)

After we had identified the articles we were using we extracted some of the key information in the form of a table of study characteristics, a snapshot of this spreadsheet can be seen on the slide and indicates the first author of the article, the year published, the country or countries where data was collected, a link to the article, and for each product we have included the area under the curve as a measure of accuracy as well as sensitivity and specificity, accuracy was also found in some articles, however, not all thus we will likely not be using that information in further analysis and the final report.

1. qXR v InferRead (Edward)
2. qXR v LUNIT v VUNO (Andersen)

For Lunit, there are many comparative articles that support me to finish this diagram

From this diagram we can see that qXR outperforms Lunit in most cases. Regardless of the article, qXR has a better AUC, which means qXR is more capable of identifying abnormal cases. But in specificity, the capabilities of Lunit and qXR are not much different.

Due to the lack of analysis on public/3rd party forums, I can't find experiments comparing VUNO with other products. While the existing results are from VUNO official, the results are too good, so we assume it is biassed. that is why we gave up comparing VUNO

ChestEye

there are only a few articles for ChestEye to make the diagram.

According to the diagram, qXR performs much better than ChestEye in most circumstances. qXR is more capable of identifying abnormal cases as it has a much higher AUC than ChestEye which is around 10 percent. Also, qXR has a higher specificity than ChestEYe which means qXR has a higher capabilities.

1. Also explain why drop vuno
2. Excluded Products (Kiran)
3. UX: PRISMA and limitations (All user experience)

* Introduction (Bronte)

The second primary focus area for our evaluation is usability and integration. This section of the ÉCLAIR guidelines has 6 questions, and we are currently working on answering 3 of them. These are; how exactly does qXR impact the workflow? How can qXR be integrated into your clinical workflow? And what are the requirements in terms of IT infrastructure?

For the first question, we are undertaking a systematic review and meta-analysis of user experience and acceptance studies to identify factors which are having an impact on the workflow.

For questions two and three, we are looking at the hardware, software, and storage requirements for qXR and the method for integrating it into the clinical workflow.

* qXR and limitations (Bronte)

We began our usability evaluation of qXR by undertaking a systematic review using PRISMA. On the slide here we have our PRISMA flow diagram which shows that from our searches, we identified 27 articles and 17 remained after duplicates were removed. We searched Google Scholar, ProQuest and PubMed, and used keywords such as user experience, user acceptance, usability, and workflow integration, along with qXR. 16 of the articles were excluded after screening their titles and abstracts and the final article was excluded after a full text reading because it was not a user experience study. The inclusion criteria was that the article was a study of user experience, user acceptance, or impact on workflow about the product qXR in a standard clinical setting.

Therefore, we have identified a major limitation for our evaluation, that is, there were no user experience studies of qXR identified in our searches. Our solution to this limitation is to expand our evaluation to AI in radiology and apply the assumption that the evaluation will identify generic factors that could apply to qXR.

* PRISMA TB
* Expanded PRISMA UX broader
* Evaluating UX for other medical imaging AI products
* My research covers the user experience evaluation targeting AI products used in medical imaging other than qXR. After we changed the scope of our project, we decided to take our evaluation to a broader horizon. The purpose for this part of the evaluation is that we could explore the general attitude of the radiology community against the use of AI. More importantly, we could identify the key factors the users are looking for when choosing the AI product and potentially take the factors back to qXR for examination. Also, it is the best that we could build up a plan of improvements when it comes to the use of AI products instead of only taking the developers perspective. For example, what kind of training is necessary for radiologists before they use medical imaging products at work and what kind of suggestion could we form and provide to qXR after this research.
* The research followed the PRISMA process strictly, I tried to collect articles as thoroughly as possible. The database chosen is PubMed, ProQuest and Google search engine. I chose PubMed and ProQuest based on their highly recognized academic reputation. Apart from that google search engine could help increase the diversity of the articles. Initially, 78 articles were collected from these platforms using key words including user experience, user satisfaction, UTAUT, TAM, acceptance, AI, and medical imaging. 2 duplicated articles were removed, 76 were left after the identification stage. After reviewing the titles and abstracts of the 76 articles, 56 articles were excluded mainly because of the lack of relevance and missing key concepts. This left me with 20 remaining articles. After reading the full text of all of them, 15 were excluded and the major issue was that there were not enough participants and the quality of the evaluation did not meet our ideal standard. Finally, there were 5 articles left.

Next slide

* Some interesting results from the remaining articles are listed as bullet points, and it is my belief that they will be further integrated in order to form the former suggestion. The results indicate that only 55.1% of the radiologists are familiar with the basic concept of AI and only approximately 46% of them had received AI related education. These statistical results are suggesting AI related training and education is lacking among the radiology community, and for the better implementation of AI products like qXR, making sure that the radiologists using it understand the basic concept of the development of the product is necessary. A user experience evaluation done targeting a chest-CT AI product TRx used the system usability scale to evaluate the perceived ease of use. Based on the result, the perceived usefulness and perceived ease of use were highly rated, however, the output quality can be poor when it comes to lung opacities. In another theory of planned behaviour model, factors like subject norm, perceived self-efficacy and personal relevance are the factors affecting the behaviour intention, and the behaviour intention has a direct influence on the actual behaviour when it comes to the use of AI. Another study integrated a TAM model with a trust-based model resulting in some interesting findings as well. It demonstrates that the initial trust and social influence are the most important factors affecting behaviour intention.
* Because of time limitations, I won’t be going through each result and factors in more detail. There will be more specific analysis once I integrated all the results and it will be present on the final presentation and documented on the final report. With these results and factor analysis, we could go back to qXR and its competitors and form a more valid result.
* Additional Articles and study characteristics (Bronte)

In addition to the articles we have identified through the systematic review following the PRISMA process, we have also included 5 articles which were provided to us by Simon. They meet our inclusion criteria because they are studies of user experience or acceptance in the topic of AI in radiology. To begin our process of analysing the included studies, we have started extracting the study characteristics. We have designed this table based on the SPICE framework which is used for qualitative research evaluating a product or intervention. For each study included for analysis we are recording study name, year, type, setting, perspective, number of respondents, Intervention, comparison, and evaluation.

* How to integrate into workflow (software, hardware, DB) (Cailin)
* Hardware, Server, Price: we have emailed partner@qure.ai waiting for information.
* Software: Ubuntu 18.04 is preferred
* Validation: qXR can generalize to any X-ray manufacturer (CR/DR) and has been tested with over 20 leading X-ray manufacturers globally.
* *Integration:*
* It is possible to integrate qXR with:
* the client’s legacy standard reading environment PACS or Directly to the X-ray device
* Integration can be done via AI marketplace or distribution platform and be done on a Stand-alone web based
* Additionally, each deployment of qXR can be coupled with access to qTrack to facilitate viewing of results and managing the screening workflow.
* Integration Modes
* API based:
  + - Dicom images can be uploaded to our REST APIs through a token-based authentication. Results can be downloaded using a separate API endpoint
* PACS based:
  + - PACS integration based on DIMSE protocol to transfer raw scans and receive outputs. Our gateway can be added as a Dicom node in your network which will receive images, anonymize them, upload them using the APIs, download results and send these results back to your PACS.
* *Hosting Options*
* Cloud-based:
  + - Customers can use our cloud servers to upload scans and receive results for both modes of integration.
* API Based:
  + - Private cloud
    - Local IT infrastructure, which offers deployment of products on local IT systems at hospitals & radiology centers.
* With either option, a cloud-hosted setup in which scans are uploaded to Qure.ai’s servers or a locally hosted, on-premise solution that uses off-the-shelf hardware, it will grant radiologists access to a workflow management platform that supports patient registration and tracking, as well as follow-up visits and X-ray screenings scheduled for those visits. From a dashboard, admins get an overview of registered patients with the size and location of their abnormalities, plus their bacteriological test results and radiology reports.

1. Testing plan (Bronte)

As you have discussed with us, we know that you are intending to use our evaluation of qXR to inform your decision whether or not to proceed with purchasing or trialling qXR. So we have designed a comprehensive testing plan to ensure that our project is reliable and of high quality. The plan is divided into 6 areas.

First is article quality. To ensure article quality we are using databases and Google Scholar to search for articles, The articles are limited to journal articles, and to reduce the risk of bias, we are excluding articles whose studies are funded by the company whose product is being studied.

Second, is systematic review. We are following the PRISMA process to ensure that the systematic review is of a high quality. This means we are designing search queries with keywords relevant to the research question, including all results of search queries to ensure nothing is missed, defining a clear inclusion criteria relevant to the question, screening the title and abstract of articles using the inclusion criteria and reading the full text of the articles and using the inclusion criteria to select articles to be included. We are also providing justification for the decision of each article.

For meta-analysis, we are having a second person check the data extraction completed to reduce the risk of data entry errors. The data is also being extracted to spreadsheets which are designed to capture all the relevant data in a uniform and systematic manner. For diagnostic test accuracy, the data visualisation tool Tableau is being used for statistical analysis, which will include constructing forest plots as well as other relevant graphs. For user experience, we will be using the TAM and UTAUT model to provide the basis for the factors we identify.

(Next slide)

To ensure evaluation quality, we are following the ÉCLAIR guidelines which are a practical set of questions to evaluate commercial AI solutions. The five areas of evaluation are Relevance, Performance and validation, Usability and integration, Regulatory and legal aspects, and Financial and support services considerations. As we are focusing on the diagnostic test accuracy and user experience of qXR, the two questions relevant to these are answered in-depth through a systematic review and meta-analysis. For the other questions, a high-level response will be provided.

For report quality, the structure will follow the evaluation areas of the ÉCLAIR guidelines, with an additional executive summary, introduction, recommendation, and conclusion. We will also use academic writing and provide complete and consistent referencing.

Finally, for our presentation we will provide a complete overview of our evaluation, our slides will have a professional look and include visual representations of our findings, and our presenters will be prepared to answer any questions about the evaluation.

1. Next steps (Zheyuan)

For the remainder of our project, we need to finish our meta-analyses, the remaining ECLAIR

guidelines questions, determine our recommendation, and submit our report and presentation.

For the meta-analysis technical side, this will include creating forest plots and other relevant graphs. And for the UX side this will include completing the extraction of study characteristics and grouping together similar areas for analysis to identify which factors are impacting on the workflow. ‘’

We still have a range of questions to complete for ECLAIR, including legal aspects, financial and support services considerations, relevance, and sub-questions for performance and validation and usability and integration.

Once we have completed the evaluation of qXR, we will make a recommendation whether to purchase qXR, trial qXR, or not proceed with qXR with a clear and logical justification.

Finally, we will submit a report and conduct a presentation containing our evaluation and recommendation.

1. Q&A - record feedback