# Presentation Slides (1 min per section)

## Quality of the overview of the project - Cailin

P56: Benchmark Commercial AI product for Medical Imaging Services Provider

The project involved an:

* Evaluation of the commercial AI product qXR by conducting a meta-analysis and then providing our recommendations to our clients which will help their decision whether or not to invest in qXR.
* Identification of factors that affect user acceptance when AIs are introduced to medical imaging practices from a thematic analysis to help help limit the impact on the workflow when the new technology is introduced

These goals were achieved by our team using PRISMA guidelines for source finding and ÉCLAIR guidelines for the structure of our evaluation

Our stakeholders include: Simon Poon and Neysa Petrina (Information System researchers) and

ABC Company (Medical health care professionals and their patients )

Risks include:

* Insufficient evidence
* could not benchmark the diagnostic test accuracy with every certified CAD for TB product, which limits our analysis and risks there being an alternative product that is better performing that we would not know about
* The user experience analysis is from pooled data of multiple products (due to their being no studies on the qXR’S UX), therefore there lies a risk that qXR’s user experience is different to the AI’s we study.
* Limited time with the research process and lack of support from academic staff (due to us not being postgraduate/honours students).
  + The zero support and small time can result in having potential data quality issues with data collection for our final report as our team is unable to dedicate time to checking the values before analysis due to the tight deadline provided

## Quality of system specification- Kiran

Our system specification was informed by the creation of 26 user stories, these informed both functional and non-functional criteria for our end product. Our functional requirements were majorly defined by the ECLAIR guidelines which will be elaborated upon in later slides, further to this, a major requirement was the benchmarking of qXR against other AI products as well as identifying the efficiency of qXR. Non-functional requirements identified included the report being comprehensive and covering all requirements set out by the client, meeting time demands of the client and utilising reliable research. We identified some technical restraints that prevented us from fully realising this project, consistency between different papers as well as the products being analysed at differing versions between papers had some restraints in final analysis, biases existed toward certain products in many articles and these were identified and caused the article to be excluded and finally the availability of articles on these products was limited due to the novel nature of the products and lack of certification of many. Here you can see a few examples of user stories including the number of the user story, a short description, acceptance criteria, a definition of what requirement it is and its status, here you can see our user story on comparing accuracy defined by AUC was incomplete, this is because we opted to use specificity and sensitivity in our final analysis, which will be elaborated on in the demo.

## Remaining User Stories - Kiran

This slide shows the other 23 user stories as well as their acceptance criteria, these are more comprehensively discussed in the group report.

## System Architecture and Design - Zheyuan

The System Architecture is basically split into 5 parts, first, the team will collect the AI radiology related articles online, the team will research through academic websites to find articles able to answer the ECLAIR questions. Then, two systematic reviews were undertaken, the team evaluated the performance and how the products impact the workflow. The team drafted a systematic review protocol to explain the whole systematic review process. The third part of the system is data synthesis. The team extracted the data from those articles through systematic review process and synthesized them, the technical team followed meta analysis process to sythesise the data and prodced forest plots for each products, the user experience team used thematic analysis to perform their analysis. Afterwards, the team followed the ECLAIR guidline, the team use comprehensive evaluation of qXR to answer the 20 ECLAIR questions, first is an executive summary, an overview of the foundings will be given, and an introduction will identify why the team undertake the report and the objectives. Then relevance is introduced, it is a more detailed definition of qXR, for example the benefits of AI in radiology. Afterwards, the technical team analyse the performance and Validation while user experience team analyse the usablility and integration of qXR. Next, the evaluation explain the regulatory and legal aspects to ensure qXR is legally available. Furthermore, the report evaluate the financial and support services considerations to inform the client

of how to pay for the product. And finaly a conclusion is provided.

All to the end, after all the research process is done, a final recommendations of what actions to take and issues when investing will give to the

client.

## Quality of the whole work - Bronte

To assess the quality of the work done in the project, we adopted two methods of testing, quality assurance testing and acceptance testing. Since our project does not include programming, software testing methods are unsuitable, however, we believe that the testing methods selected are sufficient for testing the project. The quality assurance tests begin with the smallest component of our project, the articles we collected, and progressively move through the project to the final report as a whole, therefore, as each component of the project is added, the project is tested to determine its quality. The acceptance tests use the user stories which were created to cover the entire project scope, therefore, assessing whether they are implemented sufficiently determines whether the project scope was fulfilled. A limitation of these tests is that they are manually tested and likely contain a level of subjectivity. This could have been mitigated by having our report peer-reviewed by an external person, however, this was not able to be completed within the limited time to complete the project.

The quality assurance tests are divided into seven sections, each which covers a specific component of the project. Each section has a criteria that the project was tested against multiple times over the course of the project by manually reviewing the work completed and making improvements where necessary.

## Quality Testing - Bronte

In the interest of time, I will only go through the article quality testing. The articles we selected are foundational to the quality of our project. They are our research for conducting the evaluation, so if they are of a poor quality, our entire evaluation and resulting recommendations would be unreliable. We developed 3-part criteria for testing article quality and then tested this multiple times over the course of the project. As you can see in the table, we reported when the testing took place, the test results, and the improvements made. In our first test, we had some articles that were found through Google Search and some that were not journal articles. When undertaking the systematic review process again, we made specific improvements to improve article quality. These were that we did not use Google Search and we decided to exclude Google Scholar as well because it doesn’t define its criteria for including an article, and therefore, can include poor quality articles and results that are not journal articles. Additionally, group members used filters where possible in database searches to limit the results to journal articles. The second time the articles were tested, some problems were found with specific articles and these were rectified by removing the article or re-identifying it through another source. Therefore, the final time our articles were tested, they satisfied the quality criteria.

## Acceptance Testing - Bronte

Acceptance testing was formally completed in the final week of our project to determine whether all of our user stories were implemented and therefore, the project scope fulfilled. For each user story, one or more acceptance tests were defined and tested by manually reviewing the work.

Our first user story is, as a medical imaging provider, I want a comprehensive report so that I can justify my use of qXR. For the first two tests, every ECLAIR question was assessed to check whether it was answered and contained references, and analysis. For the third test, the conclusion section which contained our recommendations, was compared to the previously assessed ECLAIR questions to determine whether the recommendations were based on them. The acceptance testing for this user story and 24 others were successful.

However, the user story, as a radiologist, I want to have an accurate AI product to detect TB so that I can give the correct diagnosis, was incomplete. This was because of issues with the technical data collected and a lack of time to search for more data, meaning some meta-analysis steps could not be completed. This issue is clearly explained in the report as a limitation of our analysis.

## Quality of the application of discipline knowledge and tools - Andersen

In this project, the knowledge we need are ISYS3401 background knowledge, PRISMA (a protocol for Systematic Reviews and Meta-Analyses), Thematic analysis (a common form of analysis within qualitative research), ECLAIR (a commercial static code analysis tool) , ROC and AUROC (a performance measurement for the classification problems at various threshold settings), and Heterogeneity (it is very common in meta-analysis which means there is variability in your data)’

Talk about tools, they can be divided into 3 aspects.

Since we are a hybrid team, communication tools are very important, we used slack for text contact, and slack also provided a grouping function, which can help us increase our productivity.

We also use zoom for different types of meetings, such as tutor meetings and client meetings.

For file management, we used google drive and bitbucket. Bitbucket is the official one, so we didn't upload any draft. For those early-stage works, we stored them in google drive

For Data analysis and visualization, some of us try to use R studio, but tableau also can satisfy project requirements, and tableau has better eases to use, which is why we choose tableau.

## Quality of response to key changes (scope changes) - Edward

For the key changes requested by our client, the main change is the scope change in our project. At the first stage, the team is required to conduct a systematic review and meta-analysis to benchmark the diagnostic test accuracy studies on different certified AI products to detect TB. The team needs to find out which AI product performs better in detecting TB and make a summary for our client. At the second stage, in order to make the work more clear. Our client suggests us to divide into 2 sub-teams which are technical team and user experience team separately. Specifically, the task for technical team remains same which is using systematic review and meta-analysis to benchmark the DTA study for different AI products to detect TB. The user experience team will need to use systematic review and meta-analysis to measure the factors affecting user acceptance of AI used in medical imaging diagnosis. It’ll be better to extract articles related to AI in medical imaging, and even better for AI in TB diagnosis. However, there are few articles related to user experience using AI to detect TB. So the ux team will expand their search range to all medical imaging studies. At the third stage, the technical team is suggested to focus one certified AI product which is qXR and we are going to benchmark qXR against all other competing AI products in TB diagnosis. The scope for ux team remains same but will look for user experience studies specific to qXR as well. In addition, our client provided ECLAIR guidelines for us to structure our client report. This guideline proposes a framework on what we need to analyse when assessing an AI solution in medical imaging. At the final stage, the ABC health service company has hired the group as consultants to prepare a professional report evaluating the product and the team is also required to provide a recommendation of whether using qXR or not. ECLAIR guidelines are the method for evaluation report and the report will focus on DTA studies and user experience specifically. That’s all for the key changes in our project.

## Demonstration (video or live) of the final product/design (most of the team)

## Introduce project: ECLAIR - Bronte

Our client is acting as a representative for the health services company ABC, who is considering whether to adopt qXR, an AI tool which is used to interpret chest x-rays to diagnose diseases such as tuberculosis. At their request, we have evaluated qXR using the ÉCLAIR guidelines which are a practical set of questions specific to commercial AI solutions in radiology. This informs our recommendation to the client of whether to invest in qXR.

## ECLAIR recommendations - Bronte

The ECLAIR guidelines are divided into five sections and have a total of 20 questions. I’ll briefly go through the recommendations that arose from the questions and then you’ll hear more about our primary focus areas from Kiran and Benjamin.

The first section is relevance, which considers qXR’s benefits and risks. We recommend that ABC invests in qXR to take advantage of its benefits, including improved efficiency, workflow optimisation, and reduction in misdiagnoses. Something which sets qXR apart from its competitors is its suite of supporting products, such as qTrack, which is an end-to-end disease management platform. However, there are also potential risks such as automation bias, legal liability, restricted access to data, and weak cybersecurity. For this we recommend maintaining a risk register to monitor and manage the potential risks.

Next is performance and validation which investigated the performance of qXR and details about its algorithm. qXR’s performance will be elaborated on later and for the algorithm, it was found that qXR had been extensively trained, identified biases and accounted for them where possible so we recommend using qXR.

For usability and integration, we looked into multiple methods of integrating qXR into the workflow and recommend that using an AI marketplace vendor would be the most simple method. This is because they can take care of many of the integration aspects. Additionally, ABC should provide staff with support and resources to equip them with the knowledge to understand, trust, and use qXR.

For the legal aspects, qXR is currently not approved for use in Australia as a medical device. We recommend that ABC seeks a sponsor for qXR who can complete the application process with the Therapeutic Goods Association for it to be approved.

Finally, for the financial and support services considerations, we recommend that ABC determines the number of annual studies, concurrent users, and sites so indicative pricing can be sought.

## Show the tech data analysis - Kiran

## Forest Plots

## We used forest plots mainly as a test of heterogeneity, this is the process of comparing the results of these products and making a judgement on whether the results across studies were homogenous or heterogenous following the guidelines in Diagnostic test accuracy methods for systematic review and meta-analysis. This is an example forest plot showing the specificity and sensitivity that were extracted from one of our articles, in total there were 6 pairs of forest plots created, we then met and decided that the data across articles was homogenous which indicated we believed that the data was fairly consistent and we were indeed able to conduct our meta-analysis. Forest plots are also a good tool in general for our data visualisation as they helpfully show confidence intervals unlike the second visualisation I will show you, this is important as it is necessary in making our informed decision in a recommendation to our client.

## SROC graph

## The data visualisation that most informed our final recommendation can be seen here, this SROC graph, that is a summary receiver operating characteristic graph, this plots all sensitivities and specificities of our articles in the same space, this was important in determining the overall diagnostic test accuracies of our products, as can be seen in the graph, qXR and CAD4TB significantly outperformed competitors, perhaps harder to see is the fact that qXR outperforms CAD4TB by a small margin, this is very important in DTA as client’s require only the most accurate products especially in the medical field. Further analysis using this graph could have involved an SROC curve which would have given further proof of the consistency of our data, however, it was found this would require further data from our articles which was not collected and further not present in all articles. Therefore, the tech team’s final recommendation in terms of DTA was for our client, ABC, to pursue qXR for their medical imaging needs.

## Explaining that there needs to be heavy importance on knowledge as that is the only way to build trust - - ben

This is a model we constructed according to the codebook. And this is a direct demonstration of the summary of the thematic review results. There are four constructs reflecting data from different studies that we adapted from the traditional UTAUT model. However, we added the personal attitude construct to record the evaluations of the general attitude of the participants. The factors under each construct are also labelled. The three factors under performance expectancy are efficiency, effectiveness and Extrinsic motivation. Perceived ease of use, ease of learning and complexity are under effort expectancy. The factors under social influences are subjective norms, social factors, and image. The factors under facilitating conditions are perceived behavioural control, support, and compatibility. Most importantly, acceptance, willingness, trust and risks are recognized as personal attitudes.

We found that the effectiveness of AI tools are being highly recognized by radiologists all over the world, but they also realised that their current workflow will be changed significantly. Their paycheck is likely to be affected, more training is expected, and knowledge of AI is required. We also found that surprisingly, effort expectancy in the use of AI is not related to personal attitude. More importantly, social influence is the key factor which impacts trust, and trust is one of the major factors of personal attitude. This puts social influence in a vital position when it comes to user acceptance.

In short, some conclusions can be drawn that the general trend of using AI in medical imaging is highly recognized by doctors globally, and they are willing to embrace the innovation. However, the most concerning issues for doctors when it comes to the use of AI tools in the medical imaging community are jobs being replaced by AI and legal trust. If more knowledge can be comprehended by radiologists, these concerns can be reduced significantly. Also, more experience brings more confidence in the use of AI in medical imaging and social influence has been underestimated when it comes to acceptance of AI products.

## Final recommendation: choose qXR for the pilot -zheyuan

After considering the needs of the ABC Company and conducting a systematic review as well as a meta-analysis and thematic analysis,

The AI product the team recommended to ABC Health Services is qXR.

In addition, our team suggests that the ABC should put a heavy amount of emphasis on the knowledge of the users to help improve the user's acceptance towards the new technology.

Furthermore, we suggest our clients to do a pilot of the qXR first due to lack of user experience information,

We recommend our client to identify how the radiologist's opinion of using qXR during their diagnosis and also

the efficiency and effectiveness of the practical qXR performance.

## Quality of group processes - Men men

In general, the group cooperation for our team is efficient and proactive, the main reason is that we were able to construct a team structure according to the project aims and objectives. Thus, unlike the traditional coding heavy project, we splitted the team into two sub teams in order to conduct evaluations correspondingly. The subteams are called tech team and UX team based on the tasks. Each team has a manager for the general management of the team, including progress tracking, task allocation, quality control. There is also a manager appointed for the team as a whole who is in charge of client communication, meeting organisation and progress management. Kiran is the manager of the tech team, Anderson, Edward, and Zheyuan are the members. Bronte is the manager of the UX team, Cailin and Benjamin are the members, and Cailin is the team manager. Although we were not able to adapt to the XP roles, we were inspired by it and modified it based on our team’s unique condition. The role of Customer Liaison is Cailin, because she is in charge of client communication. Bronte and Kiran were appointed as programmers since they are managers of the two subteams and the responsibility collapsed. The rest of the group take turns in the other roles weekly, especially for manager, we decided that manager takes the responsibility of taking meeting minutes. As for team communication, the team was able to utilize slack and organize two group meetings online using Zoom. Also the manager used the SCRUM meeting questions to track the process the individual tasks, the questions include, qrevious work, work plans, and obsticales. Each of the team member demonstrated a proactive attitude and the atmosphere for team communication is always friendly. We also organized client meeting once a week to receive feedbacks and adjust according to their instructions. There two issues we encountered throughout this project, both had been identified and fixed. The first one is that the standards for the evaluation including meta-analysis and systematic review was not clearly provided by our clients, the quality of these process was once hard to control and there were no protocol to follow. We quickly decided we could write our own protocol for this specific evaluation, and the standard had been cleared since then. The other one is that we did not established an effective process tracking machemism, and this issue was fixed by the SCRUM meeting method mentioned previously.

task tracker spreadsheet listed out all the assignments and their associated tasks, which listed the team member allocated to the task, status and task due date, This spreadsheet ensured the team was up to date with deadlines and allowed for team members to easily update the status of their allocated tasks.

This is end of this presentation, thank you for listening.

# MARKING GUIDE

The presentation and demo should cover the points below (including but not limited to) in the following order:

* Overview of the project including goals, objectives, achievements, stakeholders and scope (concise and clear)
* System specification and design including:
  + Summary of key **functional and non-functional requirements** expressed in the format of **user stories**
  + Examples of **representative user stories** with the corresponding **acceptance criteria**
    - Each user story should be the smallest unit of work for specific functional or non-functional requirement instead of a combination of several user stories
    - Each user story should be clearly described and designed based on client's needs
    - Each user story should be testable and estimable
  + **Technical and other constraints**
  + Final (complete) **system architecture and design**
    - May include high-level designs (e.g., mock-ups, UIs, system/software components and its interactions)
* Quality of the whole work including:
  + **Testing plans** such as user acceptance testing, system testing, API testing and/or UI testing
    - Use of relevant **testing types** (e.g., unit tests, regression testing, integration, system, usability and/or acceptance testing)
    - A **justification** for why these tests are sufficient should also be included.
  + Use of relevant **testing techniques** (e.g. Equivalence partitioning, boundary analysis) to design test cases
  + Any other quality aspects or testing-related work **specific to the project nature** (e.g., API testing, UI testing, stress testing)
  + Demonstrate **execution of designed test cases** and corresponding **results** (e.g., bugs, bug fixes)
    - Test cases should be consistent with user stories and cover different aspects of all user stories
    - Explain why the current test cases are enough for the project
  + Demonstrate enough **test coverage** and explain why it is enough for the scope of the project
  + **Summary** of the significance and limitations of tests
* The application of discipline knowledge and tools (refer to the detailed requirements in '[Final Project Report (Group)](https://canvas.sydney.edu.au/courses/42688/assignments/373160)') :
  + Use and application of **discipline knowledge**
  + **Tools** used to build the systems
* Responses to **key changes** requested by the client (especially after the first client deployment) and how well these changes were handled or implemented as improvements or new features. (Zichen Li)
* **Demonstration (video or live)** of the final product/design:
  + It should cover all the implemented elements in the system/software/design or any work required to be completed by the client
  + It should match the user stories/requirements described above
  + The demo should be included in the submission along with the presentation slides.
  + It will be assessed based on how well (quality) the most crucial user stories are implemented
* **Group processes** including collaboration and roles, client interaction, and reflections on key improvements after the first set of assessment
  + Summary of collaboration and teamwork
  + A systematic process of work - weekly plan and task allocation
  + “Issues” tracking and progress;
  + Use of bitbucket and slack and other previously unfamiliar tools
  + Work with clients - group interaction with client
  + Summary of group member's contribution to the presentation (1 slide listing role/s of each member in the presentation)
  + **Critical assessment** and **Reflections**

**Note:** Make sure the presentation outline follows the same structure and order shown above.

**Note:** if any of the above topics is irrelevant to your project/work, you will need to explicitly justify it. Also, any other details that might be specific to the nature of your project/work and do not fit under any of the above parts should be explicitly added to the presentation.

### Requirements

* Each group has 20 minutes to do the presentation and show their demo followed by a Q&A section lasting 3 to 5 minutes with your tutor and other groups.
* All members must participate in the preparation and delivery of the presentation and the demo. All presentations and demos will be recorded. (Absent students will not get the group mark and may receive 0 marks if they do not contribute to the preparation and the presentation.)
* The maximum of slides is 20 (excluding the front page and references if there's any).
* There must be a cover page including the information of team members (names, SID, and Unikeys), formal group name (e.g., SOFT3888\_T17\_03), formal project name (e.g., P01 - Autonomous Car Using Neural Networks and Computer Vision), client name, your tutor name and tutorial time. (For group members who are doing multiple capstone units/projects, add a note to indicate the student name, and the other group name and the project name.)
* The presentation and slides should be concise, clear, formal and attractive to audiences.
* Be sure you have appropriate body language and eye contacts.
* The overall effect of the presentation and your own performance will be measured by tutors during the tutorial. But the finalization of assessment will take some time after the tutorial.

### Submissions

* Each group has to submit your slides and prototype demo (.zip file) through the submission link provided in this Canvas page by 23:59 pm, 24th October.
* There should be one submission per group
* Name your submission using the group name and project number (e.g., COMP3888\_TU17\_02\_P02.pdf)

### Marking Guide

* Quality of the overview of the project (10%)
* Quality of system specification and design (15%)
* Quality of the whole work (15%)
* Quality of the application of discipline knowledge and tools (10%)
* Quality of response to key changes (5%)
* Demonstration (video or live) of the final product/design (15%)
* Quality of group processes (5%)
* Quality of the presentation (15%):
  + Presentation is well-structured and has logical flow of information
  + Presentation is well-prepared and well-rehearsed
  + The delivery is within the time limit
  + Fully and appropriate utilization of presentation tools (slides, animation, visuals, etc.)
  + The slides are easy-to-read, clear and concise with appropriate titles and styles and without too much text
* Q&As and group contribution (10%) - convincing and sensible answers that show understanding of the work done and all members have sensible contributions to the presentation and demo

**Note:** No-contributing and/or absent students will not get the group mark.