**Investigating Usability of AI in Radiology to Identify it’s Impact on the Workflow: A Systematic Review Protocol**

# Introduction

## Rationale

The systematic review process is currently taking place in Sydney, participated by seven undergraduate students from the University of Sydney where two of them are located in China, and conducting this review remotely. The general scope of this project is to consult a company on the matter of choosing proper medical imaging AI products. The goal of the UX team is to explore the general user experience of the medical imaging community regarding the use of AI. More importantly, identify key factors in AI products that are critical to the current clinical workflow. Then, with this systematic review, the team could take these evaluations and factors back to qXR which is the product this project is aiming to investigate. Further, evaluate qXR user experience based on the highlighted factors. To conduct a systematic review properly is important for this project because it is necessary to review evidence regarding the usability of AI products in radiology as thoroughly as possible. Hence, in general, this review is aiming to answer the question of ‘How exactly does the product qXR impact the workflow of medical imaging?’

## Objectives

The objectives of this systematic review are to gain insight into how radiologists are reacting to AI technology being used in the medical imaging industry. Also, what are some significant characteristics of such products that are influencing the workflow when it comes to radiological practices? According to this aim, this systematic review should be able to answer the following questions:

1. What is the attitude of radiologists when it comes to the use of AI in medical imaging?
2. What is the radiologists’ perspective on the future role of AI in medical imaging?
3. What are some factors in medical imaging AI products that radiologists find important?
4. What characteristics are preferable in medical imaging AI products for radiologists? And what characteristics are not preferred?

# Methods

## Eligibility criteria

There are seven conditions listed as eligibility criteria below for the studies that this review is intended to collect.

1. This review would only collect studies from 2018. The reason behind this is that the AI products used in medical imaging have been innovating drastically only for around the last decade, and user experience evaluations done more than five years ago could not reflect the current attitude of radiologists. Hence, the criteria for the acceptable timeliness for this review is set to five years.
2. Journal articles are the only valid study form for this review. After consultation, the reviewers believe that the period for reviewing books will be too long for this review, and the quality and quantity of books on the topic of user experience evaluation on medical imaging AI products, are doubted. Journals, on the other hand, are the ideal form for the targeted studies because they contain valuable information and would not take too long to review.
3. The appropriate studies for this review should be taken in standard clinical settings with standard medical imaging equipment. Since the client of this review is from a comparatively developed urban area, studies from remote areas would not be considered relevant.
4. The data collected from the studies should be from primary data sources because the ideal evaluation for user experience should be based on the form of a survey or interview. Secondary data source fails in the aspect of providing first-hand responses.
5. The studies for this review will only be focusing on the perspective of people who are working or will work in radiology. Patients, on the other hand, are not the proper target for the studies of this review because they are not likely to produce valid opinions on their medical imaging results thus not considered users.
6. For the purpose of determining a suitable minimum sample size, the studies will be divided into two types. The first is a survey, which is defined as standard-form questionnaires which require limited effort to be completed, for example multiple choice or entering number. For these the sample size should be at least 120 participants, any study with less than 120 participants is considered insufficient because the smaller the size of the study is, the more likely it is for the result to be biassed by demography features, for example, the nationality of the radiologists.The second type is focus group which is defined as an interview or study where participants must complete a task before completing a questionnaire. For these the sample size should be at least 20. This is because these types of studies are expected to be of a smaller size and high-quality data can be extracted from them.
7. The subject of the studies for this review is limited to AI products in radiological practice and their usability. Subjects other than this are considered deviations from the objectives of this review.

## Information sources

The sources from which this review chooses its studies are being evaluated strictly by all the participants together. The database to search the studies should be academically approved and widely used by the scientific community in general. It should be recognized by both the medical community and the artificial intelligence community, based on the nature of this review. Thus, the sources of this review are listed as such:

1. PubMed. A highly rated database from the National Library of Medicine and the National Center for Biotechnology Information of the United States of America. PubMed is a well-known database that contains over 34 million citations for biomedical literature from multiple medical journals. It includes Medline which is a premier database for medicine and health services. - Ben
2. ProQuest. ProQuest is also a platform built specifically for researchers. The platform holds a great academic reputation among the scientific community, 98% of the top 400 universities rely on ProQuest when it comes to research.
3. Embase. A premier biomedical and health care database that is European-oriented. This database was recommended by the University Library website when searching for medicine and health related databases.
4. ScienceDirect. A science database which includes many scientific areas and specifically has good coverage of health sciences and medical research. This was chosen because it has a broader view to capture studies that may not have been included in medicine-specific journals.
5. ‘The Lancet’. ‘The Lancet’ is a medical journal dedicated to publishing cutting-edge research regarding medical science. All the articles published in that journal were carefully reviewed and read by doctors and scientists.
6. The client of this review. Simon is a representative of the client of this research. In order to guide the review team which consisted of first-time participants to the systematic review, Simon provided around a dozen of valuable articles for the team to review.
7. Qure.ai official website. Qury.ai is the company of the targeted product, qXR. Though the evidence provided by the product’s official website could be potentially biassed, the website still can be a valid source when it comes to gathering evidence and user cases after close examination.

## Search strategy

The main focus of the search according to the objectives of the review is user experience. Therefore, the search strategy focuses on retrieving studies that conducted a survey toward radiologists regarding the user experience of medical imaging AI products. User experience, in this case, could also be represented by terms including ‘user satisfaction', ‘usability’, ‘TAM’ (technology acceptance model), ‘user acceptance’, and ‘UTAUT’ (unified theory of acceptance and use of technology). With these components, proper search queries could be formed and searched through the already selected databases.

Some of the search queries are:

* AI for TB “usability”
* “Usability” evaluation for AI products used in medical imaging
* “User satisfaction” evaluation for AI products used in medical imaging
* “User experience” evaluation for AI products used in medical imaging
* “User acceptance” evaluation for AI products used in medical imaging
* “UTAUT” evaluation for AI products used in medical imaging
* “TAM” evaluation for AI products used in medical imaging

## Study records

### Data management

There are many reasons why it is important to have accurate and high-quality data within an organisation. First, high-quality data is necessary for making sound decisions. If the data is inaccurate, the decisions made based on that data will also be inaccurate. Second, poor data can lead to lost customers and revenue. If customers cannot trust the data that an organisation provides, they will take their business elsewhere. Third, bad data can damage an organisation's reputation. If an organisation is known for having inaccurate data, it will be difficult to gain the trust of potential customers and partners. Finally, inaccurate data can result in legal problems. If an organisation is sued because of inaccurate data, it could be required to pay damages.

1. We will have a Google spreadsheet to store the scraped data from research papers, making sure the link to the data source is next to the data. Our reason for this is so we have a centralised database which will help our research find all the data, increasing their time and making data analysis and efficient process,
2. We will designate a data steward. Their roles will be as follows: they will standardise all the data that is the database. They will confirm the data quality (ensure accuracy etc.). If we did not have a data steward, we are at risk of having incorrect data which can lead to false findings in our conclusion when discussing the workflow impacts on AI integration.

### Selection process

We are using the PRISMA approach. The PRISMA approach is a systematic way to collect and analyse data. It is an alternative to other methods, like case-control studies or surveys, that can be more time-consuming and may not be as reliable.

* **First** we need to collect all possible studies that we will use for analysis.

1. Design search queries (research question)
2. Identify databases to search
3. Run search queries in the databases and collect results.

* **Second** we screen them to identify which are relevant for our analysis.

1. Define inclusion/exclusion criteria.
2. Exclude based on title.
3. Exclude based on abstract.
4. Exclude based on full text and if it contains the data we need.

After the groups have collected the final studies for our report we will have two reviewers, the head of technology and the head of UX independently select the final articles against the inclusion criteria. The reason for having the two leads of the different teams review is because the one will have a deep insight into the topic whilst the other one has a slight understanding but will ensure there is no bias. If there are discrepancies between the reviewers it will be addressed and resolved in the following meetings. We do these reviews to ensure the selection process can be recreated and give confidence in our team in our final list of studies.

### Data collection process

After selecting our final papers after completing all the search criterias and the review process. We will go through all the studies and collate all the different data tables. We must go through various steps before collating that data and using it for our research. We do this to confirm the data quality and data believability.

1. First we identify data from our chosen study
2. Afterwards, we ensure that this data was produced by the authors of the paper as we want to have numerous unique datasets relating to UX, so by ensuring the data was produced from the researchers will help with our data quality and analysis as all the data sets will be unique.
3. Once we have identified the data table and checked it is a primary data source we will manually scrap the data form the final papers in a Google spreadsheet to store the scraped data from research papers, making sure the link to the data source is next to the data. Our reason for this is so we have a centralised database which will help our research find all the data, increasing their time and making data analysis and efficient process,
   1. We will repeat the 3rd step with a second team member extracting the data. This is done to ensure that the data is correct and the steps can be recreated.

## Data items

For the study characteristics, the following data will be extracted:

* Study ID
* Study title
* Year
* Conflicts of interest
* Study type
* Study size
* Setting
* Perspective
* Intervention
* Comparison
* Evaluation

For extracting data for the thematic analysis, there are no planned assumptions about the data or simplifications. However, if assumptions or simplifications are made during the thematic analysis, these will be clearly documented. All data will be extracted to a spreadsheet and will likely include the following items.

* Study ID
* Question
* Response
* Study analysis
* Thematic analysis code

## Outcomes and prioritisation

The highest priority outcome for this thematic analysis is to answer the questions defined in the objectives section to answer the overall question of, “how exactly does the product impact the workflow?” It is anticipated that the themes identified through the analysis will reveal the factors that impact the workflow and their level severity on the workflow. This means the most influential elements are identified that must be heavily taken into consideration when implementing AI into radiology practices. So therefore, once identified, solutions can be made to aid in creating a better user experience for the radiologists.

## Risk of bias in individual studies

Factors that will be considered when assessing the risk of bias in individual studies are:

* Assessing if the paper is funded by a person or company who has an interest in a positive study of user experience.
* How participants were selected for the studies.
* Robustness of the study: reading the process in how the data study was created, including seeing if the results would be similar if the study was to be recreated.

## Data synthesis

User experience is defined as “user’s perceptions and responses that result from the use and/or anticipated use of a system, product, or service” (ISO, 2019). However, methods to study and evaluate user experience vary greatly, for example, questionnaires, interviews, literature reviews, and benchmark testing (Bitkina et al., 2020). Therefore, it is anticipated that it will not be possible to quantitatively synthesise the data. The data extracted from the included studies will be qualitatively synthesised using thematic analysis. Qualitative analysis will be the most appropriate option because it will be able to reveal insights that may have been missed in quantitative analysis. Further, it has been suggested that when making decisions about health systems, the most important role for qualitative evidence is representing views and experiences (Lewin & Glenton, 2018).

The thematic analysis will primarily be deductive and use an iterative process. The Unified Theory of Acceptance and Use of Technology (UTAUT) model will be used as the conceptual framework (Venkatesh et al., 2003). The UTAUT model was developed from analysing eight different models of acceptance and was thoroughly tested. The model defines performance expectancy, effort expectancy, social influence, facilitating conditions, and behavioural intention as determinants of use behaviour. While this model is an acceptance model, it is anticipated that it will provide a strong foundation for explaining how AI impacts on the workflow of radiologists. This is because this question will be answered from a user experience perspective, and determinants of user acceptance are likely to reveal factors which impact on the workflow. However, this thematic analysis is not restricted to being deductive which means if the UTAUT model is found to be unsuitable, new codes and themes will be inductively created. 

The UTAUT model will be used to create a preliminary codebook which is used as a data management tool (Crabtree & Miller, 1992). The constructs of the UTAUT model will be the main themes, and codes (sub-themes) will be created based on them. When coding the studies, if some data does not fit any of the codes, a new code will be inductively created and is not required to fit in the conceptual framework. Therefore, the codebook is intended to be refined and modified during the thematic analysis process.

Group members will familiarise themselves with the data by reading through the studies and extracting the study characteristics and data to a spreadsheet.

All group members will initially code each of the studies independently. Then they will discuss and reflect on each coding, with particular discussion about any inductively created codes and different choices for codes. This step will likely require multiple discussions and reviewing the data to reach a consensus for each code.

Once the group is satisfied with the codes, the next step is to search for themes. Since it is a primarily deductive analysis, the themes are already defined, however, for inductively created codes, they will be reviewed to determine whether they fit into an existing theme or a new theme needs to be created. Through group discussion and reflection, these themes will be reviewed and refined to ensure they correctly represent their codes.

Finally, the written analysis will include a discussion of each theme and what they reveal about the user experience of AI in radiology and how it impacts on the workflow.

## Meta-biases

Meta-biases, such as publication bias and outcome reporting bias, are of concern because they may affect the analysis and conclusions drawn. Due to time constraints it is not possible to search grey literature or translate studies into English, therefore, it is accepted that publication bias may be present because only published studies in English are included. For outcome reporting bias, the outcomes in the methods and results sections will be compared. If outcomes are missing without explanation, the study will be excluded due to outcome reporting bias.

## Confidence in cumulative evidence

The study characteristics of all the studies will be reviewed to assess the overall quality of the included studies and identify any biases present.

## References

Bitkina, O., Kim, H., & Park, J. (2020). Usability and user experience of medical devices: An overview of the current state, analysis methodologies, and future challenges. *International Journal Of Industrial Ergonomics*, *76*, 102932. <https://doi.org/10.1016/j.ergon.2020.102932>

Crabtree, B., & Miller, W. (1992). A template approach to text analysis: Developing and using codebooks. In B. Crabtree & W. Miller, *Doing Qualitative Research* (pp. 93-109). Sage Publications. Retrieved 24 September 2022, from.

International Organization for Standardization [ISO] (2019). *Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems* (ISO 9241-210). Retrieved from: <https://www.iso.org/standard/77520.html>

Lewin, S., & Glenton, C. (2018). Are we entering a new era for qualitative research? Using qualitative evidence to support guidance and guideline development by the World Health Organization. *International Journal For Equity In Health*, 17(1). <https://doi.org/10.1186/s12939-018-0841-x>

Venkatesh, Morris, Davis, & Davis. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, 27(3), 425. <https://doi.org/10.2307/30036540>