



Visualizing Future Breast Cancer Prognosis - Integrating AI Predictions into Screening Programs

Master's Thesis in Computer Science and Engineering

CLARA ENSTRÖM
MATILDA ÅBOGE ERIKSSON

MASTER'S THESIS 2024

**Visualizing Future Breast Cancer
Prognosis - Integrating AI Predictions
into Screening Programs**

CLARA ENSTRÖM & MATILDA ÅBOGE ERIKSSON



UNIVERSITY OF
GOTHENBURG



CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Computer Science and Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Gothenburg, Sweden 2024

Visualizing Future Breast Cancer Prognosis - Integrating AI Predictions into
Screening Programs

CLARA ENSTRÖM & MATILDA ÅBOGE ERIKSSON

© CLARA ENSTRÖM & MATILDA ÅBOGE ERIKSSON, 2024.

Supervisor: Josef Wideström, Department of Computer Science and Engineering

Advisor: Anna Tidstam, Aileen Health

Examiner: Morten Fjeld, Department of Computer Science and Engineering

Master's Thesis 2024

Department of Computer Science and Engineering

Chalmers University of Technology and University of Gothenburg

SE-412 96 Gothenburg

Telephone +46 31 772 1000

Gothenburg, Sweden 2024

ABSTRACT

Sweden has a long tradition of mammography breast screening programs. AI's integration into healthcare is rapidly expanding. Those AI solutions are mainly aimed at replacing one of the two radiologists who are required to read a mammogram. However, when using the AI solutions available, current breast screening programs have limited ability to prevent fast-growing tumors from becoming incurable. This is especially important in the significant portion of breast cancer cases detected between screenings, called interval cancer. The rate of interval cancer is higher among young women, women with a family history of cancer, as well as black women. In order to address this gap, the TechMed AI startup Aileen Health is developing an AI system to predict interval cancer occurrences, aiming to shift healthcare focus from diagnosing breast cancer to preventing tumors from becoming incurable. This thesis examines radiologists' interaction with the AI system in order to design an AI solution that builds trust and motivation. More specifically, the thesis aims to answer the research question:

What are key factors for trusting and integrating AI-based predictions into the current mammography screening workflow to support radiologists?

The thesis employs a Research through Design approach to establish requirements and guidelines for designing an AI system applicable to disease progression in medical imaging. It also proposes a visual representation of the AI-generated content through an iterative, collaborative design process. The evaluation identified key factors for establishing trust in AI systems, resulting in 14 requirements and guidelines across four categories: usability, trust, motivation, and technical. The design proposal is now a part of Aileen Health's product demo and landing page and will be used in further product development. The thesis emphasizes that trust in AI systems is influenced by factors beyond design, such as pre-existing attitudes toward AI, and should be considered together with external factors for comprehensive trust establishment. Overall, the thesis provides insights for implementing AI in healthcare systems, focusing on the design of AI-generated content. By doing so, there is a potential for using AI for predicting disease progression, in cancer treatment and beyond.

Keywords: AI, Breast Cancer, Interaction Design, Mammography Screening, Radiology, Research through Design

ACKNOWLEDGMENT

We would like to express our gratitude to Lars Lindsköld for his guidance and support throughout the project. His extensive network of contacts and generosity have opened doors and enabled valuable collaborations. We also want to give a special thanks to Anna Tidstam, our supervisor at Aileen Health, for guiding and supporting us through the project. Her presence and commitment have been a great asset to us. We would also like to thank Josef Wideström, our academic supervisor from Chalmers University of Technology, for his valuable input throughout this thesis. Furthermore, we would like to express our gratitude to AI Sweden for giving us the opportunity to participate in their Master Thesis Talent program and for offering support and guidance throughout the thesis process. Lastly, we would also like to thank Unilabs for participating in our study. Their cooperation has been essential to the successful completion of our research.

CONTENTS

1. Introduction	1
1.1 Research Question	2
1.2 Aim	2
1.3 Limitations	3
1.4 Stakeholders	3
1.5 Interaction Design	4
2. Background	5
2.1 Breast Cancer	5
2.2 Mammography Screening	5
2.3 AI in Healthcare	6
2.4 Related Work	8
2.4.1 Transpara	9
2.4.2 Vara	9
2.4.3 Lunit INSIGHT MMG	10
2.4.4 ProFound AI Risk	11
3. Theory	13
3.1 Human-machine System	13
3.1.1 Emotions in Human-product Interaction	14
3.1.2 Motivation	17
3.1.3 Human in the Control Loop	18
3.1.4 Mental Model	19
3.2 Designing Visual Representations	19
3.2.1 Information Processing	20
3.2.2 Usability	20
3.2.3 Typography	21
3.2.4 Color Theory	22
3.2.5 Symbols	22
3.2.6 Landing Page	22
3.2.7 Labeling AI-generated Content	23
4. Methodology	25
4.1 Project Approach	25
4.2 Iterative Design Process	26
4.2.1 Empathize	27
4.2.2 Define	27
4.2.3 Ideate	28
4.2.4 Prototype	28
4.2.5 Test	28
5. Planning	31
5.1 Time Plan	31
5.2 Ethical Issues	32
5.2.1 Transparency, Responsibility, and Biases	32
5.2.2 Patient Experience	33

6. Process	35
6.1 Empathize Phase	35
6.1.1 Literature Review	35
6.1.2 Observations at Unilabs	36
6.1.3 The Existing Product Demo of Aileen Health	37
6.1.4 Initial Usability Testing.....	40
6.2 Define Phase.....	45
6.3 Ideation Phase	47
6.3.1 Benchmarking	47
6.3.2 Brainstorming	47
6.4 Prototyping Phase	48
6.4.1 Future Multiview	48
6.4.2 Future AI Overview	51
6.4.3 Onboarding Process.....	56
6.5 Test Phase	58
6.5.1 Usability Testing with the Radiologist	58
6.5.2 Analysis of Gathered Data	63
7. Results	65
7.1 Requirements and Guidelines	65
7.1.1 Usability	65
7.1.2 Trust	66
7.1.3 Motivation.....	67
7.1.4 Technical.....	67
7.2 Design Proposal	67
7.2.1 Future Multiview	68
7.2.2 Future AI Overview	70
7.2.3 Landing Page	73
8. Discussion.....	81
8.1 Result	81
8.1.1 Requirements	81
8.1.2 Design Proposal	81
8.2 Process and Execution	84
8.3 Validity and Generalization of the Results.....	86
8.4 Ethical Considerations.....	86
8.5 Future Work	87
9. Conclusion	89
References	91
Appendix A	97
Appendix B	101
Appendix C	105
Appendix D	107

GLOSSARY

AI prediction - The process of using artificial intelligence algorithms to forecast or estimate future outcomes based on historical data and patterns

Artificial intelligence, AI - Technology that allows computers and machines to simulate human intelligence and problem-solving abilities

Breast biopsy - A diagnostic procedure for breast cancer using a tissue sample from the breast

CC view - The craniocaudal (CC) view provides a top-to-bottom perspective of the breast and is one of the two standard mammography screening views

DICOM - Digital Imaging and Communications in Medicine (DICOM), the international standard format for transmitting, storing, and sharing medical images and associated data in healthcare

Interval breast cancer - Cancer detected after a routine screening mammogram but before the next scheduled mammogram

Mammography - A medical imaging technique used primarily for the early detection of breast cancer. It involves taking X-ray images of the breast tissue to detect abnormalities such as tumors or cysts

Mammogram - The X-ray picture made during mammography

MLO view - The mediolateral oblique (MLO) view provides a side view of the breast and is one of the two standard mammography screening views

Orthanc - An open-source DICOM server designed for healthcare and medical research

Radiologist - A doctor who specializes in diagnosing and treating diseases and injuries using medical imaging techniques such as X-rays

Research through Design - An approach that leverages the strength of design as a reflective method, continually reinterpreting and reframing a situation by creating and evaluating artifacts that serve as proposed solutions

1. INTRODUCTION

Sweden has a long tradition of organizing breast screening programs with mammography for detecting breast cancer. Mammography is the most efficient method for detecting breast cancer but needs to be followed by a biopsy in order to be diagnosed. Other imaging methods are MRI and ultrasound, which are used as supplementary imaging modalities. All women in Sweden between the ages of 40 and 74 are invited to a mammography screening approximately every two years (1177, 2022). Therefore, there is a two-year interval between screenings, during which breast cancer may potentially develop. The breast cancers detected at screening exams are called *screen-detected* breast cancer, and symptomatic breast cancers occurring between screenings are called *interval cancer* (Socialstyrelsen, 2022). In Sweden, a substantial proportion, around 30 percent, of breast cancer cases arise between scheduled screenings. Often, these interval cancers are diagnosed at more advanced stages, potentially leading to a worse prognosis. The rate of interval cancer is higher among young women and women with a family history of cancer (Jatoi et al., 2022). Furthermore, the effect is not evenly distributed among all races, as black women are more commonly affected by interval cancers. Addressing interval cancers is, therefore, not only about improving the quality of healthcare but also about promoting health equity.

The integration of artificial intelligence (AI) within healthcare is rapidly growing, marking a significant shift in how medical expertise is utilized (Yu et al., 2018). Recent improvements have broadened the scope of AI applications, extending beyond traditional domains once solely managed by human experts. As a result, AI has advanced clinical diagnostics and decision-making across various medical domains, including radiology and breast cancer research and diagnostics. For instance, some AI models are trained to detect suspicious areas in mammograms, indicating potential breast cancer (Lunit, n.d). Others focus on predicting short-term risks of breast cancer development (iCAD, n.d). However, the majority of currently available AI models are trained either on detecting the location of cancer or its risk for appearing in time, never both location and time.

The TechMed AI startup Aileen Health is working on building a system that predicts breast cancer. The objective is to create an AI that generates visualizations of how and when breast cancer is predicted to appear. Knowing when and where breast cancer occurs can move the healthcare system from treating existing diseases to being able to take preventive actions, e.g., reducing the time interval until the next screening exam or treating women with other preventative actions (Regionala cancercentrum, n.d). With awareness of the risks, the women wanting to know their risks are motivated to conduct regular self-examinations and detect interval cancer earlier.

For a successful implementation of the AI system, specialist doctors in radiology, commonly referred to as radiologists, must have confidence in the system's capabilities and be motivated to adjust their established workflows. To achieve this, it is necessary to understand and analyze the factors that influence how these professionals interact with the AI system. Therefore, the forthcoming master's thesis aims to address this challenge

by analyzing and evaluating the interaction between radiologists and AI systems, focusing on how AI-generated visualizations should be presented. The master's thesis will be conducted in collaboration with Aileen Health and will explore radiologists' experiences and reactions to using Aileen Health's AI system in their daily practice through observation and interviews. By delving into these interactions, key factors influencing the creation of trust and motivation among users will be identified.

Based on the identified factors, the visual representation of the AI system will be enhanced. The objective is to create a visual presentation that effectively communicates the AI system's insights and results while also establishing credibility and trust for radiologists. The focus here is on the interaction between the different users and the system. This requires consideration of user needs and ethical concerns to create a technically accurate, fair, and reliable presentation. Bias in data and algorithms, responsibility, and integrity are all critical issues that must be taken seriously and addressed proactively to ensure an ethical and trustworthy implementation of AI in radiology.

1.1 Research Question

AI-driven predictions offer a significant opportunity to detect interval cancer earlier, thereby preventing cancers from becoming incurable. However, successful integration depends on factors like trust in AI predictions and seamless workflow integration. This research explores user and system interaction, aiming to provide insights into essential elements for successful AI integration. Consequently, the following research question was formulated:

What are key factors for trusting and integrating AI-based predictions into the current mammography screening workflow to support radiologists?

1.2 Aim

The upcoming thesis aims to design a visual representation of the content generated by an AI model intended for integration into an existing healthcare system. A significant focus will be investigating factors influencing healthcare professionals' trust and motivation to implement the AI system into their current workflow. By encouraging radiologists to use the AI system, the focus of healthcare can transition from treating to preventing. This means not only treating existing illnesses but also proactively working to prevent them through early detection. The AI system will also serve as decision support, aiding radiologists in making recall decisions for further investigation, thereby reducing false recalls and unnecessary patient anxiety.

Through a literature review and evaluations involving radiologists, requirements, and guidelines for visual representations of AI-generated content will be developed. The aim is for these specifications to apply to similar projects within healthcare, particularly in image processing. Additionally, the project aims to inspire other areas within healthcare and

countries beyond Sweden to improve the comprehensibility of AI-generated information, promoting more extensive utilization and potentially saving more resources and lives. However, it is essential to be aware of significant differences in healthcare systems worldwide, as each country has its own laws, regulations, and available resources, which can affect both the purpose and acceptance of AI systems.

Finally, a design proposal for the visual representation will be developed in an iterative and collaborative design process. The proposal will demonstrate how the identified requirements and guidelines can be applied and implemented to create a visual representation of the AI-generated data. The aim is to test and evaluate the design proposal with stakeholders and users, resulting in insights regarding trust in this particular type of AI-supported system.

1.3 Limitations

In the thesis, a limitation concerns the willingness and availability of radiologists to participate. As of today, Sweden has only 71 breast radiologists available, leaving a shortfall of 55 specialists to fulfill the needs within the healthcare system (Cardona Cervantes, 2024). The willingness relates to their prior experience of AI, whether they feel confident enough to give feedback on an AI solution, and their perceived threat of being replaced. Additionally, radiologists' busy schedules and heavy workloads contribute to this limitation, making it challenging for them to prioritize participation in a student's thesis project. Consequently, the study will be constrained by the number of radiologists available for participation and the time they can allocate for interviews and observations. Furthermore, Aileen Health's AI solution is under development and, therefore, cannot be tested clinically. Instead, a product demo is used to get feedback from radiologists.

1.4 Stakeholders

The project involves multiple stakeholders, including the TechMed AI startup Aileen Health and its CEO, Anna Tidstam. Aileen Health was founded in April 2024 and is developing an AI model that predicts the occurrence of interval cancer by generating a future prognosis. The visualizations of these predictions are the focus of the forthcoming master thesis. Anna Tidstam is the primary expert in the system's AI technology, with more than 15 years of expertise in the field of artificial intelligence. She will also act as the company's supervisor, providing guidance and direction to the master's thesis team members. Tidstam primarily works remotely from Austria, which is the main site of Aileen Health.

An additional stakeholder is AI Sweden, the Swedish national center for applied artificial intelligence. The organization promotes and advances the development and use of AI in Sweden and serves as a collaborative platform where various stakeholders connect to exchange expertise, resources, and best practices related to AI. The upcoming thesis is part of AI Sweden's Master Thesis Talent program, offering support and guidance throughout

the thesis process. As a member and partner of AI Sweden, Västra Götalandsregionen (VGR) will also have a substantial role in this project, mainly due to their responsibility for healthcare services within VGR, such as mammography screenings. Furthermore, since breast radiologists will be the primary users of the product, they are also classified as stakeholders in the project. As the ultimate users, their opinions must be carefully considered since their attitude plays a significant role in determining the adoption and usage of the product.

1.5 Interaction Design

Several factors must be considered when implementing AI into an already existing healthcare system to facilitate a seamless implementation. During the implementation, it is crucial to guarantee a high level of usability of the AI model to ensure the usage of the system. Radiologists should find it easy to interact with and understand the recommendations provided. Low usability may result in heightened mental workload, stress, or frustration, which could cause a non-utilization of the system and a significant loss of resources (Norman, 2013). The field of human-computer interaction, also called interaction design, is therefore highly relevant to implementing AI in healthcare systems efficiently, effectively, and satisfactorily. It also ensures that the system is customized for the specific requirements of the user.

When adding an AI model to a current system, the designer's role is to ensure the system is user-friendly, motivating the radiologists to use it and add it to their workflow. The system must guarantee that it does not cause the radiologists any more significant mental workload, stress, or frustration while fostering trust and satisfaction to accomplish this goal. This can be achieved by ensuring that the system is easily discoverable, that users clearly understand existing features and locations, and how the product is intended to be used (Norman, 2013). Secondly, the information needs to be presented and visualized in a way that is easy to understand and interpret to provide efficient and effective knowledge transfer (Sharp et al., 2019).

2. BACKGROUND

This chapter presents background information on the project and the current state of knowledge in AI within healthcare and breast cancer. It includes details about breast cancer, mammography, and the current applications of AI in healthcare.

2.1 Breast Cancer

Breast cancer refers to developing a cancerous tumor in the breast (1177, 2023). There are different types of breast cancer, each requiring various treatments. However, when detected early, breast cancer is generally highly treatable. Each year in Sweden, 9,000 women receive a breast cancer diagnosis, equating to more than one woman every hour. About 1,400 of these women lose their lives to the disease (Karma, n.d.). However, the number of women being diagnosed is increasing. Why a person gets cancer can never be entirely sure (1177, 2023). However, several factors can raise the risk of breast cancer; some can be avoided, while others are beyond the individual's control. Increasing age is an example of such a factor. A breast lump or thickening in the breast or armpit that feels different from the surrounding tissue is a common symptom of breast cancer. This is particularly true if the lump does not disappear in a few weeks and the individual is above 40.

If breast cancer is detected, treatment varies depending on the type of breast cancer and how far it has progressed. A standard procedure is surgery, where part or all of the breast is removed. After surgery, it is also common to receive radiotherapy and chemotherapy to reduce the risk of the tumor returning or spreading. Another standard treatment is hormone therapy. However, this is not suitable for all subtypes of breast cancer. Many people affected by breast cancer can continue their life as it was before the diagnosis, with no immediate physical or practical barriers (1177, 2023). However, some people may feel vulnerable and anxious for quite some time after the treatment is finished. The disease and treatments can also affect an individual's self-esteem and sexuality.

Breast cancer detected after a negative routine screening mammogram but before the next scheduled mammogram is called *interval cancer* (Hovda, 2022). In Sweden, interval cancers account for about 30 percent of cancers detected among screening participants (Socialstyrelsen, 2022). Some interval cancers are not visible initially but grow rapidly, becoming detectable through symptoms such as a lump in the breast. Additionally, some interval cancers are also unintentionally discovered during other radiologic examinations for non-breast-related symptoms (Hovda, 2022).

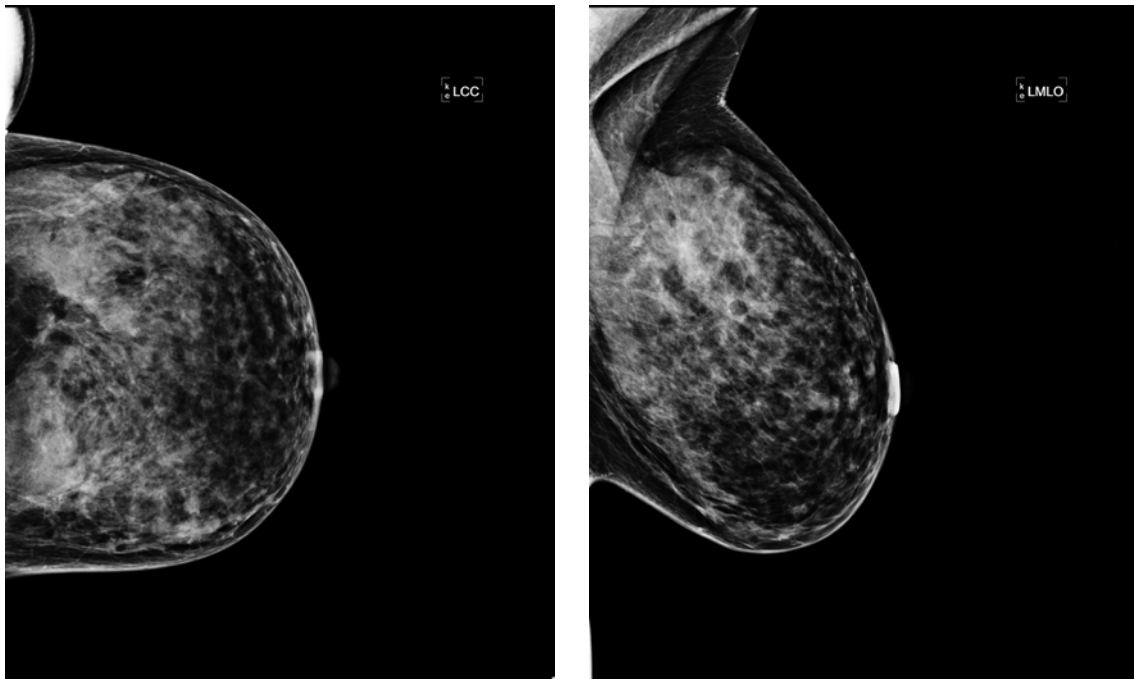
2.2 Mammography Screening

Mammography screening is a type of examination that is done to detect breast cancer before the woman has experienced any symptoms. In this context, screening means a systematic examination of a part of the population to identify women with potential risk

for breast cancer (1177, 2022). Screening imaging consists of four X-ray images of the right and left breast in both mediolateral oblique view (MLO) and craniocaudal view (CC) (Mohamed et al., 2018). MLO and CC are two standard mammographic views, as seen in Figure 1.

Figure 1

Pictures showing a breast in CC view to the left and MLO view to the right



After the examination, two radiologists examine the mammograms independently, a procedure called *reading*. The mammograms are also compared with prior images, if there are any, to make the examination more certain. Patients might need to re-do the examination for further investigation of the breast or abnormalities, also known as *recall*. Further investigation with other methods, such as biopsy or ultrasound, might also be needed (1177, 2022).

The advantage of mammography screenings is that the procedure can detect early signs of breast cancer before any symptoms are showing and, therefore, prevent the disease from spreading and becoming more aggressive. However, there are still a few disadvantages to the procedure. There is a risk that the X-ray does not detect the breast cancer, leaving the patients with a false negative. An explanation can be due to dense breast tissue. Additionally, there is also a risk that patients who do not require cancer treatment receive it since mammography and other tests of the cancer tumor are not able to tell if the tumor will be dangerous for the patient in the future (1177, 2022).

2.3 AI in Healthcare

In the past years, the implementation and growth of AI technology have been under significant growth. AI is currently being applied to several places within the healthcare

system. There are ongoing studies about how AI can be implemented in radiology. Previous studies show that AI can be applied without interfering with the radiologist's work and increase their performance. For instance, in a study by Rodríguez-Ruiz et al. (2019), they examined radiologists' performance in breast cancer detection independently and compared it with an AI system. The study measured the radiologist's performance using several factors, such as reading time. The study's result indicated that the radiologists supported by the AI system had a higher performance rate than those without the AI support system. However, there was no significant difference in reading time when the radiologists used the AI tool. The study's results indicated that AI could be relevant to use within the healthcare system and that the AI tool had a more significant positive effect on less-experienced radiologists. However, the authors argue that this can be due to two reasons. Firstly, more experienced radiologists have a higher capability of detecting breast cancer than less experienced, leading to a not-as-significant change in their performance with an AI support tool. Secondly, more experienced radiologists are less prone to adapt to new techniques and tools that can affect their performance. The author also argued that the learnability and learning time of the system could be a reason why there was no difference in reading time. Reading time might be reduced when radiologists are more familiar with the system (Rodríguez-Ruiz et al., 2019).

Another study by de Vries et al. (2022) showed the radiologists in the UK's perspective of implementing AI into breast cancer screenings. It examined their attitude toward incorporating AI into mammography screenings and how the AI information should be presented through a questionnaire. The participants were also asked what evidence they would need to convince them to introduce AI into the workplace. The questionnaire presented four scenarios for incorporating AI into the radiologist's workflow; partial replacement, total replacement, a triage scenario where a radiologist would only examine patients where the AI predicts there is a high risk, and a companion scenario, where the radiologists have access to the AI as support.

The results showed that the majority had a positive attitude towards implementing AI into breast cancer screenings, although only approximately one-third of participants understood how AI works. Participants expressed that AI would be beneficial to help with healthcare staff shortages. Regarding the scenarios, a majority preferred partial replacement and rejected a total replacement. However, concerning the evidence required to convince radiologists to implement AI, national guidelines and studies utilizing nationally representative datasets were especially preferred. Participants also expressed that validation of the system and a user-friendly interface are essential to adopting AI. The results from the ranking of preferred AI representations showed that a majority preferred a region of suspicion superimposed on the image, e.g., graphically showing information. In summary, the study shows that most screen readers would like to incorporate AI into the screenings as an assistive tool or partial replacement (de Vries et al., 2022).

Capio S:t Görans Hospital in Stockholm is the first breast center in the world to replace one of two radiologists with AI (Cardona Cervantes, 2024). Since June 2023, the Hospital has seamlessly transitioned from research to clinical practice, significantly reducing recall

rates and waiting times. The journey began with consultant and breast radiologist Karin Dembrower and her team’s study, ScreenTrust CAD, which screened 55,500 women between April 2021 and June 2022. The results showed that the combination of AI and one radiologist found more cancers and recalled fewer healthy women. Motivated by the promising results, Capio S:t Görän’s Hospital approved the implementation of AI in its daily operations. 7-8 months later, they started clinical practice.

The mentioned studies focus on implementing AI into reading X-rays to assist radiologists when diagnosing mammography images or replacing one or both radiologists during screenings. However, there have not been as many studies investigating the use of AI as a preventative tool before breast cancer becomes noticeable during screenings, aiming to detect early symptoms and reduce severe disease and mortality.

2.4 Related Work

The recent expansion of AI in healthcare has been substantial over the past few years, reshaping operations across multiple medical domains (Yu et al., 2018). AI has advanced to the extent where it is able to serve as a tool for healthcare professionals and possible increase performance rate (Rodríguez-Ruiz et al, 2019). Radiology emerges as one of the prominent fields benefiting from AI advancements. Breast cancer manifests through various signs detectable by breast radiologists, such as masses, calcifications, architectural distortions, and nipple retraction. There are currently two types of AI available for breast cancer: one type is trained to detect the features mentioned, while the other is focused on classifying the risk of getting breast cancer within the short term. Several companies leverage the different types of AI in various capacities for breast cancer detection and risk prediction. The most established ones in this field are summarized in Table 1, providing an overview of the AI applications. These companies were examined to assess how they have managed their AI solutions and interfaces, as well as how they present the information generated by AI.

Table 1

An overview of the different AI models, their training data, and the specific findings they can detect or predict

AI Findings	Transpara	Vara	Lunit Insights MMG	ProFound AI Risk	Aileen Health
Cancer detection	✓	✓	✓	—	✓
Cancer prognosis	—	—	—	✓	✓

2.4.1 Transpara

Founded in 2014, ScreenPoint Medical is working on a product called Transpara that enables clinical improvements for the rapid prognosis of breast cancer (ScreenPoint Medical, n.d). Transpara enhances mammography reading accuracy, assists in interpreting suspicious areas, increases confidence in both standard and suspect cases, and speeds up the reading process for both 2D and 3D mammograms. Utilizing AI technology, Transpara identifies abnormal calcifications and soft tissue lesions and provides an objective region score for assistance with diagnosis. Furthermore, it assesses the overall likelihood of cancer in mammograms, classifying tests on a 10-point scale to enhance diagnosis accuracy. Transpara is notably relevant to the forthcoming thesis as it provides an assessment of cancer risk for patients, presenting this information visually to radiologists, similar to the functionality of Aileen Health’s product. Transpara’s interface is shown in Figure 2.

Figure 2

The interface of the product Transpara



Note: How it works (Photograph). ScreenPoint Medical.
<https://screenpoint-medical.com/exam-score/>

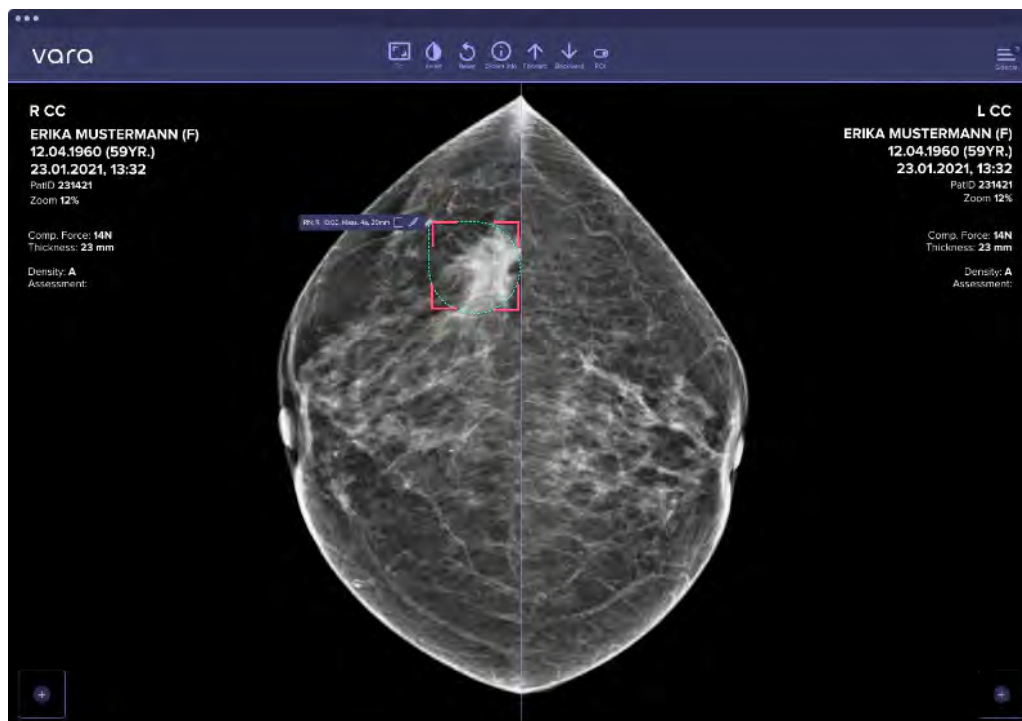
2.4.2 Vara

The German company Vara is working to create safe opportunities for reliable and accessible breast cancer screening (Vara, n.d, a). Their technology aims to deliver superior AI and a superior experience for radiologists in an all-in-one workflow solution. The software previews mammograms, completes structured reports, and reviews previous mammograms for potentially missed signs of cancer. The company works mainly with a decision support system, aiming to combine the strengths of radiologists and AI (Vara, n.d, b). The goal of decision aid is to support the radiologist with AI to improve both sensitivity and specificity. However, AI can be imperfect and not produce 100 percent correct predictions in all cases. Therefore, decision referrals strive to combine radiologists’ human expertise with AI’s technical capabilities.

Vara and the University of Lübeck have initiated a prospective observational study known as The PRAIM Study. This initiative aims to assess the utilization of an AI application and workflow software to assist radiologists in breast cancer screening (Vara, n.d, c). The study directly observes radiologists' interaction with technology, and critical screening-related measurements can be compared between mammograms read with AI assistant and those without it. Vara's product, like Aileen Health's, strives to serve as a decision support tool by combining the intelligence of AI with the expertise of radiologists. Moreover, Vara and the PRAIM study concentrate on understanding the interaction between radiologists and AI systems, drawing parallels to the primary focus of the thesis. Vara's interface is shown in Figure 3.

Figure 3

The interface of Vara's AI system for breast cancer screening



Note: Increasing cancer detection rate (Photograph). Vara.
<https://www.vara.ai/products>

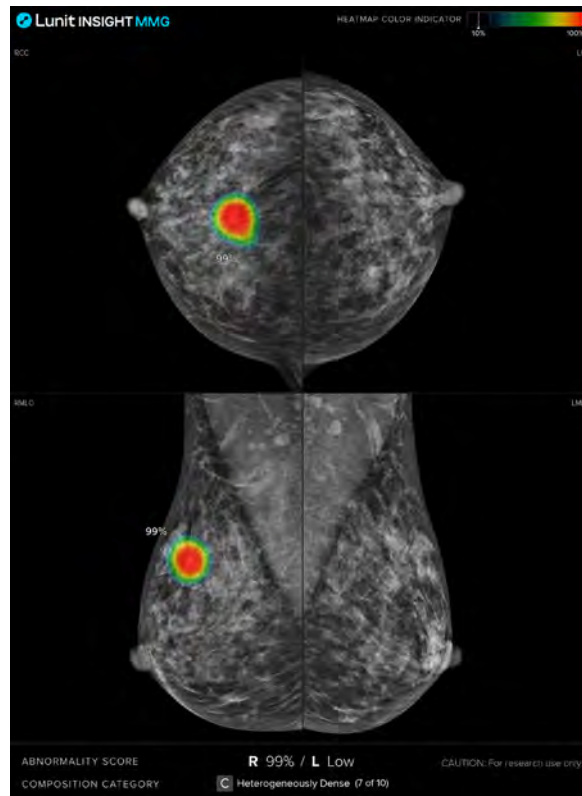
2.4.3 Lunit INSIGHT MMG

Lunit INSIGHT MMG is an AI created by the company Lunit to assist radiologists in interpreting mammograms (Lunit, n.d). The AI detects suspicious areas on the mammograms, including mass, calcification, distortion, and asymmetry. It analyzes and generates the breast cancer location represented as a heatmap or contour map, an abnormality score indicating the likelihood of the presence of breast cancer, and an evaluation of breast density. In alignment with Aileen Health, Lunit's program also

provides AI-generated information to assist radiologists. Therefore, valuable insights can be drawn from their visualization methods for the AI's estimations. Lunit's interface is shown in Figure 4.

Figure 4

The interface of the AI Lunit INSIGHT MMG



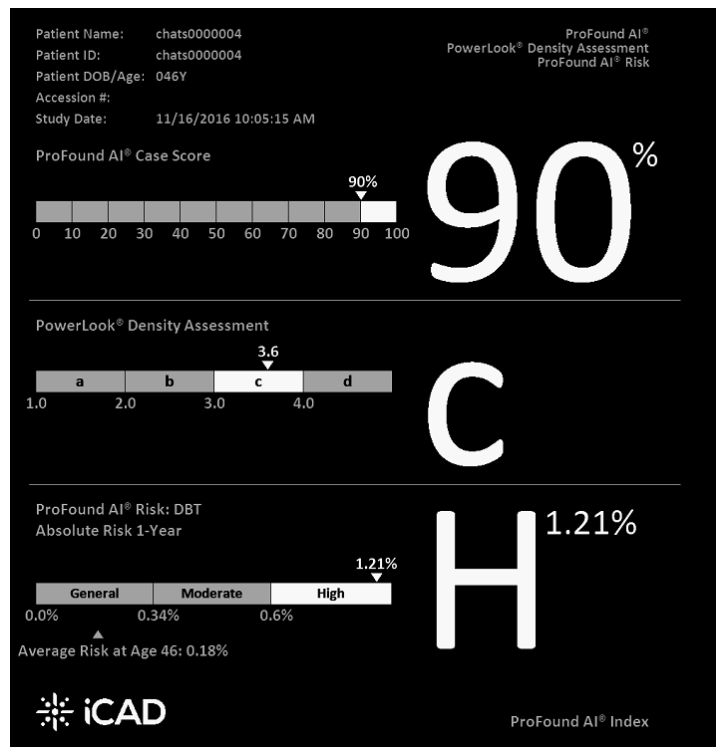
Note: Less Doubts, More Confidence (Photograph). Lunit.
<https://www.lunit.io/en/products/mmg>

2.4.4 ProFound AI Risk

The company iCAD has licensed ProFound AI Risk, an AI to predict the risk of breast cancer during an annual mammography screening (iCAD, n.d). The AI is a clinical decision support tool that evaluates several data points in the image and calculates a short-term risk score for developing breast cancer (Figure 5). Based on the result, it categorizes the risk as low, general, moderate, or high. It can estimate breast cancer risk over one and two years. The AI aims to identify women who may benefit from additional screening or a personalized screening interval. The ProFound AI Risk solution by iCad is relevant to the forthcoming thesis, as it offers estimations of breast cancer risk within a short timeframe. Consequently, this system can provide valuable insights into effective strategies for presenting cancer prognosis predictions for the upcoming thesis. The interface of ProFound AI Risk is shown in Figure 5.

Figure 5

The interface of the ProFound AI Risk



Note: Meet the ProFound Scorecard (Photograph). ICAD.
<https://www.icadmed.com/breast-health/>

3. THEORY

This chapter explains the theories relevant to this project. It contains information on emotions, trust-building in products and AI applications, motivation, data visualization, and design principles.

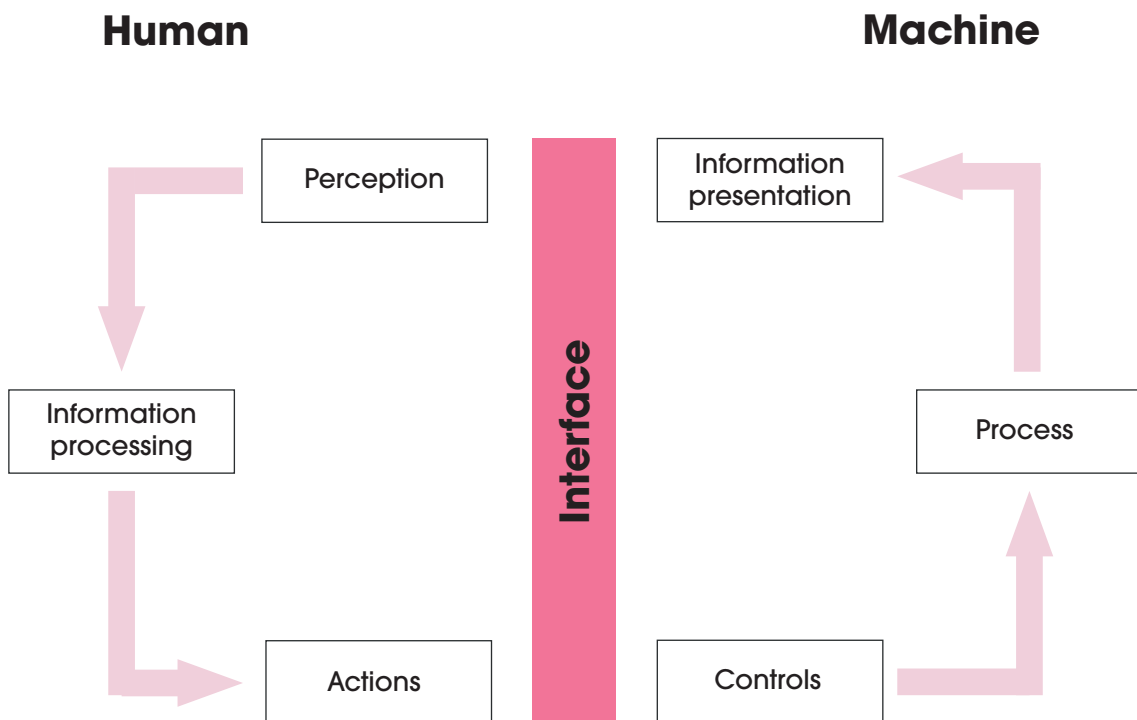
3.1 Human-machine System

A human-machine system consists of humans and machines interacting in a particular environment to complete defined tasks to fulfill the system's goals (Bligård, 2012). In such a system, the interaction is fundamental. The exchange of information happens at the human-machine interface, where humans guide or control the machine's actions. Introducing the user interface within the human-machine system creates a cyclical model where information flows between humans and machines, as seen in Figure 6.

A well-functioning system leads, among other things, to increased competitiveness, reduced risk of accidents, and improved human well-being (Osvalder & Ulfvengren, 2015). The objective is to leverage human capabilities while reducing limitations. Understanding human cognitive conditions, abilities, and limitations in information processing and decision-making is crucial for the optimal functioning of such systems.

Figure 6

A model of a human-machine system



3.1.1 Emotions in Human-product Interaction

According to Plutchik (1984), an emotion is an inferred complex sequence of reactions to a stimulus. He proposed that eight primary emotions are the foundation for all others; anger, fear, sadness, disgust, surprise, anticipation, trust, and joy. Hekkert (2006) believes that emotions are critical to the human-product interaction. To enhance the interaction, it is therefore essential to carefully evaluate and understand the emotional reactions evoked by the product. The experience of a product is influenced by several factors, including how well the senses are satisfied, the overall meaning of the product, and the feelings and emotions it evokes in the user.

Hekkert (2006) emphasizes that emotional reactions occur when evaluating events or situations. This means that the event or situation itself does not necessarily evoke emotions; instead, the interpretation of them gives rise to the feelings. Since emotions act as a motivating force for action, it is crucial to understand the underlying reasons for our preferences for specific sensory impressions. Considering these emotional aspects in design and interaction processes increases the product's usability and creates a more satisfying user experience.

Trust in Products

Building trust is essential in the development of smart consumer products (Michler et al., 2019). Studies have shown that factors such as control, transparency, security and protection, product performance, product handling, brand, onboarding, and information can affect trust. A high level of trust is required when delegating responsibilities to products. If trust does not exist, it can cause a sense of powerlessness among some users. It is, therefore, essential to consider control as a factor and give the user a sense of power over the situation. For example, the product's interface might allow users to change settings and select from various alternatives regarding data gathering (Michler et al., 2019).

Another trust-building factor mentioned is transparency, which refers to data collection, storage, and use. A transparent strategy involves informing customers about what data is being gathered, where it comes from, and its purpose. The product's decision-making process should also be transparent to help consumers anticipate the choices made by the smart product. However, transparency should come at a manageable technical and complexity level. Decisions that are not understandable to the user can significantly decrease trust (Michler et al., 2019).

Security and protection are also essential factors for building trust in smart products. Adequate security measures against unauthorized access and attacks increase user confidence, and the product is perceived as more reliable and trustworthy. It is also essential that the user is informed about the different security measures to address their safety concerns. Furthermore, product performance can also affect the sense of trust, as users tend to consider products with good performance more trustworthy than competing products with poor performance. The product must, therefore, also meet the expectations generated by sales and communication activities and provide added value (Michler et al., 2019).

Another trust-building factor is product handling, which refers to the product's usability and ease of use. A product should prioritize simplicity in its usage, with complexity carefully hidden from the user to minimize negative experiences. Therefore, finding a balance between simplicity and number of features is essential. A suitable number of features can create trust, while an overwhelming array of features or limited functionality can have a negative impact. Furthermore, if the user finds the product and its design appealing, trust can be further reinforced, as such products are often associated with positive characteristics (Michler et al., 2019).

The brand is also a contributing factor to the establishment of trust. A brand can potentially influence a product's perceived quality and performance positively. When consumers trust a brand, it reduces concerns associated with product use, enhancing confidence and likelihood of adoption. The users can depend on their expectations being met and their trust not being violated. Furthermore, onboarding and providing information to new users should be considered crucial in establishing trust in a smart product. Providing information about a product and its usage can decrease uncertainty and perceived risk, ultimately fostering an increase in trust. Supporting users in understanding the product becomes especially crucial when it is highly technical. It may, therefore, be beneficial for manufacturers to consider designing specific smart product onboarding and information activities. This might entail product training, such as a video tutorial, to share knowledge about the product (Michler et al., 2019).

Trust in AI-related Applications

Research in AI has also demonstrated the importance of establishing trust for adopting AI-related applications (Bedué & Fritzsche, 2022). Despite widespread awareness of AI's potential and functional aspects, limited progress is being made in its adoption due to uncertainties over its long-term effects. However, factors such as trust and social influence have been shown to impact the acceptance of AI adoption. Trust becomes particularly crucial when there is a lack of knowledge and prior experience with the innovation. In such cases, individuals lean on the social trust placed in experts or authorities when making judgments either in favor of or against technologies. Also, contextual conditions of implementation, as well as public debate surrounding its use, are essential. Consequently, trust in emerging technologies is sometimes established even before actual experiences with their application (Bedué & Fritzsche, 2022).

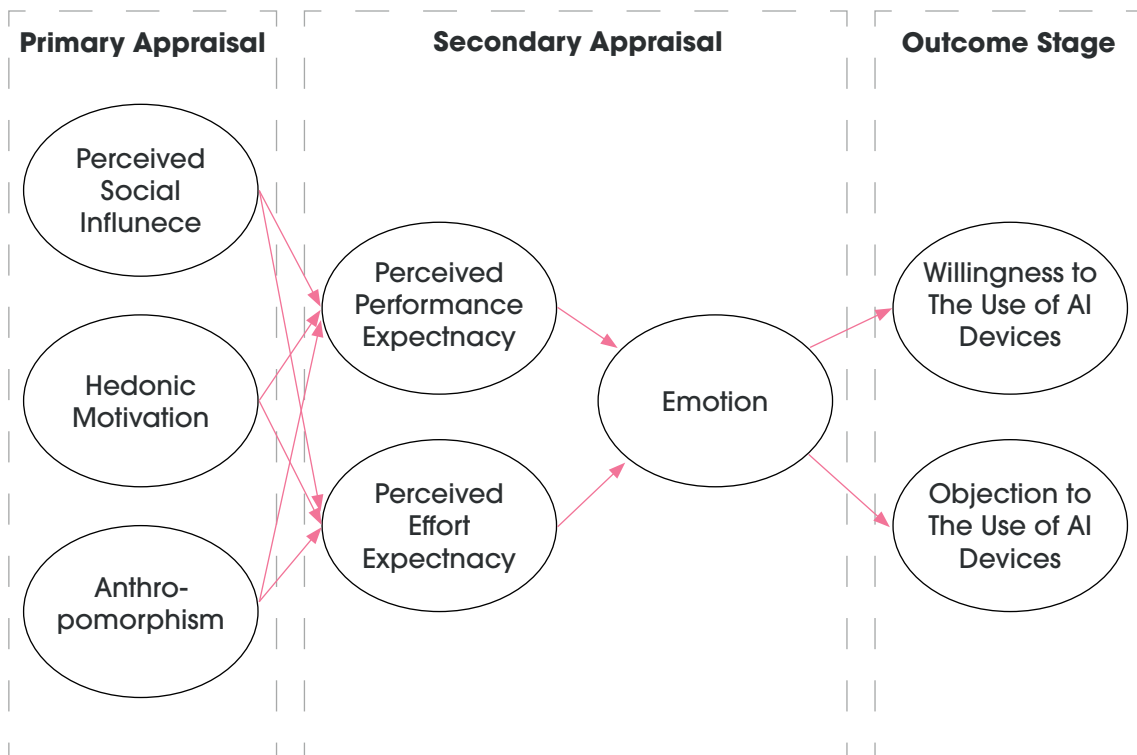
Certain vital factors must be considered to establish trust in AI systems. Firstly, ensuring access to information about AI systems, including their capabilities and limitations, is fundamental for building stakeholder trust. Transparency is another critical aspect. Openly explaining the functionality of AI algorithms and decision-making processes is vital for enhancing trust in AI systems. Furthermore, the factor of explainability must be considered. Providing clear and understandable explanations of algorithmic decisions contributes significantly to the augmentation of trust (Bedué & Fritzsche, 2022). Research on AI recommendations further suggests that information with high estimation precision is more accepted and appreciated (Kim et al., 2021). This indicates, for example, that individuals prefer a precise number, such as 19.42 percent, over a more rounded

figure like 20 percent. More accurate information is associated with a higher quality of communication and content, contributing to the establishment of trust. Trust also drives acceptance and adoption of AI by giving users confidence that utilizing a device will accomplish their intended objectives (Kelly et al., 2023). To fully utilize the potential of a device, individuals must embrace and accept the technology. A lack of approval could lead to reduced user adoption of AI, causing inefficient resource utilization and a potential decline in technological innovation. Therefore, evaluating user acceptance is crucial for stakeholders as it provides insights into the variables necessary to optimize technology adoption in different situations. As a result, various models have been utilized to evaluate users' acceptance of AI (Kelly et al., 2023).

One of these acceptance models is Artificially Intelligent Device Use Acceptance (AIDUA). AIDUA explores the user's willingness to accept using AI devices by studying the experience in three stages; primary appraisal, secondary appraisal, and outcome. These appraisals and outcomes are illustrated in Figure 7. In the primary appraisal, users decide how vital the AI device is by considering things like social influence, pleasure, and human-like qualities. In the secondary appraisal, users think about how well the AI device works and how easy it is to use. This process will affect and result in emotions toward the AI. Finally, based on these considerations, users decide whether to use the technology or prefer human service (Kelly et al., 2023).

Figure 7

The acceptance model, Artificially Intelligent Device Use Acceptance (AIDUA)



Trust in Forecasts

When relying on forecasts as a decision-support tool, one opens oneself up to potential negative outcomes. After all, forecasts are essentially predictions of the future and inherently carry uncertainty. Yet, despite this uncertainty, trust is placed in them. According to Gönül et al. (2012), several key factors contribute to fostering trust in forecasts. Firstly, the intentions of the forecast creator play a crucial part. The perceived competence of the creator significantly influences trust levels. This perceived competence can either enhance or diminish trust, depending on the level of expertise perceived. Moreover, clarity in explaining the forecast is crucial. Providing a transparent and comprehensive explanation of the forecast's components enables users to form their own understanding and trust in its validity.

Additionally, the presentation of the forecast is of great importance. Even if the creator possesses good intentions and high competence, poor presentation can reduce trust. Effective communication of expectations and alignment with user needs is essential for maintaining trustworthiness. Therefore, it is crucial to investigate optimal methods for presenting forecast information to users and to gain insights into their expectations. By addressing these factors, trust can be enhanced in forecasts and maximize their utility as decision-support tools.

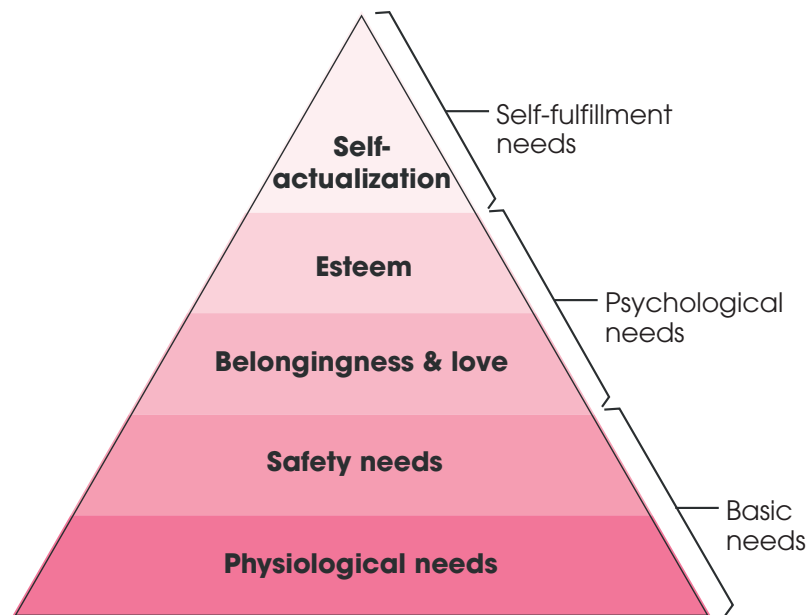
3.1.2 Motivation

Motivation is influenced by conscious and unconscious factors, including the intensity of desire, encouragement, and expectations of the individual and their peers (Ganta, 2014). Most workers require motivation to foster job satisfaction and achieve peak performance. While financial incentives drive some employees, others are motivated by acknowledgment and rewards. Furthermore, the level of motivation in the workplace directly influences employee productivity. Since employees show varied motivations, employers should invest time in understanding each individual's preferences and needs compelling motivation.

Abraham Maslow's hierarchy of needs is an explanatory model in psychology that shows how people prioritize their needs. The model is visualized as a pyramid with five different levels of needs, as seen in Figure 8. According to Maslow, individuals are motivated to fulfill a set of needs, which can be arranged in a hierarchy from survival needs to more creative and intellectual needs. The hierarchy suggests that individuals are motivated by unmet needs, and as lower-level needs are satisfied, higher-level needs become the focus of motivation (Ganta, 2014).

Figure 8

Abraham Maslow's Hierarchy of Needs



3.1.3 Human in the Control Loop

When automating a system or including an AI, questions like who is in control are crucial to consider to ensure a safe implementation. Within a system, there is a control loop, which explains how the system operates and what factors affect its actions. When a system is being moved from entirely manual by the user to partly or fully automated, the control loop of the system alternates. When a system is entirely manual, the user is in the center of the control loop, acting based on the feedback of its actions, leaving the user in complete control (Andersson, 2010). This placement of the user in the loop is also called human-in-the-loop (Rahwan, 2017). When automating a part of the system, the tasks for the user change.

However, if the system is partially automated, the user can still be inside the control loop. When a system is entirely automated, there is a risk that the user falls out of the control loop, creating a black box where they cannot tell how and why the machine acts like it does. This results in a loss of transparency for the user and other stakeholders (Rahwan, 2017). This transparency applies to the algorithm and the training data in AI. Transparency is essential for mainly two reasons. Firstly, if the user is in the control loop, they can identify errors or misbehavior. If the algorithm is performing incorrectly or is not properly trained, it is critical to be aware to correct the errors. If the user is out of the control loop and the algorithm is a black box, it is not possible to correct the mistakes that are happening. Secondly, someone must be held accountable for the AI decisions. If humans are still in the control loop and understand what is happening, trust in the system can be built. If the AI model were doing something wrong, someone would be held accountable for the mistakes and be responsible for the AI model (Rahwan, 2017).

When incorporating AI, it is essential to maintain transparency by ensuring the user remains within the control loop. Even though numerous studies show that AI can be applied to healthcare, obstacles such as the black box still must be tackled to maintain safe AI implementation. Safe implementations remain unattainable without transparency regarding the AI model's data and algorithms. Furthermore, biases within algorithms and training data may arise and remain uncontrollable without transparency. These biases can be caused by human biases that are transferred into the algorithms or from the training data. Therefore, these aspects must be considered when using AI as a decision support system to build trust (Rahwan, 2017).

3.1.4 Mental Model

When interacting with the world around us, mental models are created. Mental models are representations of how individuals experience the world around them. They can be of systems, people, or the surroundings and can be described as small-scale models of reality (Andersson, 2010). These mental models are developed while interacting and are flexible and constantly updated. Mental models do not have to be accurate of the object; they are just a representation of the object we have in our mind (Andersson, 2010). The purpose of mental models is our ability to reason, solve problems, predict the consequences of actions (Andersson, 2010), and understand the object and the world (Norman, 2013).

When designing a system, it is the designer's job to make sure that the system matches the user's mental model for it to work successfully. If the user has higher expectations in their mental model than the system can achieve, dissonance occurs between the mental model and the system. If there is dissonance, the user's trust in the system can decrease. The same applies if the user's mental model has lower expectations of what the system can achieve.

3.2 Designing Visual Representations

Studying data visualization is essential for designing visual representations and interactive elements that help people perform tasks more efficiently (Munzner, 2014). Designing visualizations involves navigating multiple trade-offs, where decisions often entail sacrificing one aspect for another. Therefore, gaining insight into the purpose, content, and methodology of visualization usage is essential. It involves comprehending user needs, analyzing data presentation, and assessing the design approach, all of which contribute to creating meaningful and impactful visualizations.

Furthermore, it is crucial to acknowledge that digital communication carries a risk of ambiguity, where messages can be interpreted in multiple ways, leading to confusion (Cambridge Dictionary, n.d). However, ambiguous communication is sometimes seen as beneficial in design work because it opens up possibilities for interpretation that can stimulate creativity (Hisarciklilar & Boujut, 2009). On the other hand, ambiguous communication can often lead to misunderstandings, which can have serious consequences. To avoid this, clear and precise messages are crucial. This can be achieved by establishing a clear visual hierarchy of information, selecting appropriate typography, and utilizing

colors to evoke specific emotions, among other strategies (Watzman & Re, 2007). However, it is crucial to be aware that factors such as culture can influence color choices and the construction of information hierarchies (Osvalder & Ulfvengren, 2015). For example, different cultures may have different perceptions and associations with specific colors, which can affect how they are interpreted and experienced. The same applies to the construction of information hierarchies, where reading orientations and strategies may vary. Taking these variations into account is necessary to ensure that the design is effective and understandable for the intended audience.

3.2.1 Information Processing

A critical consideration in designing and evaluating human-machine systems is understanding human mental abilities and limitations (Osvalder & Ulfvengren, 2015). While humans possess significant potential to absorb and process various information from the system, it is also essential to recognize and comprehend the obstacles that may arise in the interaction between humans and technology. Understanding both the potential and challenges in the interaction between humans and technology is fundamental for designing a user-friendly interface. For instance, it is essential to have a well-planned visual presentation, incorporating clear cues that guide and support users throughout the process to facilitate user interaction with a product. Various senses enable individuals to perceive information from their surroundings. Vision is the most common of these senses, accounting for about 80 percent of all sensory data obtained. Consequently, designing interfaces with clear structures and patterns proves beneficial, as the visual sense actively seeks such elements. Additionally, integrating high contrast into the interface simplifies the differentiation of lines and shapes on objects, enhancing user perception and interaction. Furthermore, when creating visual information, the design should primarily be in grayscale, and color coding should be limited and used with consideration. Ideally, color coding should only be used as redundancy, that is, to reinforce already established information (Osvalder & Ulfvengren, 2015).

The process of visual information intake can be categorized into searching and scanning. The goal is unknown when searching but is defined for the user during scanning. Two main aspects affect the result when searching for objects; how the user expects the object to look and how eye-catching the object is. It is, therefore, essential to avoid burdening the interface with too much information that looks the same. Furthermore, placing critical information high up and to the left can also be beneficial, as scanning is usually done from top to bottom and from left to right. However, this can vary depending on culture (Osvalder & Ulfvengren, 2015).

3.2.2. Usability

Usability can be seen as a quality attribute that assesses how easy a user interface is to utilize. According to ISO (2018), International Organization for Standardization, usability is the extent to which a system, product, or service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified

context of use. Satisfaction represents the extent to which the user's physical, cognitive, and emotional responses that result from the use of a system, product, or service meet the user's needs and expectations. Effectiveness refers to the accuracy and completeness with which users achieve specified goals, and efficiency refers to the resources used in relation to the results achieved.

Patrick Jordan (1993) expanded the definition of usability by introducing a framework that includes changes in the interface's usability with user experience. The framework consists of five components that aim to describe how users' performance on a task might change with experience:

1. Guessability - The effectiveness, efficiency, and satisfaction with which specified users can complete specified tasks with a particular interface for the first time

2. Learnability - The effectiveness, efficiency, and satisfaction with which specified users can achieve a competent level of performance on specified tasks with an interface, having already completed those tasks once previously

3. Experienced user performance - The effectiveness, efficiency, and satisfaction with which specified experienced users can achieve specified tasks with a particular interface

4. System potential - The optimum level of effectiveness, efficiency, and satisfaction with which it would be possible to complete specified tasks with an interface

5. Re-usability - The effectiveness, efficiency, and satisfaction with which specified users can achieve specified tasks with a particular interface after a comparatively long time away from these tasks

Jordan suggests that the framework should contribute to clarity in defining usability goals from the designer's perspective. When designing for usability, it is important to consider whom the interface is intended to be usable for, and a crucial part of this is the expected level of user experience. For instance, when designing software for well-educated users who frequently utilize the program, prioritizing intuitive design may be less critical.

3.2.3 Typography

Typography is essential for effective digital communication. The text's form should enhance communication rather than hinder the reader's engagement with the content. Hellmark (1994) outlines several factors that influence readability and legibility, with typeface selection being one of them. Typefaces typically fall into two categories: Serif and Sans Serif. Serif fonts are distinguished by the small lines or feet that extend from each letter. In contrast, sans means without, meaning fonts that lack these lines. When readability from a distance is a priority, Sans Serif typefaces offer advantages due to their more straightforward letter construction. They are particularly well-suited for headlines or text presented on screens.

3.2.4 Color Theory

Color is the most strongly emotional element in visual communication, provoking immediate and intense reactions (Watzman & Re, 2007). Strategic use of color enhances information absorption and facilitates differentiation between different types and hierarchies of information. When selecting colors for a product, considering the symbolism associated with each color is essential. Colors, such as red, orange, and yellow, commonly evoke notions of warning or danger and are frequently employed as standard indicators across various contexts, including guidelines established by organizations like the American National Standards Institute (ANSI) and the Society of Automotive Engineers (Braun et al., 1995). In a study conducted by Braun et al. (1995), findings revealed that product labels presented in color were perceived as more hazardous and increased readability compared to those in black and white. The authors underscored in their article the efficacy of color coding in enhancing visual search tasks, emphasizing its advantage over alternative coding strategies such as numeric or shape-based.

3.2.5 Symbols

In a user interface, symbols can function as effective means of communication with users. A symbol is essentially a sign, shape, or object that is used to represent something else (Cambridge Dictionary, n.d). However, for effective communication, symbols must be universally recognized and easily interpreted by users (Osvolder & Ulfvengren, 2015). When chosen appropriately, symbols offer distinct advantages over text. They are visible from greater distances, overcome language barriers, and communicate information quickly and precisely.

Furthermore, presenting information in multiple ways, also referred to as a redundant form of presentation, significantly increases the likelihood of its correct interpretation (Osvolder & Ulfvengren, 2015). Therefore, employing diverse visual presentations, such as combining text with images or integrating both color and shape, enhances clarity.

3.2.6 Landing Page

As previously discussed, establishing a transparent onboarding procedure is crucial for fostering trust when introducing new products (Michler et al., 2019). One way to accomplish this is by creating a landing page. A landing page is the initial webpage a visitor encounters after clicking through from an email, advertisement, or other digital platform (McCormick, 2023). Its primary objective is to provide visitors with comprehensive information to facilitate confident decision-making and stimulate interest in the brand or product. A clear headline on the landing page is crucial to attract a visitor's attention, stimulate interest, and promote further reading (Teodorescu & Vasile, 2015). Additionally, it assures users that they have arrived at the correct destination and that the content is relevant to the link they clicked. Furthermore, to enhance the credibility and trustworthiness of the landing page, it is crucial to leverage social proof, such as reviews from other customers and visitors. This reinforces the sense of shared experience and

fosters confidence in making the correct decision. When creating a landing page, it is also essential to consider colors, as they can affect mood and emotions, steering visitors towards a specific mindset.

Maslow's hierarchy of needs guides the design of landing pages according to visitors' emotional needs (Teodorescu & Vasile, 2015). Basic needs related to survival motivate potential customers when they see product features. As customers read about product features, they may instinctively or consciously associate them with how the product can fulfill their fundamental needs, thus strengthening their motivation to purchase or use it. Trust is crucial for consumers, especially regarding their safety needs. In addition to traditional reviews, landing pages can feature other reliable indicators, such as messages explaining the necessity of providing email addresses and assurances regarding data privacy. To address the need for belongingness, the design and text of the page should foster a sense of community. Furthermore, emphasizing the product's advantages, not just its features, is crucial when addressing esteem needs in the hierarchy of needs. This approach appeals to stakeholders and demonstrates how the product or service improves quality of life. Achieving self-actualization on a landing page involves treating visitors with transparency, respect, and attention to their needs and preferences. When visitors feel valued and understood, they are more likely to trust the brand or product and feel motivated to engage further and ultimately purchase.

3.2.7 Labeling AI-generated Content

The use of AI-generated content is becoming increasingly common across various forums. With its fast-growing pace, it becomes imperative to differentiate between real content and AI-generated content. According to Arechar et al. (2023), the choice of a label to denote AI-generated content depends on the label's goal. For instance, the selection of the label should vary based on whether the objective is to communicate that the content is created by AI or to caution users about potential misinformation. Arechar et al. (2023) demonstrated through their research that the phrase *AI generated* is preferred for indicating content created by AI. Alternatively, *artificial* serves as another viable option. However, the authors emphasize the importance of contextual evaluation of different phrases to ensure their effectiveness.

4. METHODOLOGY

The following chapter describes essential design methodologies crucial to the project's success. It explores an approach to design research and explains the iterative design process. Additionally, the chapter summarizes the specific phases selected for this thesis, offering a comprehensive understanding of the chosen methodology.

4.1 Project Approach

Aileen Health aims to achieve significant changes rather than solely gathering knowledge. The company has pinpointed a notable issue concerning a high occurrence of interval cancers and is actively collaborating with stakeholders to analyze the system and implement suitable changes. Given its emphasis on collective action and enhancement, the project aligns with the principles of action research. Action research is an iterative process in which researchers and practitioners collaborate in several steps (Avison et al., 1999). These steps typically involve interpreting a problem, taking action to address it, and reflecting on the outcomes to learn and improve. This iterative approach allows for continual refinement and adaptation based on real-world experiences and insights gained throughout the process.

Moreover, the master's thesis project also strives to follow a Research through Design (RtD) approach. RtD leverages the strength of design as a reflective method, continually reinterpreting and reframing a situation by creating and critiquing artifacts that serve as proposed solutions (Zimmerman & Forlizzi, 2014). The approach is suitable for research that involves designing technologies that are not yet well-understood, as well as for investigating how people interact with specific designs (Aydın Baytaş, 2022). Four criteria have been established to serve as a structured framework for evaluating and formalizing contributions in interaction design research conducted through the Research through Design approach (Zimmerman, 2007).

Process: Ensures the quality of the research by emphasizing the carefulness and transparency of the methods employed. It aims to enable the reproducibility of the research process and encourages transparency in documenting methodologies.

Invention: Focuses on the innovation and significance of the contribution, requiring researchers to integrate various subject matters uniquely. The criterion encourages a thorough literature review to position the work within the existing research landscape and assess its potential impact.

Relevance: Shifts the focus from traditional scientific validity to real-world relevance. Researchers are encouraged to frame their work within practical contexts and express the preferred state their design aims to achieve.

Extensibility: Ensures that the research outcomes are extendable, allowing the broader community to apply the research process to future design problems or leverage the knowledge created by the resulting artifacts.

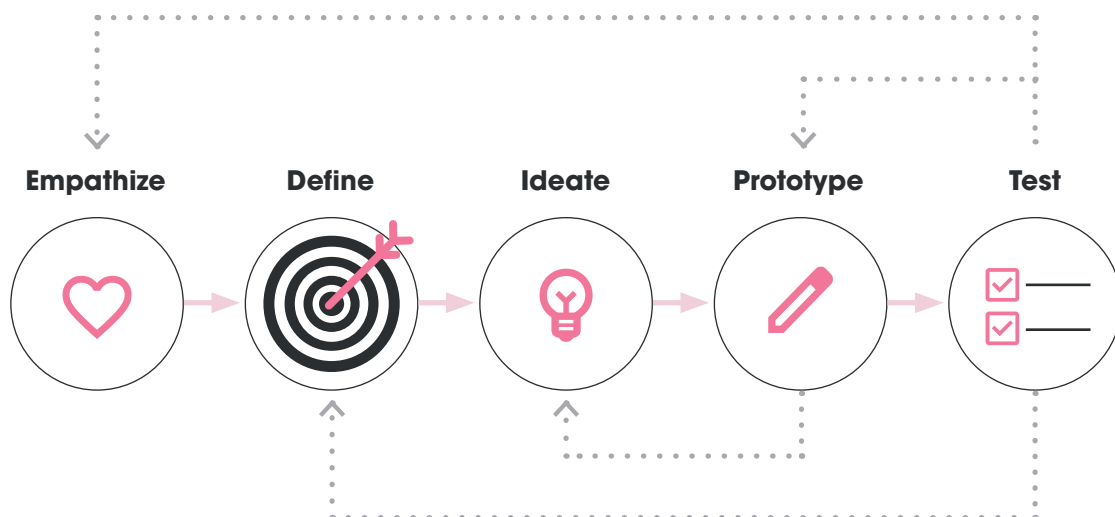
4.2 Iterative Design Process

This thesis will employ an iterative design process characterized by its feedback-driven approach to development. The iterative design process is often cost-effective and efficient since it does not require the full development of a product before receiving user feedback (Interaction Design Foundation, 2023). Instead, prototypes can be generated and assessed with users at various process stages, promoting a user-centered outcome. This iterative approach allows continuous refinement and optimization, contributing to a more responsive and adaptable project development cycle. This approach is suitable for the upcoming thesis, given the focus of evaluating the interaction between AI systems and radiologists. Therefore, adopting a feedback-driven process becomes essential to implement the insights gained from usability tests and iteratively refine the design. Utilizing a linear approach would limit the incorporation of insights from usability tests and hinder the refinement of the prototype, resulting in a less user-centric design.

However, design thinking does not adhere to a singular definition or a fixed process. In this project, the chosen approach is the Design Thinking model, as presented by the Interaction Design Foundation (2016). It is a non-linear, iterative process used by teams to comprehend users, challenge assumptions, redefine problems, and develop innovative solutions via prototyping and testing. Therefore, this methodology is particularly advantageous when dealing with ill-defined or unknown problems, such as in the forthcoming thesis. As seen in Figure 9, the model contains five distinct phases; Empathize, Define, Ideate, Prototype, and Test. However, the stages in design thinking are not always sequential; teams often run them in parallel, out of order, and repeat them as needed for a flexible and adaptive approach.

Figure 9

The Design Thinking model



4.2.1 Empathize

The first phase is called Empathize. During this stage, the aim is to understand the problem, typically through user research (Interaction Design Foundation, 2016). Empathy plays a vital role in design thinking, enabling designers to put aside preconceived assumptions about the world and gain valuable insights into users and their needs. Therefore, conducting a literature review is a valuable method for gathering background information and gaining insights into the current state of knowledge on a particular topic (Bligård, 2015). Another fundamental approach for gathering user information is through interviews. This method aims to gain insights into how users think and reason. There are three different interview techniques: structured, semi-structured, or unstructured. A structured interview involves pre-formulated questions and is preferable if more quantitative data is desired. If the discussion takes place freely around a topic, the interview is considered unstructured and results in more qualitative data. A semi-structured interview combines elements of both, providing a structured question framework while allowing for a more open discussion. The flexible structure of an interview enables a deeper exploration of the user's thoughts by asking follow-up questions to clarify the information further. This also reduces the risk of misinterpreting responses. Consequently, this method is particularly suitable for gathering more subjective insights into feelings and opinions.

Observation is also a commonly used method, allowing investigators to understand the user's situation through firsthand experience by observing, listening, and engaging (Bligård, 2015). The objective is to understand the user's situation without influencing the user's actions. The strategy is primarily used to identify needs and potential areas for improvement by providing insights into users' existing workflow and their interactions with the system. Observations can occur either in the field, representing real-use situations, or in a controlled environment like a lab with carefully arranged experimental conditions.

4.2.2 Define

Define is the second stage of the Design Thinking model. The phase aims to identify the core problems and create problem statements by analyzing the gathered information (Interaction Design Foundation, 2016). The KJ analysis is a suitable method for analyzing a substantial amount of qualitative data and gaining an overarching perspective (Bligård, 2015). The method involves documenting the results from the data collection on Post-its, with each Post-it representing a single data unit. Afterward, the Post-its are grouped based on common themes and assigned corresponding headings. Consequently, the method progresses from a detailed level to a holistic perspective of the collected data, where themes can be identified. This strategy is advantageous because it provides a deeper understanding of the data and facilitates the identification of patterns and themes.

Based on the results, the identified needs can be compiled into a list of requirements. A requirements specification is a suitable tool for effectively describing and communicating project objectives (Bligård, 2015). By creating a requirements specification, one can systematically identify and document what needs to be achieved with the project. This includes establishing which functions, features, and performance criteria are required for

the project to be successful. It is essential to define the what question before moving on to the how (Cooper et al, 2014). The specification contains requirements (R) and guidelines (G) (Bligård, 2015). A proposal must fulfill all specified requirements to be deemed a possible solution. Moreover, the requirements should be objectively assessable through yes-or-no questions: if subjective judgment is required, guidelines are utilized. However, the requirements specification grows continually throughout the development process, with the level of detail gradually expanding since it is only feasible to establish some essential aspects at the initial stages.

4.2.3 Ideate

The third phase is called Ideate. This phase aims to create innovative solutions to the problem statements (Interaction Design Foundation, 2016). Brainstorming is an established technique that fosters creative thinking (Bligård, 2015). This method involves exploring the solution space and generating many ideas. Through joint inspiration, unexplored ideas can emerge, allowing for a more extensive exploration of the solution space. It is beneficial to start from different focus areas to avoid getting stuck in conventional thinking. Another technique for generating ideas is the Crazy 8's exercise. In this activity, participants are tasked with quickly sketching eight different ideas in just eight minutes (Google, n.d). The objective is to promote generating multiple solutions within a limited timeframe. Furthermore, conducting a benchmark could be a valuable approach to exploring how others have solved similar problems. Analyzing competitors provides valuable insights into features, functionalities, user flows, and the overall user experience when engaging with alternative products (Guimaraes e Equipe Aela, 2022).

4.2.4 Prototype

Prototype is the name of the fourth phase in the Design thinking model (Interaction Design Foundation, 2016). The objective is to identify the best solutions by creating different versions of the product. During this experimental phase, various ideas can be explored and investigated. A prototype can be described as a concretization of a concept and can be used to communicate ideas in different phases of the project (Bligård, 2015). Prototypes are defined by their level of fidelity (Hanington & Martin, 2019). Early in the ideation process, low-fidelity prototypes, such as sketches and storyboards, are often used. This type enables quick feedback for iterative changes. During the later stages of the design process, high-fidelity prototypes are frequently utilized. These prototypes are more polished and typically capture the look and feel of the final product, sometimes even containing basic functionality. They prove valuable for collecting feedback from customers and users, who can now assess aesthetics, form, interaction, and usability.

4.2.5 Test

The fifth phase is called Test (Interaction Design Foundation, 2016). During this stage, the prototypes are tested with the actual users to assess their usefulness in addressing the

identified problem. The outcome may reveal new insights, leading to potential refinements to the prototype or a revisit to the Define stage to reassess the problem. Usability testing is a method that allows the investigator to observe an individual's experience with a digital application while performing a specific task through various steps (Hanington & Martin, 2019). The tasks are typically concrete and reflect the specific goals of the end user, often contextualized with scenarios. This method's primary goal is to identify interface elements that may cause frustration or confusion, allowing necessary adjustments to be made before the application is launched.

The Semantic Differential Scale can also be a favorable method for investigating user experience and attitudes (Hanington & Martin, 2019). The tool is used to measure the individual's attitude towards the product. Typically, the scale contains pairs of opposite adjectives, often antonyms, representing contrasting qualities related to the subject of evaluation. Respondents are asked to place a mark on a scale, generally seven points, indicating their assessment of how well each pair of adjectives describes the target. Another technique for gathering information on emotions and subjective feelings is the Hesselgren Emotional Scale. Using this scale, participants can rate the emotions the system evokes (Romice & Uzzell, 2003).

5. PLANNING

This chapter presents the proposed time plan for the project and the ethical considerations that may affect the implementation of the work.

5.1 Time Plan

The thesis will begin with a comprehensive literature study to understand the subject and the current state of knowledge in AI and breast cancer research. Simultaneously, a planning report will be created, including a summary of the literature review, relevant theories, proposed methodology, and ethical considerations. Additionally, initial requirements and guidelines will be formulated based on the literature review findings.

After this, user studies will be conducted to learn more about the target users and their work environment. The existing demo of the system will then be evaluated through usability testing. Given the anticipated constraints on time with the radiologist, significant effort will be dedicated to planning the test to guarantee a reliable and valuable outcome. Post-testing, the data will be compiled and analyzed, formulating additional requirements and guidelines to serve as a foundation for the design proposal. Later on, brainstorming sessions will be conducted based on problem areas identified during the initial usability tests. Prototyping will be carried out using Figma to visualize ideas emerging from brainstorming and facilitate evaluation with users. The prototype will be used in usability testing, and iterative redesigning will occur based on their feedback.

Report writing will continue throughout the project, emphasizing fine-tuning towards the end. However, given the iterative nature of the project, it is essential to acknowledge the potential for significant changes. Consequently, the plan should be viewed as a proposal and initial draft, and it should be open to adjustment as the project evolves. The time plan has been visualized as a Gantt chart in Figure 10.

Figure 10

A Gantt chart of the planned process



Task	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Planning report	Pre-study																			
Literature review		Empathize																		
User studies				Empathize																
Preparation of usability test						Empathize														
Usability testing									Empathize			Test								
Analysis of tests									Define					Define						
Ideation										Ideate							Ideate			
Creation of prototype											Prototype									Prototype
Writing report					Report															

5.2 Ethical Issues

This section discusses ethical issues regarding the implementation of the AI created by Aileen Health in mammography screenings. It involves the technical aspects of implementing AI in healthcare, such as bias in data and algorithms, responsibility, and the patient’s product experience.

5.2.1 Transparency, Responsibility, and Biases

When implementing an AI model or any automatization into a system, there is a risk of creating a black box if the training data and algorithms of the AI model are not kept transparently (Rahwan, 2017). Maintaining the transparency of the training data and algorithms is essential to building trust in the system and preventing future issues. Transparency allows not only the first party to control the AI but also external parties to increase the reliability of the AI model. This is crucial when developing an AI model to ensure safe implementation and development that can be controlled. Responsibility is another critical issue to consider while implementing AI. When no specific person is in charge, the reliance on the system might decrease (Rahwan, 2017). If someone can be held

accountable for the AI decisions and the implementation of the AI model, trust in the system has the right conditions to be built. Unclear responsibility can lead to problems being overlooked when they arise.

Even though AI has the potential to make healthcare more effective, there is still a risk when it comes to choosing the data and algorithms for the AI model that can have prominent consequences. Biases in the AI model can lead to a prominent difference in the performance of care between patient groups, such as ethnic minorities or gender. This can, for example, be misdiagnosis for people of minorities since the AI model has not been trained on data that includes these groups of minorities. If not acknowledged, there is a risk for inequalities within healthcare to increase further (Norori et al., 2021).

Recognizing the importance of ethical considerations is essential. Although technical factors like selecting training data and algorithms are beyond the scope of this thesis, it is crucial to acknowledge them and maintain transparency with users. Therefore, as an interaction designer, it is essential to carefully consider how this information is presented to users.

5.2.2 Patient Experience

An additional ethical consideration concerns the patient's perspective of including AI in the mammography process, as a patient might feel vulnerable while undergoing a mammography examination. However, how patients feel during the process varies a lot. Some might experience emotions such as nervousness or anxiety, and others are not as affected.

For the patients, there are mainly two challenges to consider when implementing the AI. The first challenge is how to present the AI-generated information in the mammography result letter in a trustworthy way to the patient to make sure they still feel comfortable. Patients' attitudes and knowledge about AI can differ significantly and are very individual. Therefore, carefully considering how the data and the other information are presented is crucial to maintaining the patient's trust. Secondly, there is a variation in the patients wanting to know if they are more prone to getting breast cancer in the near future or not. Certain patients may avoid being aware of such knowledge due to the uncertainty associated with AI-generated information and the lack of guarantees. Therefore, patients might be given the option of whether or not to include AI-generated images as a feature, giving them a sense of control.

However, it is crucial to emphasize that this thesis primarily focuses on the radiologist's role in interpreting information rather than focusing on the patient. Therefore, matters about the patient are not within the scope of this master's thesis.

6. PROCESS

This chapter describes the design process in chronological order. It is structured into sections corresponding to the distinct phases; Empathize, Define, Ideate, Prototype, and Test.

6.1 Empathize Phase

This section presents the execution and results of the methods employed to comprehensively understand the issue, the users involved, and their needs.

6.1.1 Literature Review

A literature review was conducted to collect background information on relevant topics, such as mammography screenings and breast cancer research. Additionally, it aimed to gain insights into the current state of knowledge regarding the intersection of AI and healthcare. Google Scholar and previously studied course materials were primarily utilized to source relevant literature. Furthermore, keywords such as AI Mammography, AI Healthcare, and Trust AI were employed to conduct comprehensive searches. The results of the literature study can be found in Chapters 2 and 3. However, the literature review also led to a list of requirements and guidelines. They were divided into four categories; Usability, Trust, Motivation, and Technical. The initial specifications are presented in Table 2.

Table 2

An overview of the requirements and guidelines identified through the literature review.

	Criteria	Requirement or Guideline
1.	Usability	
1.1	Should be easy to understand	G
1.2	Should not result in increased workload, mental and physical	G
1.3	Should meet users' expectations and their mental model	G
1.4	Should feel consistent with the rest of the system	G
1.5	Should be compatible with multiple computer systems used by radiologists	G
2.	Trust	
2.1	Should provide users with sufficient control over the product and its features	G
2.2	Should provide access to information about the AI system, its capabilities, and limitations	G

	Criteria	Requirement or Guideline
2.3	Should provide access to background information about the product, such as training data, founders, etc	G
3.	Motivation	
3.1	Should encourage the radiologist to act on the AI information	G
4.	Technical	
4.1	Should be able to be generated using the programming language utilized by the company, such as Python	R

6.1.2 Observations at Unilabs

An observation was conducted at Unilabs in Gothenburg to gain insight into radiologists' current work processes. Region Västra Götaland has an agreement with Unilabs, giving them the responsibility of conducting mammography screening within the region. Unilabs, one of Europe's largest diagnostic companies, currently provides mammography services for women at multiple locations across Sweden (Unilabs, n.d).

During an observation at Unilabs, the workflow of a breast radiologist was studied. It emerged that a radiologist typically dedicates 30 seconds to one minute to analyze a patient's images. Initially, the radiologist reviews information provided by the radiology nurse regarding the mammography examination, such as prior surgeries, reported lumps, or family history of breast cancer. They then quickly review the images of both breasts from the two standard views; CC and MLO. A more detailed analysis follows, during which the radiologist begins by zooming in on the right and left CC images, carefully studying them from top to bottom. This procedure is then repeated with the left and right MLO images. After that, the latest images are compared with older examinations, allowing a more precise assessment of changes over time. Subsequently, the radiologist assesses whether there are any suspicious findings or if the patient's condition is deemed healthy. This standardized process remains consistent across all patients. Figure 11 presents an example of the typical working scenario for radiologists.

Figure 11

A photograph displaying a typical working scenario for radiologists



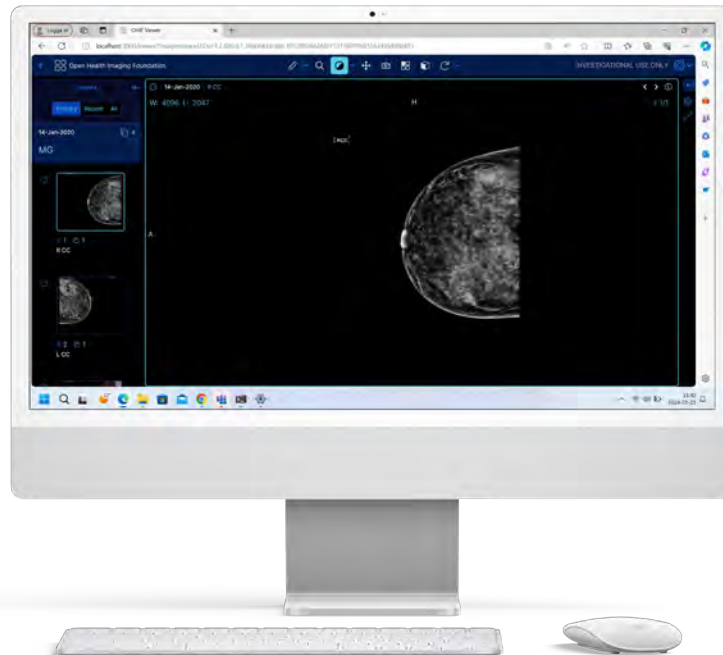
Note: From Adobe Stock (Photograph), by Framestock (<https://stock.adobe.com/se/images/professional-female-doctor>). Adobe Stock standard license.

6.1.3 The Existing Product Demo of Aileen Health

Aileen Health created a product demo at the end of 2023, before the start of the master's thesis, to display how the product will look and function (see Figure 12). The purpose was to recreate the current workflow of radiologists in combination with Aileen Health's AI system, providing them with an overview of a potential future integration. The demo consisted of images from previous and current examinations, along with pictures of Aileen Health's future prognosis. The prognosis included four images: a future multiview comparing the right and left breast (see Figure 13), two views showing the breast expected to develop cancer (see Figure 14) and a future AI overview displaying zoomed-in images of the specific area in the breast from multiple mammography screenings (see Figure 15). In the thesis, the product demo was used to assess the interaction between radiologists and the AI system, thereby addressing the thesis's research question.

Figure 12

An overview of Aileen Health's product demo



The pictures displaying Aileen Health's future prognosis were generated in Python and the DICOM file format. DICOM stands for Digital Imaging and Communications in Medicine and is the international standard format for transmitting, storing, and sharing medical images and associated data in healthcare (DICOM, n.d). To open the DICOM file, the program Orthanc was used, which is an open-source DICOM server designed for healthcare and medical research (Orthanc, n.d). Aileen Health's future prognosis is based on a dataset containing previous mammogram images. The dataset was retrieved from the Swedish National Data Service, named *CSAW-CC (mammography) - a dataset for AI research to improve breast cancer screening, diagnostics and prognostics* (2021-204-1, version: 1). The mammogram images in the dataset are from breast cancer screening at Karolinska University Hospital from November 2008 to December 2015. The dataset has an associated Excel file that compiles relevant information, such as the examination year, cancer status, age group, and tumor size. However, in the Python script, the data are made anonymous by generating new patient names and birthdays. Furthermore, due to the dataset's extensive size, a program capable of accessing and handling the amount of data was necessary. In this thesis, the Cyberduck program was used. In summary, the Aileen Health product demo consisted of pictures from previous and last exams, as well as the AI-generated future predictions.

Figure 13

A future multiview displaying a 24-month future prediction compares the right and left breast in CC view and MLO views

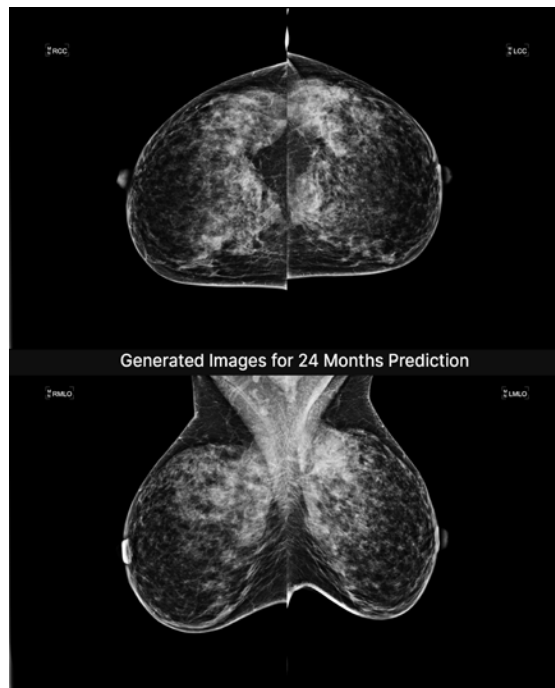


Figure 14

A 24-month future prediction of the breast expected to develop cancer in CC view and MLO view

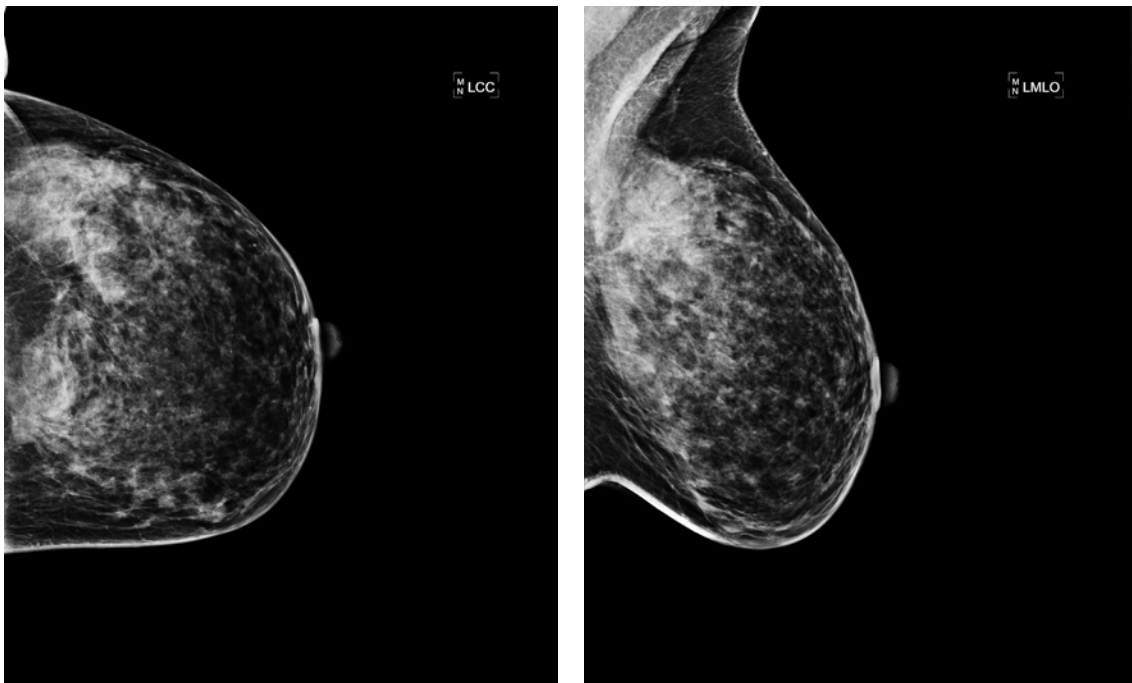
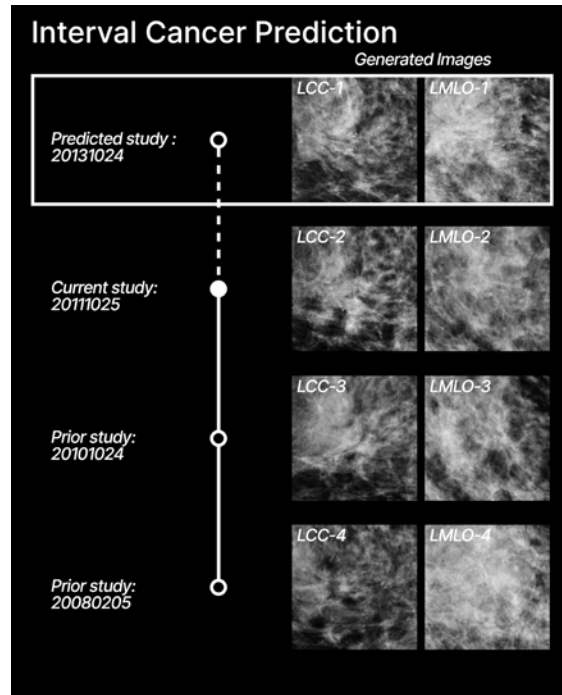


Figure 15

A future AI overview presenting zoomed-in images of the specific area of the breast where cancer is expected to occur. These images show past mammogram examinations up to the latest one, along with AI-generated predictions for future occurrences.



6.1.4 Initial Usability Testing

The thesis team collaborated with the supervisor at Aileen Health to discuss the aspects to evaluate during the initial usability testing. The primary objective was to understand the radiologists' attitudes toward the AI-generated future prognosis, along with the underlying reasons for their perspectives. Based on the discussion, five cases were formulated to serve as the foundation of the test. The first case explored whether the radiologist's actions differ when presented with a future prognosis. The second and third scenarios assessed whether radiologists are more willing to take action based on a future prognosis indicating substantial or minor changes. Lastly, the fourth and fifth cases assessed the optimal amount of information required for effective decision-making.

For this evaluation, it was essential to obtain mammography images that aligned with the five formulated cases so that the Python script could generate future prognoses. The Excel file belonging to the Swedish National Data Service dataset was used to review previous mammography examinations to identify relevant examples for the formulated cases. The associated images were later retrieved from the dataset using the Cyberduck program, and the Python script was used to generate a future prognosis for each case. The DICOM files were then uploaded to the demo using Orthanc.

Pilot

An internal pilot test was executed to identify potential pitfalls and errors in the forthcoming usability test. This involved simulating the actual usability test to uncover as many issues and areas for improvement as possible. Throughout this trial, several opportunities for enhancement were identified. First, it was noted that the presented information required greater clarity and detail. Specifically, more precise instructions about the various cases and more background information about the patient, such as age, were needed.

Furthermore, there was a need to expand the information regarding the generated future AI overview and explain the source of the zoomed-in images. As a result, a rectangle was created in the future multiview to represent the zoom-in position. The color of the explanatory text was also changed to orange to emphasize which images were AI-generated, making them more prominent. The design of these DICOM files following the pilot evaluation is shown in Figure 16. It was also discussed whether a more detailed view should be added to more clearly indicate the potential emergence of cancer. Therefore, an additional AI overview with biomarkers, measurable features that indicate signs of disease, was created. Additionally, this AI overview was made horizontally to investigate if it could enhance the understanding. It was decided to add this detailed view to the upcoming usability test to see if the information was necessary. The detailed view can be seen in Figure 17. Furthermore, it was discussed whether an AI score that indicated the likelihood of something growing in the breast would be beneficial for radiologists. Therefore, it was decided to ask this during the upcoming evaluation.

Additionally, the pilot highlighted the importance of adding a case without interval cancer, as the majority of images they encounter daily lack any signs of cancer progression. However, as the interview time with the radiologists was expected to be very limited, the number of other cases was reduced after the pilot tests to ensure time for more detailed answers. Therefore, only four cases remained for the actual interview. Consistent with prior sessions, the first case investigated whether the radiologist's actions vary when presented with a future forecast. The second and third scenarios assessed whether radiologists are more likely to take action based on forecasts indicating significant or no changes. Finally, the fourth case aimed to establish the optimal amount of information required for effective decision-making.

Figure 16

The previous design of the future AI overview and future multiview is displayed to the left, while the design after the pilot evaluation is shown to the right

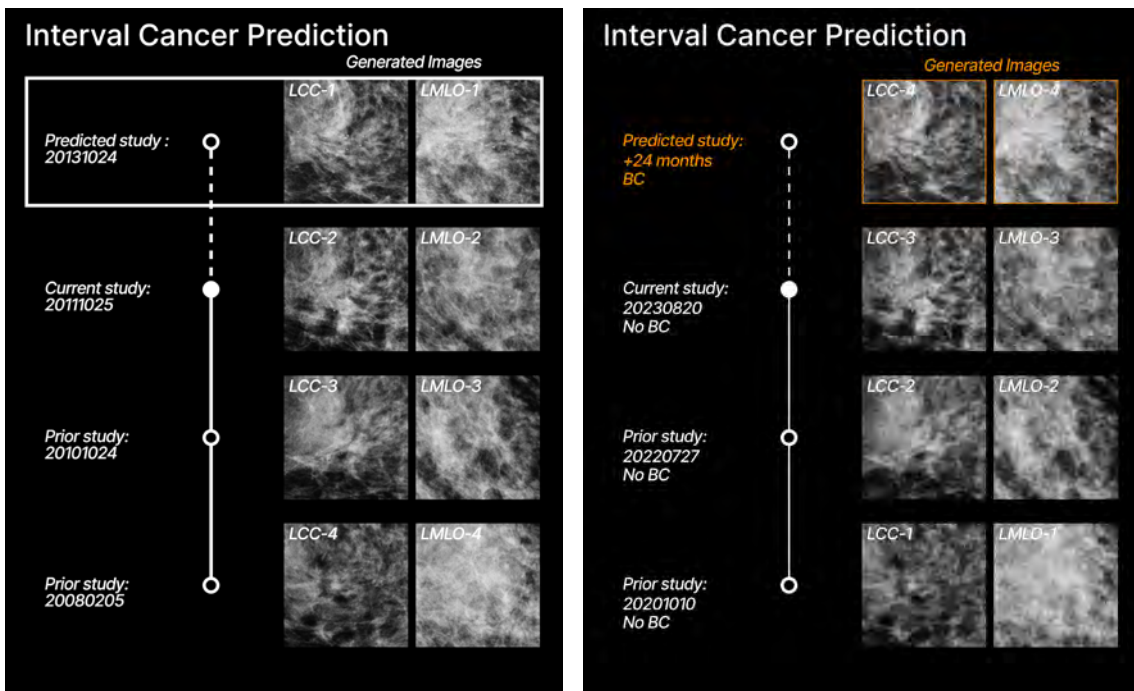
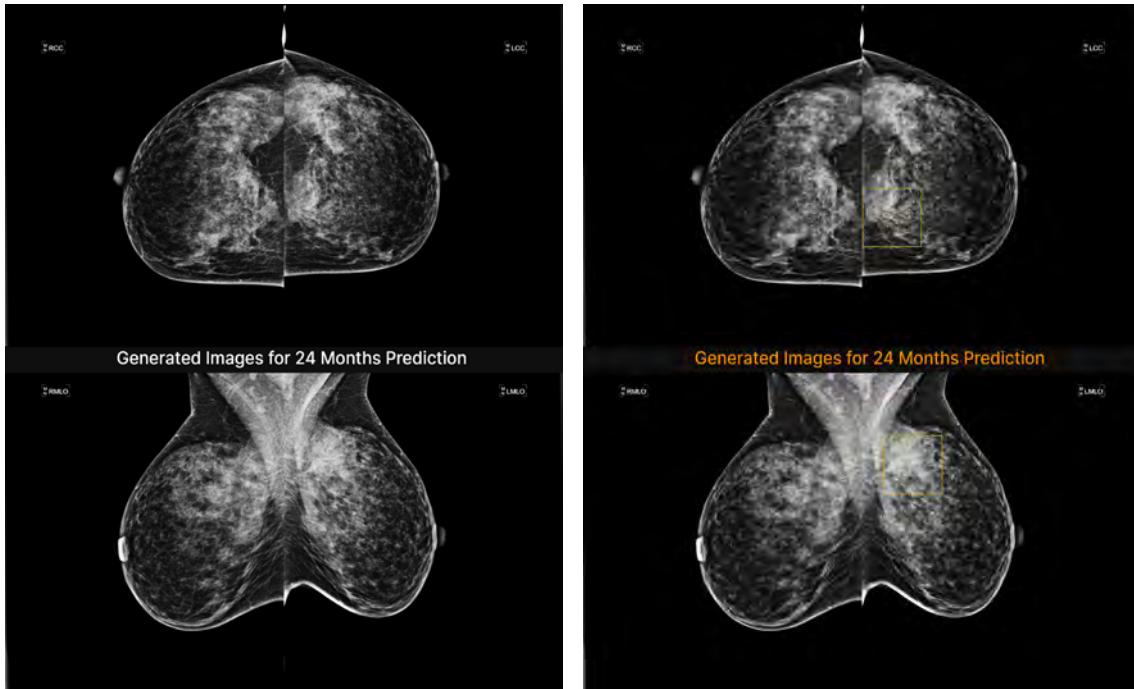
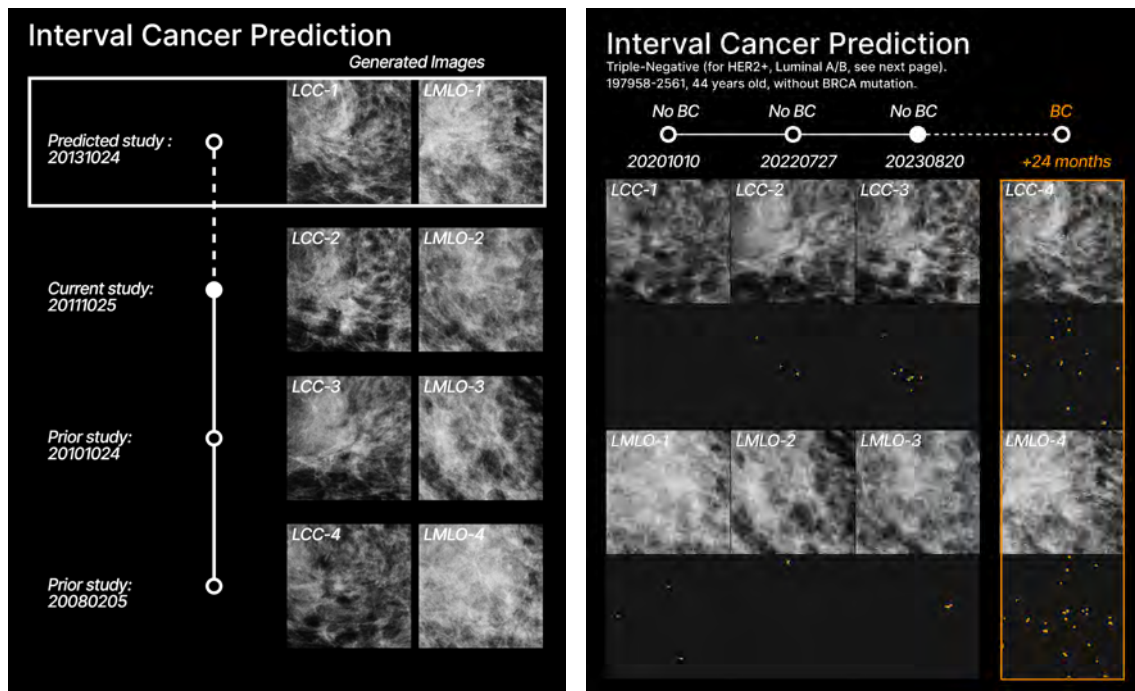


Figure 17

The previous design of the future AI overview is displayed to the left, while the design, including a more detailed horizontal view with biomarkers, is shown to the right



Usability Testing with the Radiologists

The evaluation was conducted using semi-structured interviews and observations at Unilabs with three radiologists. Due to time constraints, the evaluation was done collectively with all radiologists. The interview template can be found in Appendix A. All radiologists had at least ten years of experience in mammography. The objective of the first case study was to assess any behavioral changes following the presentation of AI-generated predictions. To initiate the case, a presentation of the demo was provided to the three participants, allowing them to navigate through the demo and examine the images. Subsequently, they were asked to diagnose the patient based on the presented information. During this part, the participants expressed confusion. They emphasized technical limitations, such as the small size of the pictures, the screen, and its resolution, which hindered their ability to provide accurate diagnoses. Consequently, this left them feeling confused about the demo.

After expressing their initial opinions on the first case, the radiologists were presented with AI-generated prediction. Subsequently, they were queried on whether their diagnosis and plan of action would be altered after seeing the AI prediction. It was noticed during the observation that the radiologists were confused. When questioned about the presentation

“It’s not obvious, and if there’s any artificial, so to speak”

Participant 1

of the AI prediction and whether it was apparent to them which file was AI-generated and which was real, it was not immediately evident to them. Despite including textual cues such as *predicted study* and *generated image*, along with a color alteration from white to orange after the pilot study, the

radiologists encountered confusion distinguishing between the AI-generated image and the real ones. Due to the technical limitations, the radiologists could not answer if they would change their diagnosis after looking at the AI-generated predictions since they could not diagnose the patient in the first place.

The second and third cases followed the same structure as the first. Initially, the participants were presented with X-ray images without the AI-generated predictions and

“Screening is screening. If we also need to look at a future image, then it doesn’t take a minute to screen through”

Participant 2

were asked for their diagnosis. Subsequently, the AI-generated prediction were shown to assess any differences. The second and third cases aimed to investigate whether the radiologists’ actions would vary depending on whether the future prediction showed a significant or no change. In these cases, the same issue emerged regarding the radiologists’ concern that they lacked the necessary technical resources to perform optimally and accurately diagnose patients. However, the cases uncovered additional insights. It became evident that radiologists were

concerned about evaluating another image and investing more time per patient. They expressed resistance to implementing the AI system if it meant increased workload and time commitment. The fact that the AI tool would evaluate the additional images was not clearly communicated to the radiologists. They perceived these images merely as additional work rather than recognizing them as part of an assistive AI system. However, when explained that the images would also be assessed by the AI, participants expressed a more favorable reaction.

During these discussions, the question emerged whether the participants would prefer consistently receiving predictive images, regardless of whether cancer was detected or not. The radiologists described the importance of consistency in this regard. One participant expressed that if such an AI model were to be implemented, it should apply in every case across all patients. Additionally, they underscored the need for future images of both breasts rather than solely focusing on the breast where cancer risk was identified. This approach facilitates a comprehensive comparison between the breasts to detect

“It’s possible for cancer to be hidden in something one doesn’t believe is anything either, so either you have to have it on every picture or not on any”

Participant 3

abnormalities.

“So, I find it really difficult for me to answer that because I have no material to base it on. I only have this right now. If I had known what a normal pattern looks like, then I could have assessed these”

Participant 2

The fourth case aimed to investigate the amount of necessary information. In contrast to previous cases, a more detailed future AI overview was presented, which included not only zoomed-in images but also a biomarker indicating the potential emergence of cancer. The information was also presented horizontally, unlike previous cases where the timeline was presented vertically. Initially, radiologists were asked to interpret the biomarker and explain what they believed its purpose was. Opinions and interpretations varied. One participant found it challenging to

conclude because they lacked comparative data, making it difficult to establish what was considered normal. Another participant expressed concern that excessive information on the screens could lead to overwhelming feelings. The third participant, on the other hand, found the biomarker helpful as it increased and guided attention.

Subsequently, radiologists were asked for their opinions on the horizontal presentation. All participants strongly preferred the horizontal format as it resembled their usual reading structure, where they read from left to right when examining old images. Radiologists were also asked whether they would like a scoring system where the AI indicates the likelihood of something growing in the breast through a numerical value. All participants were positive about this idea, believing it would make the AI's assessment more concrete rather than just generating a new image. They also noted that if AI provides such distinct assessments, it could lead to faster evaluations. Finally, radiologists were also questioned whether AI should suggest actions, such as biopsies and ultrasounds. They argued that

"I can say this, we're overwhelmed with screening images.

We have a lot to do. And right now, this is more difficult for me to absorb than continuing to try to sort through the piles"

Participant 3

they preferred to decide this themselves as it heavily depended on the overall appearance of the breast. Instead, they preferred AI only to indicate where to look and be extra thorough.

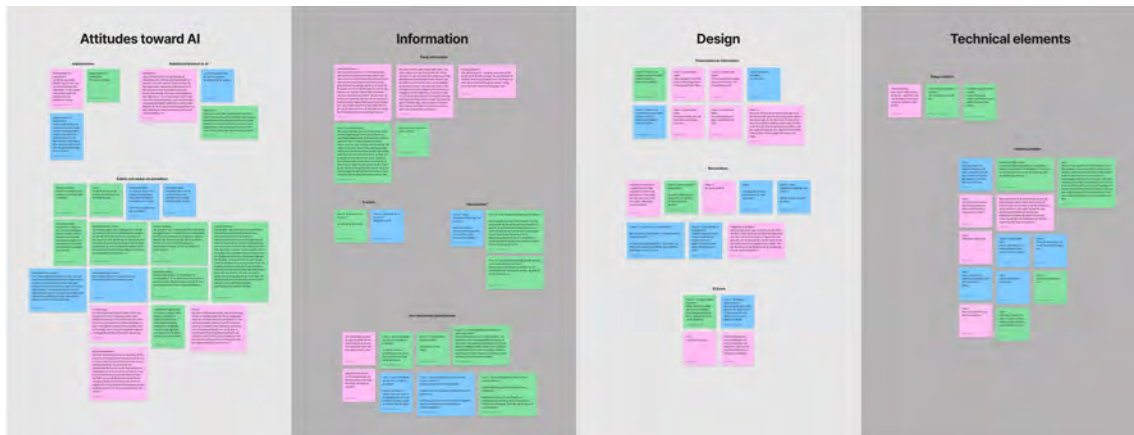
During the evaluation, concerns emerged about how patients would handle the information. They expressed that an additional visit, which might not lead to any significant findings, could cause unnecessary stress and anxiety. They also questioned who is responsible for the AI prognosis and what is expected to be done with the new information. Some feared this could lead to more workload and unnecessary recalls.

6.2 Define Phase

The interviews conducted at Unilabs were transcribed and subsequently analyzed using KJ analysis to pinpoint the fundamental issues. The KJ analysis method allowed the data to be systematically arranged and sorted according to discoveries concerning attitudes toward AI, quantity of information, design, and technical elements, as seen in Figure 18. Grouping similar ideas and information facilitated the identification of patterns and provided a deeper understanding of the problem area. Furthermore, participants were assigned Post-it notes with individual colors to simplify the distinction of statements.

Figure 18

An overview of the KJ analysis showing the identified categories



Most of the Post-it notes could be categorized into attitudes toward AI, quantity of information, design, and technical elements. However, a few notes overlapped and appeared to belong to many categories. These notes required further discussion to ensure appropriate placement. The same applied to notes that did not align with any specific category. As it was still considered essential to keep these thoughts in mind, subcategories were created to give context to the notes.

Through the interviews, it became evident that improving the clarity of the information is necessary. For instance, a more prominent indication that the information is AI-generated was needed. Since it became clear that only text was not enough, alternative methods were required to improve the clarity. Moreover, the future AI overview must distinctly communicate that it contains a timeline featuring images of past examinations, the most recent assessment, and a 24-month forecast. It also emerged that information about the AI, including details about training data and algorithms, should be accessible to those interested in learning more.

The results from the KJ analysis validated the requirements and guidelines developed from the literature review. However, additional ones were identified through the interview. These requirements and guidelines are compiled below in Table 3.

Table 3

A summary of the requirements that emerged from the interviews

	Criteria	Requirement or Guideline
1.	Usability	
1.6	Should be clear which images are AI-generated	R
3.	Motivation	
3.2	Should be time-efficient to use	G

	Criteria	Requirement or Guideline
3.3	Should not be perceived as a competitor to radiologists	G
4.	Technical	
4.2	Should have high image resolution	G

6.3 Ideation Phase

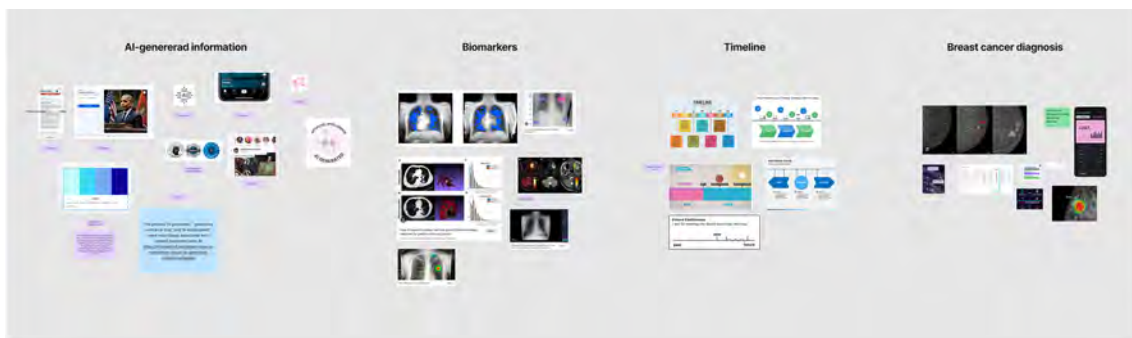
This section describes how the findings from the first usability test were used to generate new ideas for solutions to the identified problems.

6.3.1 Benchmarking

The KJ analysis pinpointed areas for improvement, such as users' comprehension of DICOM files. To enhance user understanding in this regard, several focus areas were established; AI-generated information, biomarkers, timeline visualization, breast cancer diagnosis, and the development of an onboarding process. Subsequently, dedicated benchmarking sessions were organized for each focus area to find inspiration. Using the software FigJam, inspirational materials were collected to see how existing solutions address challenges such as labeling AI-generated content and visualizing diagnoses (see Figure 19).

Figure 19

Inspirational material collected during the benchmarking session



6.3.2 Brainstorming

During the ideation phase, brainstorming sessions were conducted to generate novel solutions for the issues identified in the usability test. Sessions were performed for each focus area, ensuring comprehensive idea generation for all domains. The Crazy 8 method was employed during these sessions, where the thesis team had eight minutes to produce eight ideas. This approach fostered creativity and encouraged thinking beyond conventional boundaries. Subsequently, the most promising ideas were selected through discussion and evaluation.

Following the Crazy 8 sessions, a mindmapping exercise was conducted to refine and develop the most favorable ideas for each focus area. The thesis team sketched concepts within a one-minute time frame, passing their sketches to each other for further elaboration. This collaborative process resulted in sketches of each focus area that formed the basis for subsequent prototyping sessions.

6.4 Prototyping Phase

This section explains the process of refining the initial concepts from the ideation phase into two prototypes; the Future Multiview and the Future AI Overview. The prototypes were developed in the program Figma, using the sketches from each focus area as the basis. The aim is to discover an optimal solution by creating multiple versions that can be utilized in upcoming evaluations.

6.4.1 Future Multiview

In the Future Multiview, the main challenge revolved around effectively communicating to radiologists that the file shows an AI-generated image with a future prediction rather than an actual X-ray image. This concern was identified as one of the focus areas during the ideation phase discussions.

During the brainstorming session, it was decided to apply labels to the AI-generated images to clearly distinguish them from the others. Suggestions for labeling the AI-generated images included a stamp (see Figure 20), a watermark (see Figure 21), multiple watermarks (see Figure 22), the brand logo (see Figure 23), and a text box (see Figure 24). The discussion also focused on establishing the appropriate text for the labels. The phrases AI-generated and Artificial were chosen since these two are perceived as the clearest according to the literature review (Arechar et.al, 2023). In summary, the future multiview prototyping session resulted in five alternatives for AI labels. However, it was determined that radiologists needed to evaluate the prototypes to select the clearest and best suited for the purpose.

Figure 20

The Future Multiview with an AI stamp

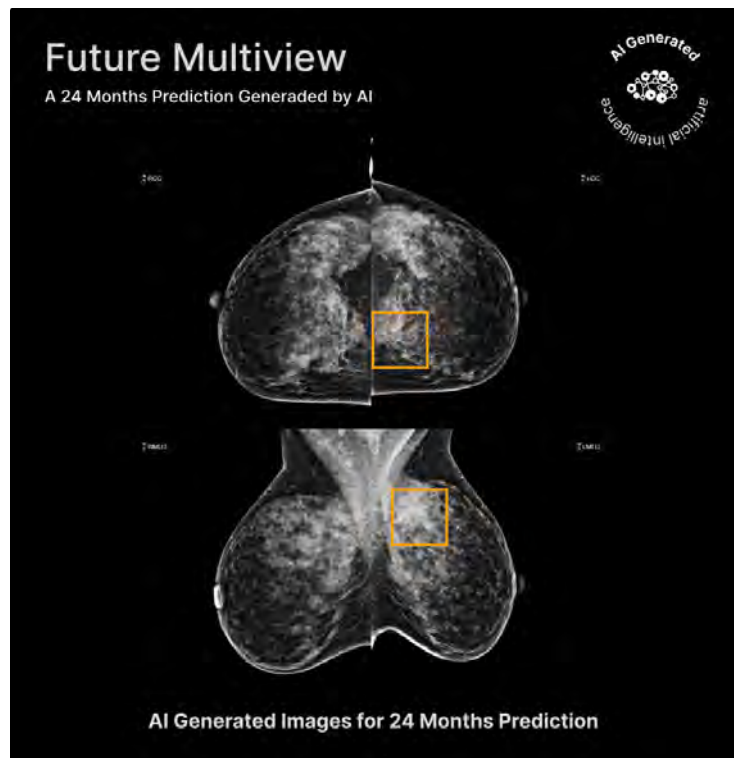


Figure 21

The Future Multiview with one watermark



Figure 22

The Future Multiview with multiple watermarks

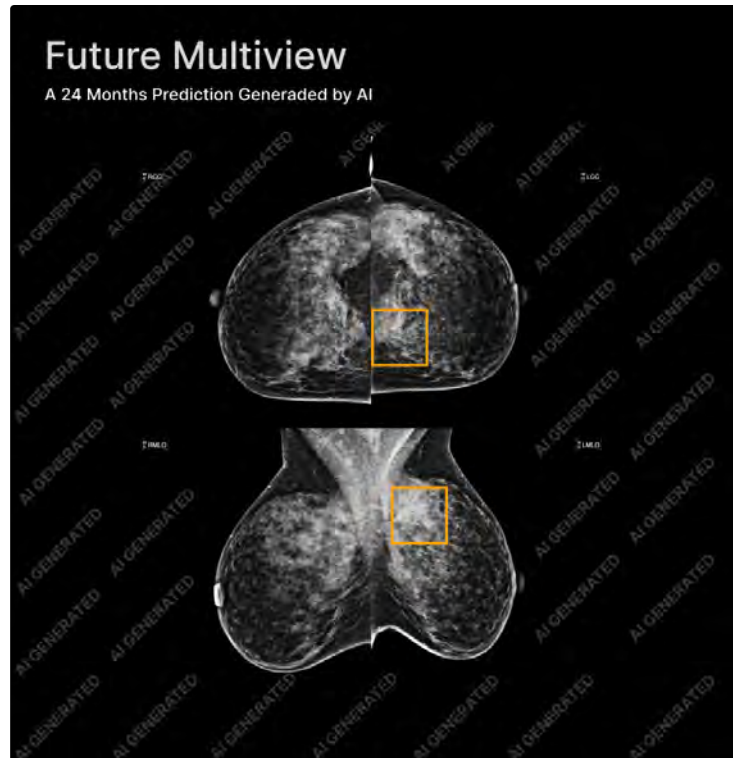


Figure 23

The Future Multiview with the brand logo of Aileen Health

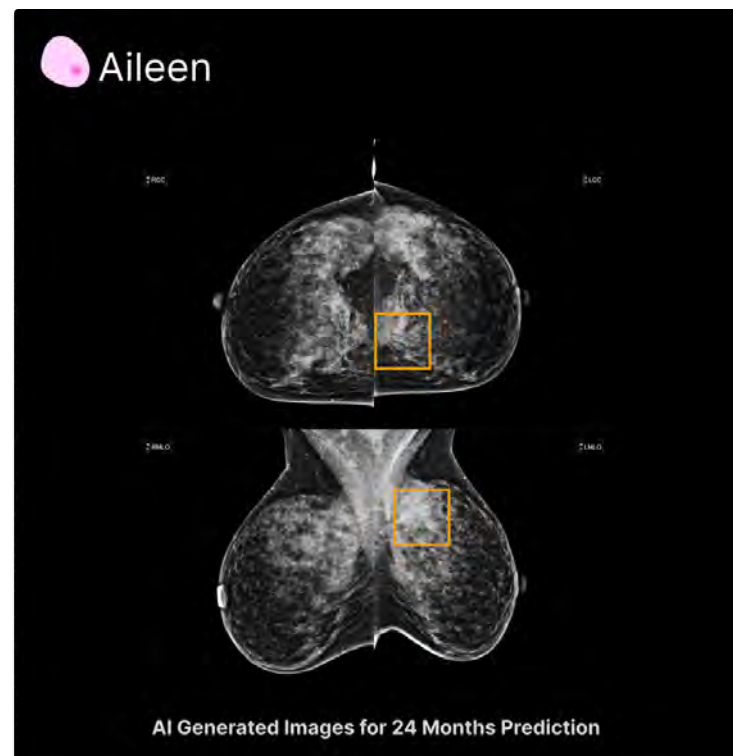


Figure 24

The Future Multiview with a text box



6.4.2 Future AI Overview

Regarding the Future AI Overview, several challenges were identified during the usability test and were subsequently selected as focus areas during the ideation phase. These included biomarkers, timeline representation, and diagnosis visualization.

Biomarkers

This focus area centered on exploring optimal methods for visualizing biomarkers. These biomarkers aim to provide quantitative insights into the progression and outcome of breast changes. During the brainstorming sessions, several presentation alternatives were explored, such as heat maps and circular markings indicating suspicious areas within the breast. However, the session resulted in two concepts concerning biomarkers. The first option involved preserving the original design with an additional view of the biomarkers, as seen in Figure 25. Meanwhile, the second alternative suggests integrating biomarkers over the existing images to offer precise localization of changes within the breast, as seen in Figure 26. However, the decision was made not to proceed with the second concept and instead to focus solely on developing the original design, as it is crucial to avoid disrupting the X-ray image.

Figure 25

The original design with an additional view of the biomarkers

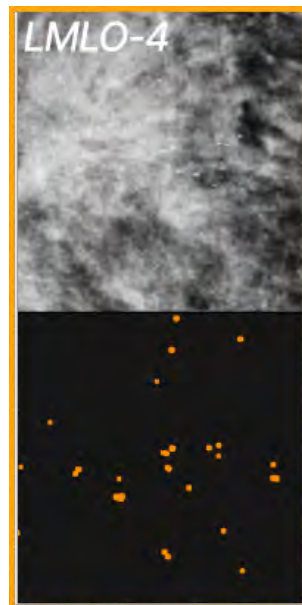
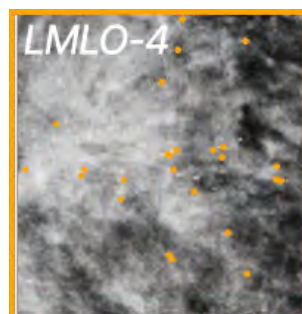


Figure 26

The second design with biomarkers over the existing image

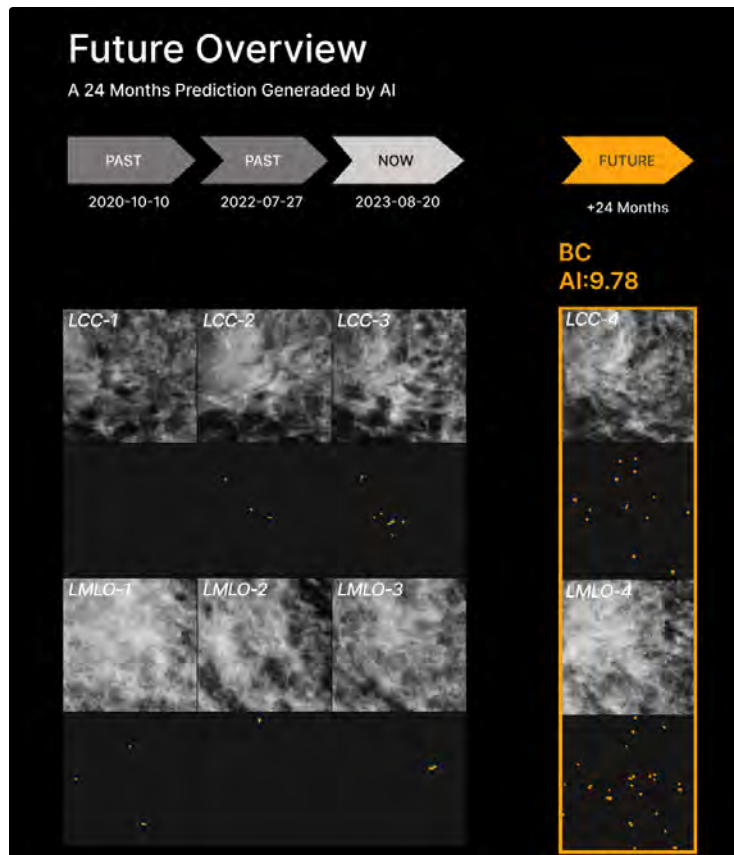


Timeline

The results from the usability test revealed a lack of clarity for radiologists in comprehending the generated AI overview’s timeline. Consequently, the brainstorming session focused on strategies to enhance the visualization, specifically clarifying the timeline. Ideas such as incorporating directional arrows, further differentiating the AI-generated image with additional space, and refining the language used within the AI overview were among the discussed proposals. The brainstorming session resulted in prototypes of the AI overview with added arrows and new headings to highlight which images are past, present, and future, as seen in Figure 27. Additionally, it was decided to employ a horizontal timeline instead of a vertical one based on the findings from the usability test, which indicated that radiologists found the horizontal option considerably easier to comprehend.

Figure 27

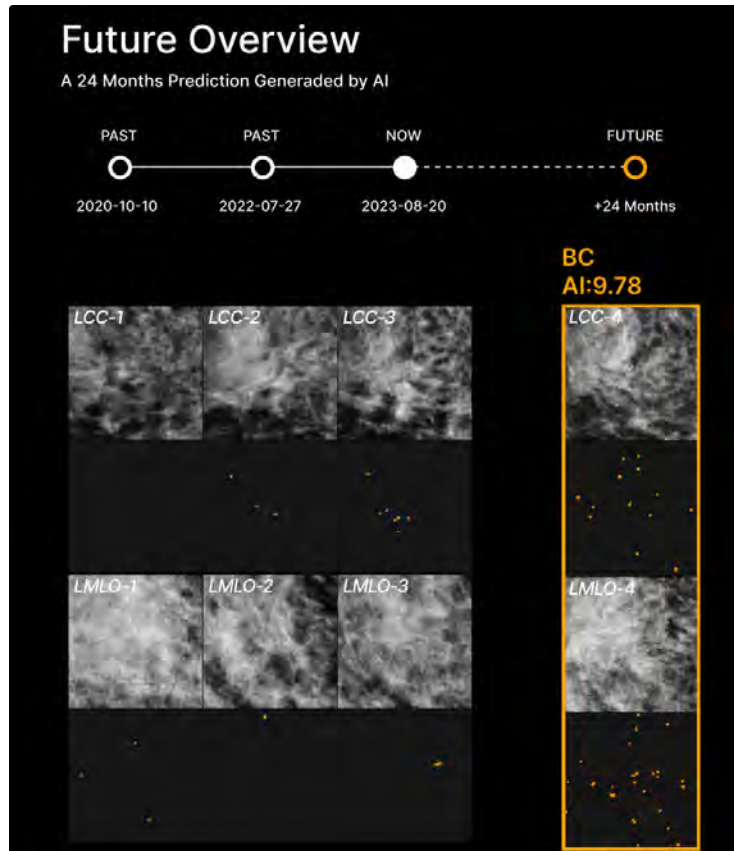
The Future AI Overview with a timeline consisting of arrows and a new heading



Another idea was developed based on Aileen Health's initial product demo, in which the concept of a timeline consisting of dots was kept but transformed into a horizontal format (see Figure 28). As in the previously presented prototype, the title and spacing of the images were also changed.

Figure 28

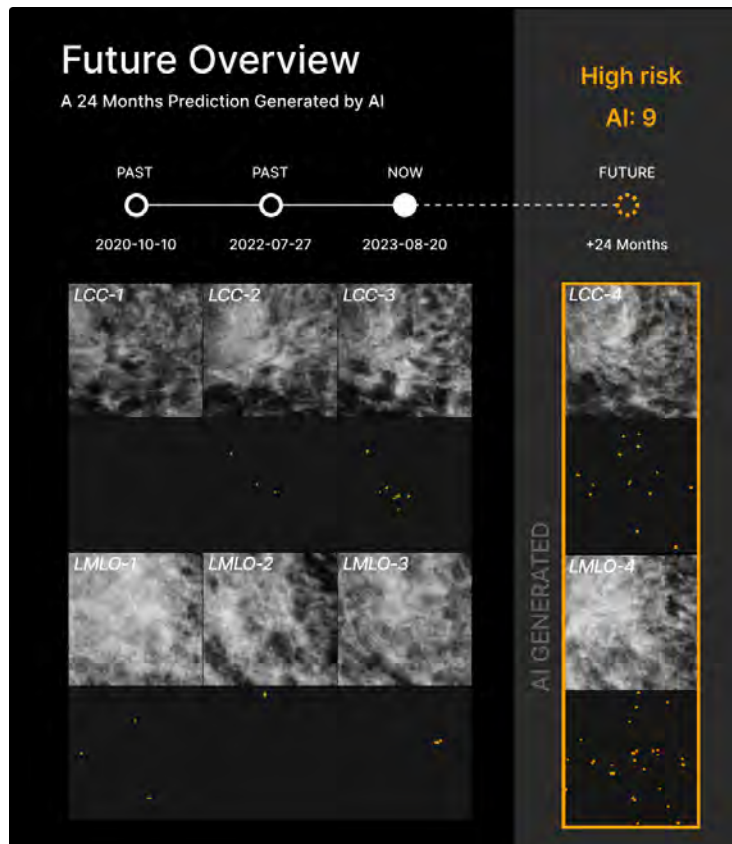
The Future AI Overview with a timeline consisting of dots and a new headline



Lastly, a further development of the prototype with a timeline consisting of dots was designed to investigate if it could enhance the clarity of the AI-generated future images. In this prototype, markers in the form of text and a grey box were added around the AI-generated content (see Figure 29). In summary, the prototyping session resulted in the presented prototypes. However, it was decided that the prototypes needed to be further evaluated by radiologists.

Figure 29

The Future AI Overview with a timeline consisting of dots and a new headline, with additional markers



Breast Cancer Diagnosis

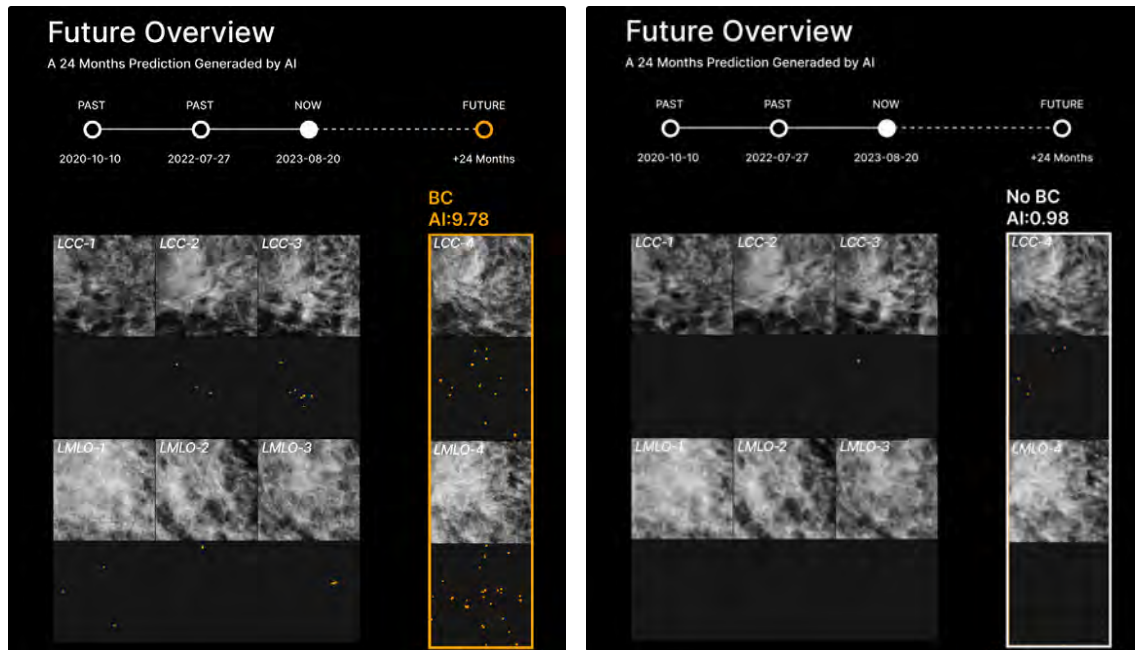
The result from the usability test revealed a common concern among all radiologists regarding the potential increase in workload if a future prognosis was added. They expressed unwillingness to examine additional images for each patient and emphasized the necessity for the AI to generate estimations of the future prognosis to decrease the workload. Therefore, the brainstorming session also focused on effectively presenting the AI's breast cancer estimation. One proposal involved adding distinct colors to the border of the AI-generated image to signify either risk of cancer or no risk of cancer. Using red or orange as a border color was suggested to signify a high cancer risk since these colors are often associated with warning signs or danger (Braun et al., 1995). White was considered as a potential color to represent low or no risk of breast cancer, as it is perceived as less aggressive and less cautionary. The difference between the two Future AI Overviews showing high risk and low risk of breast cancer, is shown in Figure 30.

Additionally, the findings from the usability test indicated that all participants expressed a positive attitude toward integrating an AI score. As the literature review suggests that more precise numbers enhance communication and content quality, one solution was incorporating decimals into the AI score (Kelly et al., 2023). To confirm that this was

also applicable in this context, an alternative solution was created using a scale of 1-10. Consequently, radiologists will assess two distinct options to study their preferences.

Figure 30

The left side shows a Future AI Overview of high-risk cancer, while the right side presents one with a low-risk cancer



6.4.3 Onboarding Process

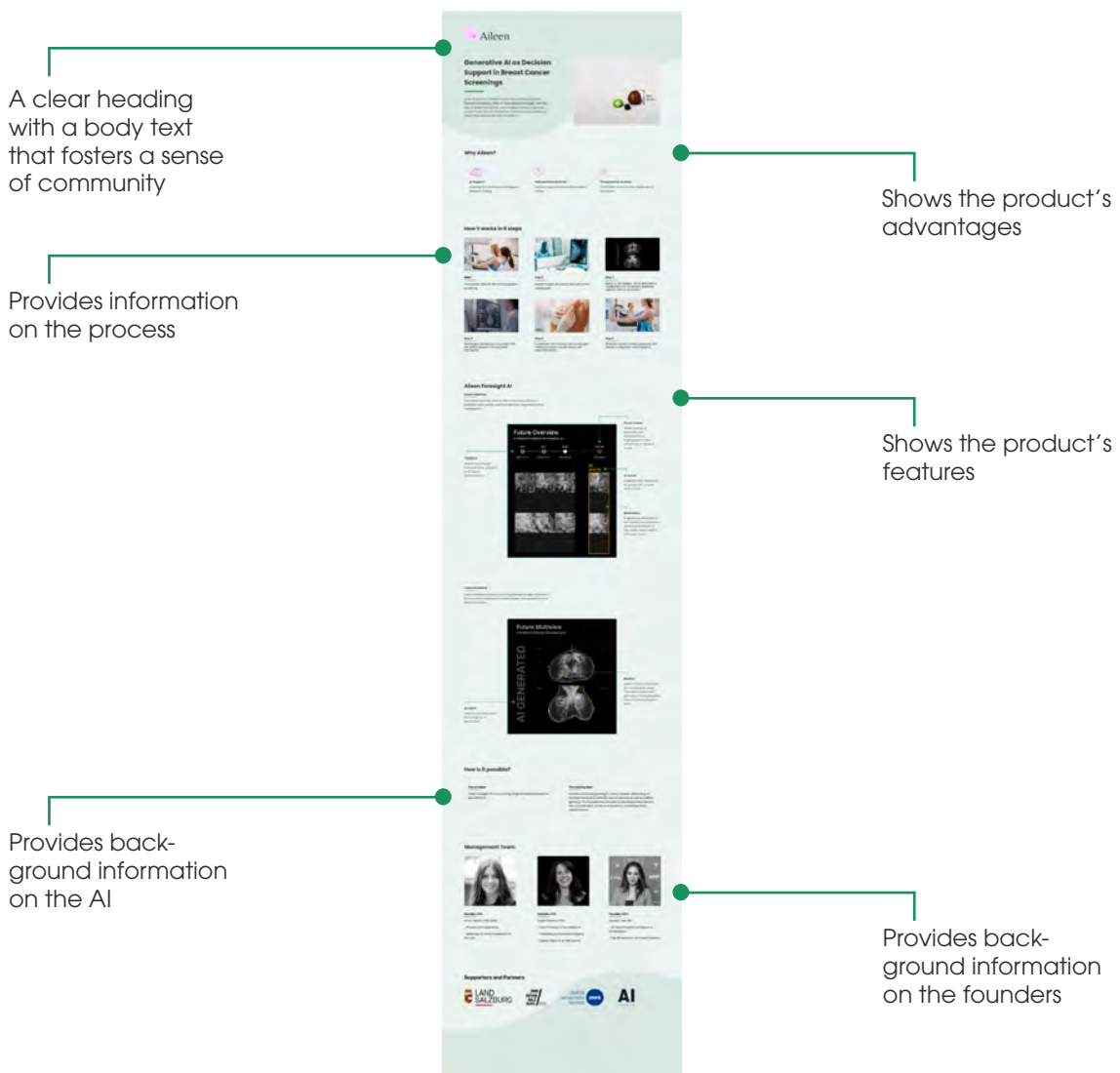
In the preliminary literature review, it became evident that the product must provide users with comprehensive information regarding the AI's capabilities, constraints, and relevant background details to establish trust. The radiologists also confirmed this during the usability test, emphasizing the importance of effective onboarding processes to ensure users have the necessary knowledge to trust and utilize the AI effectively. Several alternative solutions emerged during the brainstorming session regarding how the information could be presented. For example, an introductory movie was discussed as a suggestion, offering a visually compelling way to communicate the information. However, after consideration, a landing page on a website was chosen. This choice allowed radiologists to focus on the parts they were most interested in rather than requiring them to watch an entire movie, which could potentially cause frustration. This aspect seemed necessary, as the usability testing revealed a varied interest in understanding the functionality of the system and a variety of subjects they wished to explore further.

Based on the hierarchy of needs and the previously developed list of requirements and guidelines (see Table 2 & 3), brainstorming was conducted regarding which information should be included on the landing page and how it could be presented. These design sessions took place in Figma, where wireframes were created to visualize the page's layout and structure. An overview of the landing page's structure is shown in Figure 31.

Significant effort was dedicated to ensuring that the content was understandable and engaging to the audience. This involved refining the choice of words and incorporating images to clarify certain aspects of the information. Furthermore, the landing page maintained consistency with the Aileen Health brand by adhering to the same design language, including identical colors and fonts. In line with the requirement list, the landing page was designed to offer access to information regarding the AI system and its training data. Similarly, details regarding founders and funding were incorporated based on the provided requirements.

Figure 31

An overview of the landing page’s structure made in Figma



6.5 Test Phase

This section presents the execution and outcomes of the testing phase, during which radiologists assessed various solutions. Initially, the execution and results of the interviews are described, followed by the presentation of the compiled data analyzed through a KJ analysis.

6.5.1 Usability Testing with the Radiologist

The second evaluation was conducted using semi-structured interviews and observations at Unilabs with the same three radiologists as in the first evaluation. However, this evaluation was conducted individually with all radiologists. The interview template can be found in Appendix B. The user evaluation contained four primary segments. The initial segment aimed to explore the clarity of different AI labeling methods to ensure radiologists can effectively differentiate between AI-generated and non-AI-generated information. The second segment sought to assess the comprehension of the future overview and its intended purpose among the radiologists. Furthermore, the third segment aimed to evaluate the landing page's impact on the radiologists' attitude toward the software and verify if the provided information was sufficient to foster trust in the system. Finally, the last part aimed to investigate the perception of the product and the emotions it evokes. Most of the information used during the test was printed on paper to avoid the recurrence of technical issues experienced during the previous evaluation, where it was noted that the computer's screen size was too small.

Future Multiview

During the initial segment of the usability test, radiologists were presented with the five design alternatives for AI labels created during the prototyping phase (see Figure 32). These alternatives aimed to indicate that the X-ray image represents a future prognosis generated by AI. These alternatives included:

1. The Future Multiview with a stamp (Figure 20)
2. The Future Multiview with one watermark (Figure 21)
3. The Future Multiview with multiple watermarks (Figure 22)
4. The Future Multiview with the brand logo of Aileen Health (Figure 23)
5. The Future Multiview with a text box (Figure 24)

Figure 32

A picture from the evaluation where five alternatives were presented to the radiologists



Participants were asked to rank the alternatives based on their perceived clarity, ranging from the most clear to the least clear. While there were variations in individual rankings, overall consensus was observed. Notably, the alternative with one watermark received middle rankings from all participants, while the brand logo and multiple watermarks were consistently ranked as the least clear ones. The alternatives with the text box and the AI stamp were consistently ranked as the top two choices. Moreover, when asked to select their preferred option, participants favored either the AI stamp or the text box.

"It shines, you can't avoid looking at the AI stamp. You can't miss it"

Participant 1

Participants were additionally queried regarding the appropriateness of the headlines. All found the headlines to be fitting and illustrative of the content. However, one participant expressed a slight challenge in comprehending the headlines fully since they still lacked a complete understanding of the product and similar systems. Nonetheless, this participant stated that the headlines were likely appropriate and would probably become more understandable with increased familiarity with the product.

"If you know what it means, then you understand what it signifies, but since it's not something we work with, it becomes a bit confusing"

Participant 2

Participants were also asked to describe their understanding of the phrase *AI-generated*. All interpreted this phrase to signify that the image and the future prognosis were produced by a machine. They all agreed that the phrase *AI-generated* was more appropriate for this kind of information and provided a clearer indication that the content is *AI-generated* than the word *Artificial*.

Future AI Overview

The second segment of the usability test involved evaluating the generated AI overview. During this assessment, three design alternatives were presented:

1. The Future AI Overview with a timeline consisting of arrows (see Figure 27)
2. The Future AI Overview with a timeline consisting of dots (see Figure 28)
3. The Future AI Overview with a timeline consisting of dots, with additional markers (see Figure 29)

All alternatives of the Future AI Overview were created in two versions: one that showed breast cancer risk and one that did not. The version indicating risk was highlighted with orange details, while the other used white. For all design alternatives, participants were asked to describe how they interpreted the information presented, identify any unclear information, specify any parts they particularly liked, and indicate whether it was clear which part of the Future AI Overview consisted of the AI-generated prediction.

When the radiologists were presented with the first image of the Future AI Overview consisting of dots, all understood that it represented a timeline. They comprehended that it included past examinations, the current one, and the AI-generated future prediction. All participants appreciated the inclusion of previous exams, although the presentation of small, zoomed-in images received less favorable feedback. Notably, one participant misinterpreted the future prediction as indicative of cancer despite the AI estimation showing no signs of cancer in the specific case being examined.

"I interpret it as AI believing that something will happen in the future"

Participant 2

The second part of the test introduced the alternative design for the AI overview, incorporating arrows instead of dots within the timeline. This iteration also featured two images: one with a cancer risk and one without. Participants were queried about their preferences regarding the design alternatives. Responses varied among radiologists: one favored the dots, perceiving them as clearer, while another favored the arrows for the same reason. Conversely, the third individual perceived no significant difference in clarity between the two designs.

"The first one, the one with a clear timeline. Meaning I prefer the one with the dots"

Participant 1

In terms of distinguishing between no breast cancer and breast cancer, none of the participants were able to discern the color coding, where white signified the absence of breast cancer, and orange indicated its presence. They perceived the colors as lacking any inherent meaning and preferred the white color's simplicity. Furthermore, participants were questioned regarding their ability to distinguish the AI-generated image throughout both design alternatives. This distinction was clear to all participants, as they quickly identified the image positioned to the right as the future image. The third design option was also presented, incorporating a gray marking to distinguish the AI-generated image (see Figure 29). Without exception, all participants interpreted this marking as an additional indicator of the AI-generated image. While two participants preferred the design option featuring the marking, a third participant acknowledged the clarity of the marking but considered it unnecessary.

"No, no, it's just confusing. It becomes too precise, and it can't be that precise. So I prefer only whole numbers"

Participant 1

Finally, the AI's accuracy in estimating cancer was assessed within the same section of the test. Participants were questioned about their preference between AI scores represented as whole numbers or more precise numbers with additional decimals. Without exception, all participants preferred whole numbers and perceived no significance in including more precise numerical values.

Landing Page

In the third segment of the evaluation, participants were provided with a landing page in a website format. The landing page was introduced alongside a scenario to help them grasp the context. Participants were instructed to review the page's content and respond to questions. All participants were satisfied with the information on the website. It offered an engaging introduction to the software by stimulating interest and encouraging exploration.

"You might see it on a website the first time, but then it's good to have it printed on paper so you can have it next to you while you're working"

Participant 3

However, one participant suggested also printing the information for easy access during work sessions. The same individual also proposed that offering more insights into patient needs and their reactions to AI-generated predictions would be advantageous. They believe ensuring genuine interest from the patient's perspective is essential, as it signifies that someone wants to receive the AI-generated information. The participant argues that without this assurance, there is a risk that women might feel even worse and become increasingly anxious upon receiving such a message.

Furthermore, another participant highlighted the need to make the website available in Swedish to accommodate people's various degrees of English expertise. The same participant also emphasized the importance of underlining who was responsible for delivery, particularly in the context of screening. Nevertheless, the final participant expressed that despite the information on the landing page, they struggled to grasp the

"I still don't feel convinced about its usage, I'm not convinced that this is a good way, but it is clearly described"

Participant 2

product’s purpose. They acknowledged the clarity of the material but remained skeptical that the work strategy with AI-generated predictions was effective.

Participants were also asked to complete a semantic word scale containing seven adjectives aligned with the product’s desired attributes: credible, informative, reliable, necessary, understandable, helpful, and consistent (see Appendix C). The results are summarized in Figure 33.

Figure 22

A compilation of the semantic word scale



Finally, the participants were instructed to fill out Hesselgren’s emotional scale, rating their emotional responses to the product from 1 to 6 (see Appendix D). The scale contained primary emotions such as joy, trust, fear, surprise, sadness, disgust, anger, and anticipation, which were deliberately selected to cover a broad spectrum of emotional responses. Only two participants chose to complete the task; the results are shown in Figure 34.

Figure 34

A compilation of Hesselgren's emotional scale



6.5.2 Analysis of Gathered Data

As in the previous usability test, the interviews were transcribed, and the results were written down on Post-it notes. These Post-it notes were then used in a KJ analysis to structure the collected data. As previously, the participants were assigned specific colors on Post-it notes to differentiate their statements, and since the participants remained consistent, the same color coding was utilized. The notes were then organized based on themes, resulting in three main categories: Future Multiview, Future AI Overview, and landing page. However, additional subdivisions were made within each category to offer more specific context for the notes, facilitating a more straightforward analysis. Due to varying amounts of time available for interviews among the radiologists, the length of interviews varied greatly. Consequently, as illustrated in Figure 35, the number of statements and Post-it notes per participant differed.

Figur 35

An overview of the KJ analysis conducted to organize the collected data.



It became evident that participants grasped the information more thoroughly compared to previous tests. They particularly valued the clear labeling of AI-generated information with stamps or text boxes, finding it the most straightforward approach. Furthermore, participants universally appreciated the use of headings, which provided a clear context for the images. The effectiveness of AI-generated wording was also confirmed, as all participants indicated understanding.

Opinions were split regarding the preferred proposals for the Future Multiview. However, it was reassuring that all participants understood the timeline, addressing a previous area of concern during tests. While all participants appreciated the landing page, they noted a lack of patient perspective, emphasizing the importance of understanding patient motivations and the product's purpose. The Semantic Word Scale and Hesselgren's Emotional Scale indicated ongoing challenges in fully grasping the product's purpose and value. However, despite difficulties in fully appreciating the product, participants found it informative and understandable, as indicated by the Semantic Word Scale. The insights gained from the second evaluation were then used for the last design iteration to develop the final results.

7. RESULTS

This chapter presents the results of the thesis project in two sections. The first part outlines essential requirements and guidelines for designing a visual representation of an AI system intended for implementation in healthcare. The second section presents the final design proposal. Together, these outcomes respond to the research question of the thesis, providing both a framework for the development process and a solution in the form of a design proposal. They address key factors for trusting and integrating AI-based predictions into current mammography screening workflow to support radiologists.

7.1 Requirements and Guidelines

Developing visual representations for an AI system intended for implementation in healthcare demands precision, insight, and a profound understanding of the unique challenges and requirements within the healthcare system. By conducting a literature review, observations, and interviews, a set of requirements and guidelines has been identified to support the development process. These requirements and guidelines can be categorized into four main categories; usability, trust, motivation, and technical. Given that most identified needs demand subjective assessment and cannot be answered with a simple yes or no, they are classified as guidelines rather than requirements.

7.1.1 Usability

Ensuring high usability is essential for successful implementation when integrating AI into existing healthcare systems. Radiologists must find the AI system easy to interact with and comprehend, as low usability could lead to increased mental workload and frustration, potentially resulting in non-utilization of the system. Achieving this requires good discoverability, a clear understanding of existing features, and comprehensive knowledge of the product's intended usage. Additionally, AI-generated material must be presented intuitively for efficient knowledge transfer. Furthermore, the system must align with the user's mental model to avoid dissonance between expectations and actual capabilities, which can reduce trust in the system. The requirements and guidelines in this category are presented in Table 4.

Table 4

The requirements and guidelines falling within the Usability category

	Criteria	Requirement or Guideline
1.	Usability	
1.1	Should be easy to understand	G
1.2	Should not result in increased workload, mental and physical	G
1.3	Should meet users' expectations and their mental model	G
1.4	Should feel consistent with the rest of the system	G
1.5	Should be compatible with multiple computer systems used by radiologists	G
1.6	Should be clear which images are AI-generated	R

7.1.2 Trust

Trust is a fundamental consideration when developing a visual representation of an AI in healthcare. When users perceive that they have control over the product and are provided with transparent information regarding its functionalities, data usage, and decision-making processes, their trust in the product grows. Trust is crucial for users to accept and utilize the new technology. If users are not confident in the system, they may not use it or use it with hesitation, which can negatively affect the effectiveness and results of the implementation. The requirements and guidelines in this category are shown in Table 5.

Table 5

The requirements and guidelines within the Trust category

	Criteria	Requirement or Guideline
2.	Trust	
2.1	Should provide users with sufficient control over the product and its features	G
2.2	Should provide access to information about the AI system, its capabilities, and limitations	G
2.3	Should provide access to background information about the product, such as training data, founders, etc	G

7.1.3 Motivation

Motivating radiologists to act on AI information within the healthcare system is crucial for successful integration. The visual representation should facilitate accessing and interpreting the AI predictions, ensuring that radiologists can integrate them seamlessly into their workflow without causing delays or disruptions. Furthermore, it is also essential to ensure that the AI is not perceived as a competitor to radiologists but rather as a supportive tool that complements their expertise. Emphasizing the AI system’s supportive role and how it enhances radiologists’ capabilities can foster a positive perception and encourage acceptance of the technology. The requirements and guidelines in this category are presented in Table 6.

Table 6

The requirements and guidelines within the Motivation category

	Criteria	Requirement or Guideline
3.	Motivation	
3.1	Should encourage the radiologist to act on the AI information	G
3.2	Should be time-efficient to use	G
3.3	Should not be perceived as a competitor to radiologists	G

7.1.4 Technical

Technical requirements and guidelines were also identified during the thesis project. For instance, it is crucial that the AI-generated visualization can be developed using the programming language utilized by the company. This alignment ensures consistency and compatibility with existing frameworks, facilitating the development processes. Furthermore, high image resolution is essential for accurate visualization and interpretation of medical imaging data. Therefore, providing high image resolution in the visual representation is crucial. The requirements and guidelines in this category are shown in Table 7.

Table 7

The requirements and guidelines within the Technical category

	Criteria	Requirement or Guideline
4.	Technical	
4.1	Should be able to be generated using the programming language utilized by the company, such as Python	R
4.2	Should have high image resolution	G

7.2 Design Proposal

The design proposal contains three main components; a Future Multiview, a Future AI Overview, and a landing page. Each part serves a distinct purpose, yet collectively, they are intended to meet the specified requirements and function as a design proposal. The Future Multiview serves as the primary DICOM file for assessing cancer risk in patients, while the Future AI Overview provides additional insights into suspicious areas and the AI's assessments. The landing page facilitates the onboarding process for radiologists, offering background information about the system and its benefits. Together, these three components contribute to the project's objective of designing a visual representation of AI-generated content that enhances radiologists' trust and motivation to integrate the AI system into their workflow, eventually leading to more proactive and effective patient care.

The Future Multiview and Future AI Overview files will be incorporated into the existing workflow alongside mammography images from previous and current examinations. By integrating these components, radiologists access an extensive dataset comprising historical, real-time, and predictive data. This data gives them vital knowledge, allowing for precise diagnosis and informed decision-making. However, the landing page will serve as a valuable resource during the onboarding process, offering additional context and guidance as the AI system is implemented.

7.2.1 Future Multiview

A vital component of the solution is the Future Multiview, designed to serve as the primary DICOM file for radiologists assessing a patient's cancer risk. This file directs radiologists to suspicious areas and provides insight into the likelihood of growth. It includes several images, capturing both the right and left breast in two different views, as seen in Figure 36.

Figure 36

The final design proposal of the Future Multiview



A headline is employed to facilitate quick identification of the file, positioned in the top left corner for its perceived importance and to enhance visual scanning (Osvalder & Ulfvengren, 2015). By incorporating the term *Prognosis*, it aims to signify that the file contains data related to a forthcoming prognosis. Furthermore, the sans serif font Sofia Pro is consistently employed throughout all textual components of the file. The decision was made due to the font's versatility and suitability for digital interfaces. Furthermore, its availability in 16 different weights offers an advantage by maintaining consistency across headlines, body text, and stamps.

A stamp is positioned at the top right corner to communicate to the user that the image is AI-generated. Through usability testing, it became apparent that this method of

labeling AI-generated content was preferred. The stamp includes both a text field and an illustration. The text field features the phrase *AI generated*, chosen for its clarity based on user feedback. Initially, the stamp featured an AI illustration, but it has been replaced with the breast from Aileen Health's logo to strengthen the company's identity. This change aligns with research suggesting that a strong brand identity can foster trust among users (Michler et al., 2019). However, recognizing that some participants favored a more prominently marked text box, such a box was added at the bottom of the DICOM file. Additionally, this combination results in a redundant presentation format, enhancing the likelihood of accurate interpretation of the information.

Furthermore, rectangles direct users to the suspicious area in the breast. These rectangles are shaped to correspond with the form of the zoomed-in images in the Future AI Overview (see section 7.2.2). Additionally, the rectangle is outlined with an orange border, #FDA50F, a color commonly used to evoke associations with warning or danger (Braun et al., 1995). However, if the AI detects no risk of cancer, no highlighted rectangles are visible. Nevertheless, the Future Multiview uses minimal color apart from the rectangles. This decision aligns with the fact that radiologists are familiar with mainly seeing black-and-white information, ensuring that the visual representation matches their mental model. The high contrast also facilitates the identification of objects, such as the added stamp (Osvalder & Ulfvengren, 2015).

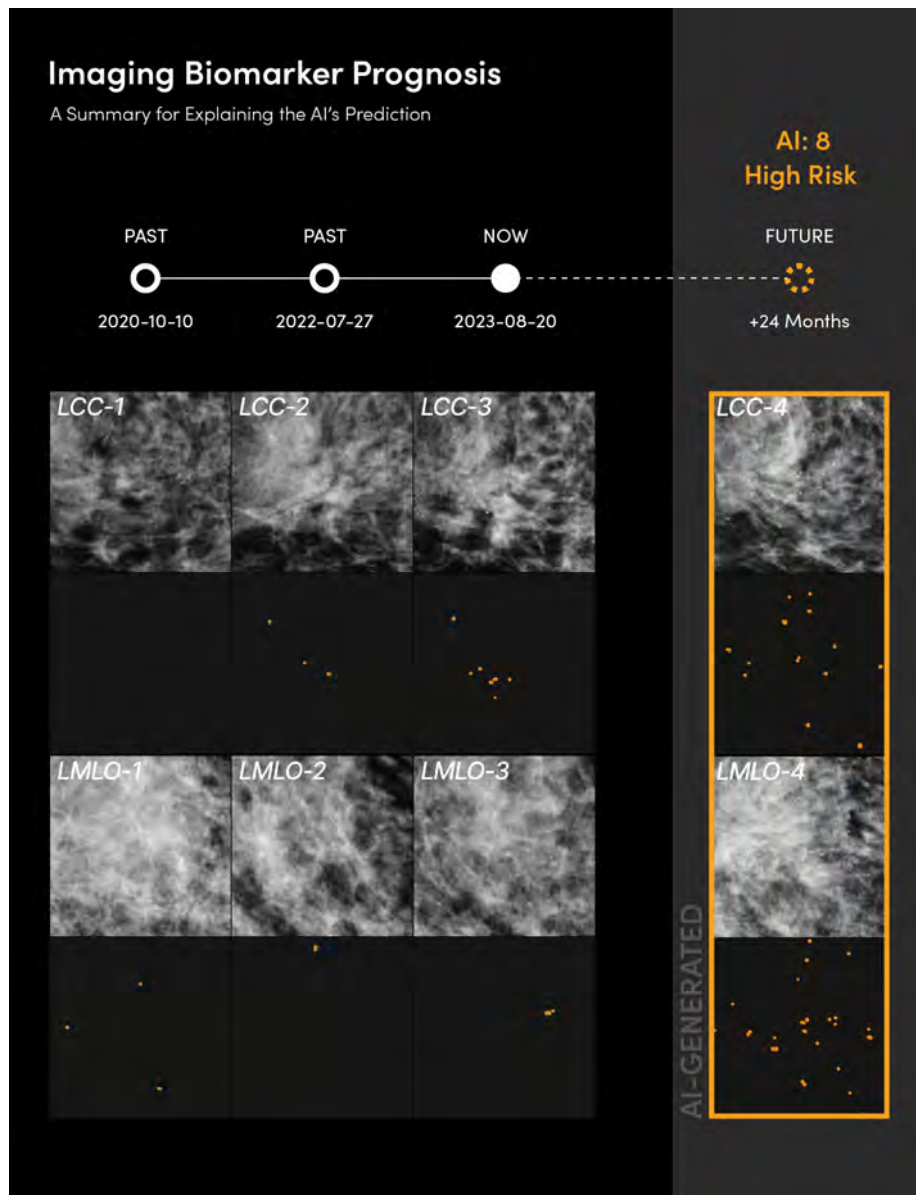
During usability testing, it became evident that the purpose of all Aileen Health AI-generated images needed to be clarified to increase understanding and reduce concern about inefficiency and extra work. Consequently, the AI score was integrated to strengthen the primary and independent role of the Future Multiview file. This approach allows radiologists to obtain the essential information from the Future Multiview file alone, potentially obviating the need to continually review the Future AI Overview file (see section 7.2.2). Furthermore, the AI score aims to optimize and reinforce the assessment process by indicating the likelihood of actual growth in the breast. Presented on a scale from 1 to 10, aligned with participants' preference for whole numbers, this feature received appreciation from all radiologists during the evaluation. The AI score is accompanied by text indicating either low, general, moderate, or high risk to emphasize the message. Previously, the terms BC or No BC were utilized, with BC representing breast cancer. However, this terminology was revised due to participants' difficulty comprehending the message.

7.2.2 Future AI Overview

Another component of the design solution is Future AI Overview, which clarifies the rationale behind the AI's assessment. After reviewing the Future Multiview, users should proceed to this overview if they desire more comprehensive insights into the suspicious area and the AI's assessment (see Figure 37).

Figure 37

The final design proposal for the Future AI Overview

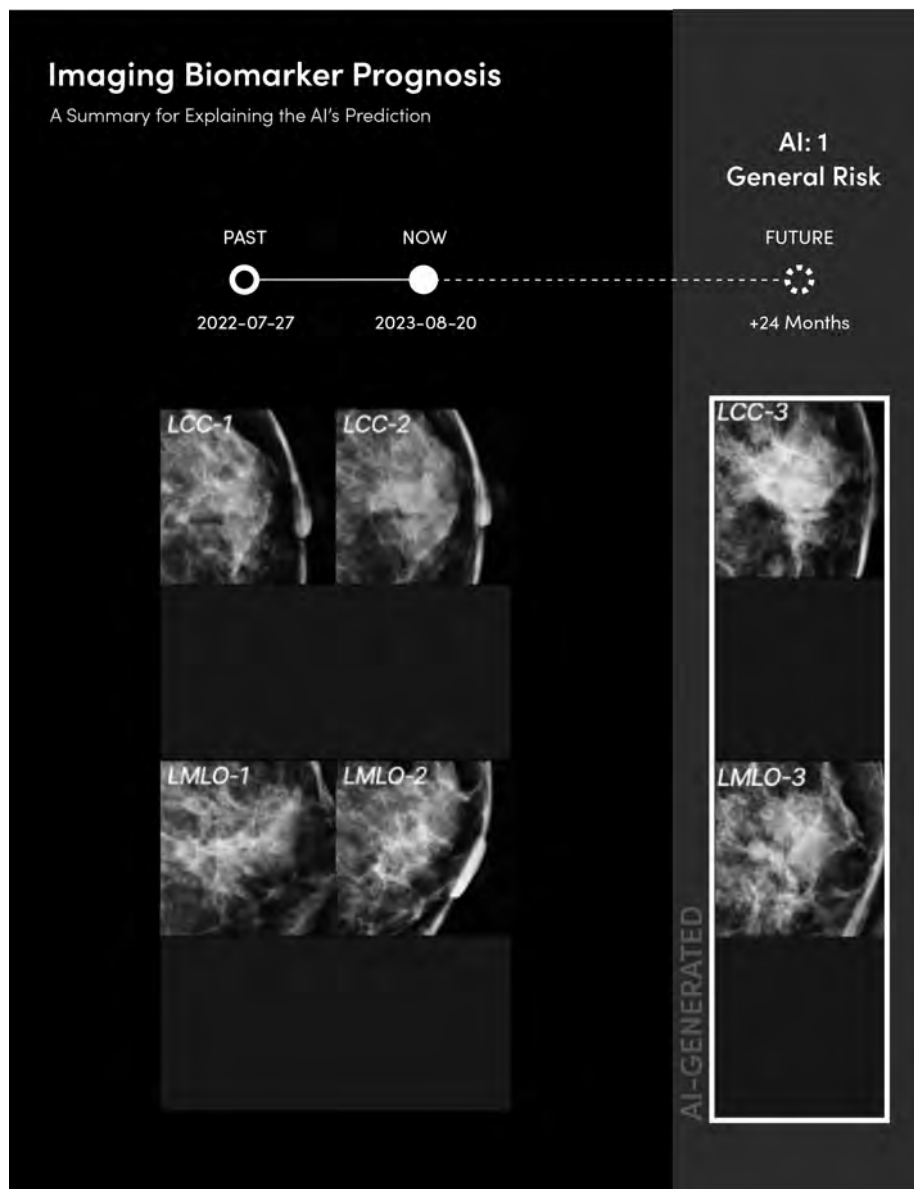


Similar to the Future Multiview, it features a headline to assist users in quickly identifying the content they are viewing. Initially labeled *Future Overview*, the headline was refined to *Imaging Biomarker Prognosis*, aiming for enhanced clarity regarding the file's intended purpose. The overview also contains a timeline where each point on the line represents a mammography examination. To denote when the examination was conducted, the terms *Past*, *Now*, and *Future* are used in conjunction with the date of the examination. This distinction is further emphasized through different circles: unfilled for past examinations, filled for current ones, and dashed for future ones. The information is presented horizontally to align with users' mental models, where the oldest image is typically on the far left and the newest on the far right.

Under each point on the timeline, there are zoomed-in images of the area in the breast that the AI has flagged as suspicious, captured from two different angles. The ability to compare breast images from different times is crucial for detecting changes or deviations over time. A view with biomarkers has also been added to make any changes in the breast more apparent. The purpose of these biomarkers is to provide quantitative information about the progression and outcome. Furthermore, a dark gray box and a text field with the phrase *AI-Generated* have been incorporated to distinguish AI-generated information from non-AI-generated content. Additionally, when the AI detects a risk of cancer, various elements are highlighted in orange, similar to multiview. Otherwise, white is employed when no or low risk is detected, as seen in Figure 38. This color system aims to enhance the assessment's accuracy and efficiency, enabling users to promptly identify potential risks.

Figure 38

The Future AI Overview when the risk of cancer is low



7.2.3 Landing Page

Many of the identified requirements and guidelines are related to background information, such as founders and training data. To meet these requirements and guidelines, a landing page has been developed to facilitate a smooth onboarding process. The landing page is a website intended to provide radiologists with the necessary information before they undergo training to learn the system. The design of the landing page is based on the hierarchy of needs and the compiled list of requirements and guidelines (see Table 4, 5, 6 & 7).

The first sections of the landing page can be seen in Figure 39. A clear headline at the top of the page sparks interest and indicates that the user is in the right place. Below the headline, a brief body text is presented in the first-person plural form to create a sense of community and collaboration with Aileen. The text aims to introduce the product while communicating a sense of participation and the opportunity to make a difference together, meeting the need for belongingness. Alongside the text, a 40-second video illustrates the product, offering an alternative means of information retrieval. This video is particularly useful for radiologists who prefer visual content or multiple presentation formats. Below the text, three benefits of the product are then presented to fulfill the need for esteem in the hierarchy of needs. These benefits aim to demonstrate how the product can improve the quality of life for radiologists in their work. Following this, the workflow process is presented step by step to reduce confusion and increase clarity in the workflow.

Figure 39

The first sections of the landing page

Aileen Health

Generative AI as Decision Support in Breast Cancer Screenings

Up to 30 percent of breast cancer cases emerge between biannual screenings, often at more advanced stages. With the help of Aileen Foresight AI, we can detect the early signs that causes those cancers.

Introduction to Aileen Foresight

30%
of breast cancer cases emerge between biannual screenings, often at more advanced stages

Why Aileen?

- AI Support**
Assisting AI for enhanced radiologists' decision-making.
- Reduced Recall Rates**
Decision support for fewer false positive recalls.
- Preventative Actions**
Shift healthcare focus from diagnostic to prevention.

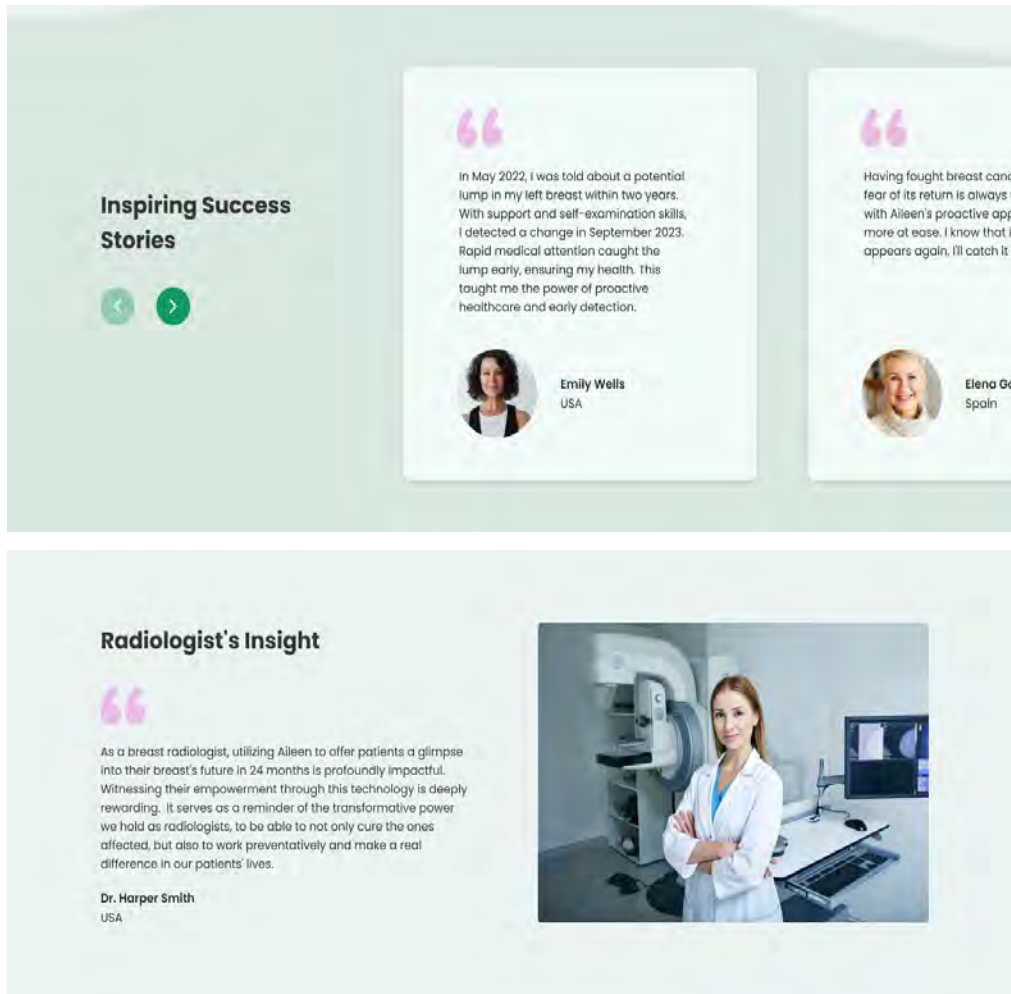
How it works in 6 steps

- Step 1**
The woman attends the mammography screening.
- Step 2**
Breast images are saved and sent to the radiologists.
- Step 3**
Based on the images, the AI generates a visualization of a 24 months prediction together with an estimation.
- Step 4**
Radiologists decide how to proceed with the patient based on the provided information.
- Step 5**
If predicted risk of cancer, the woman gets invited to primary care for breast self-exam instructions.
- Step 6**
When the woman notices symptoms, she attends a diagnostic mammography.

Furthermore, social proof is demonstrated through reviews from patients and radiologists to satisfy the need for safety, as seen in Figure 40. This addition reinforces the sense of shared experience and fosters confidence in making the correct decision. The information was added after the evaluation with radiologists, during which they expressed a desire to know how patients responded to the idea of an AI-generated future prognosis.

Figure 40

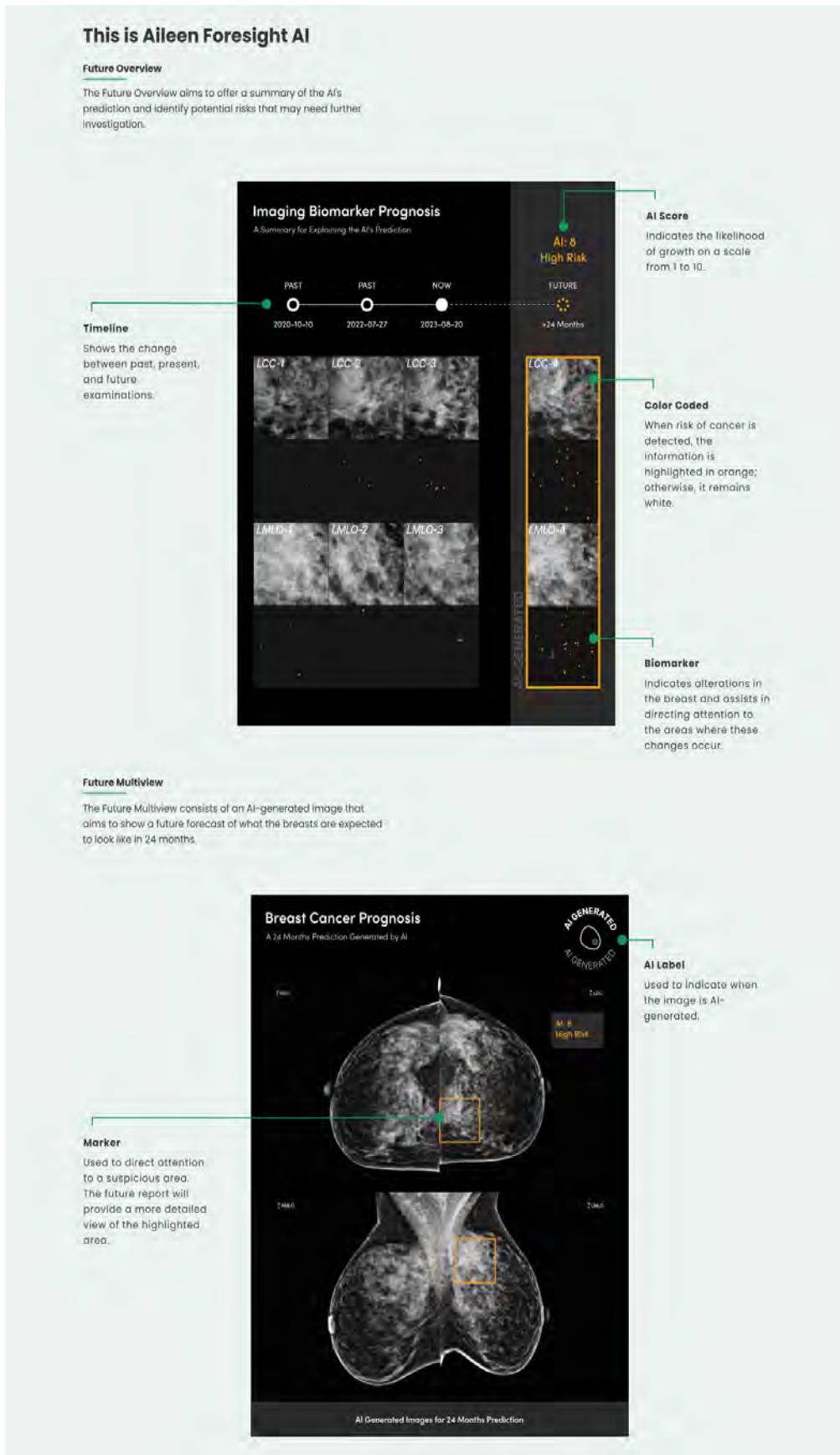
The sections of the landing page with social proofs



The landing page also describes the product's features, as seen in Figure 41. The information aims to communicate how these features can address customers' fundamental needs, enhancing their motivation to purchase or utilize the product. Furthermore, the page strives to treat visitors with transparency, respect, and care for their needs and preferences, facilitating their journey toward self-actualization. Feeling valued and understood is crucial for establishing trust in both the brand and its product.

Figure 41

The section of the landing page highlighting the product's features



Furthermore, selecting appropriate colors for the landing page was crucial as it could influence mood and emotions, navigating visitors into a specific mindset. Considering that the brand can also significantly contribute to trust, the decision was made to align the landing page with Aileen Health’s graphic profile, seen in Figure 42. This means that colors, shapes, and fonts have been selected accordingly.

Figure 42

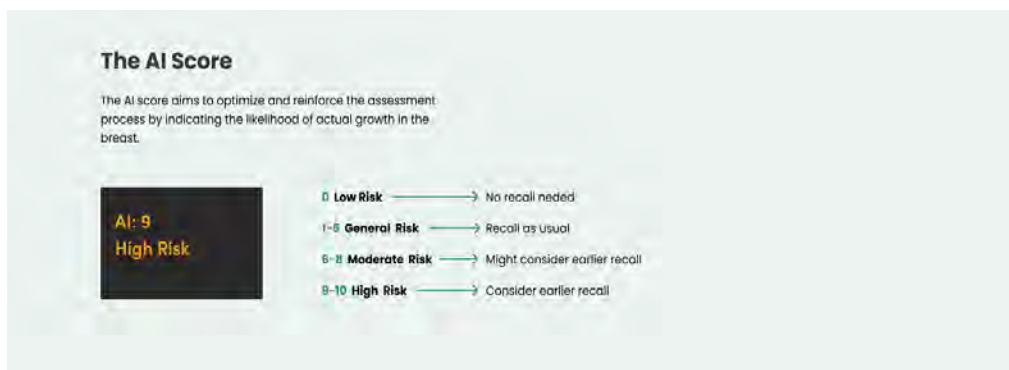
The graphical profile of Aileen Health



As previously stated, both the Future Multiview and Future AI Overview provide an AI score to indicate the likelihood of future breast abnormalities. This score should also assist in determining the timing of the patient’s following recall, providing a more precise understanding of the different score levels. No recall is necessary if the AI score is 0, indicating low risk. Scores between 1 and 5 suggest a general risk level, recommending the usual recall time. Considering an earlier recall may be beneficial for scores falling between 6 and 8, indicating moderate risk. Scores of 9 and 10, indicating significant risk, initiate an earlier recall. However, since the radiologists in the study preferred to receive only ratings without actionable suggestions, this detailed information is not incorporated into the Future Multiview or Future AI Overview files. Instead, it is presented on the landing page to acknowledge its existence while emphasizing that radiologists maintain ultimate control in deciding the course of action (see Figure 43).

Figure 43

The AI score and the following recommendations are explained on the landing page

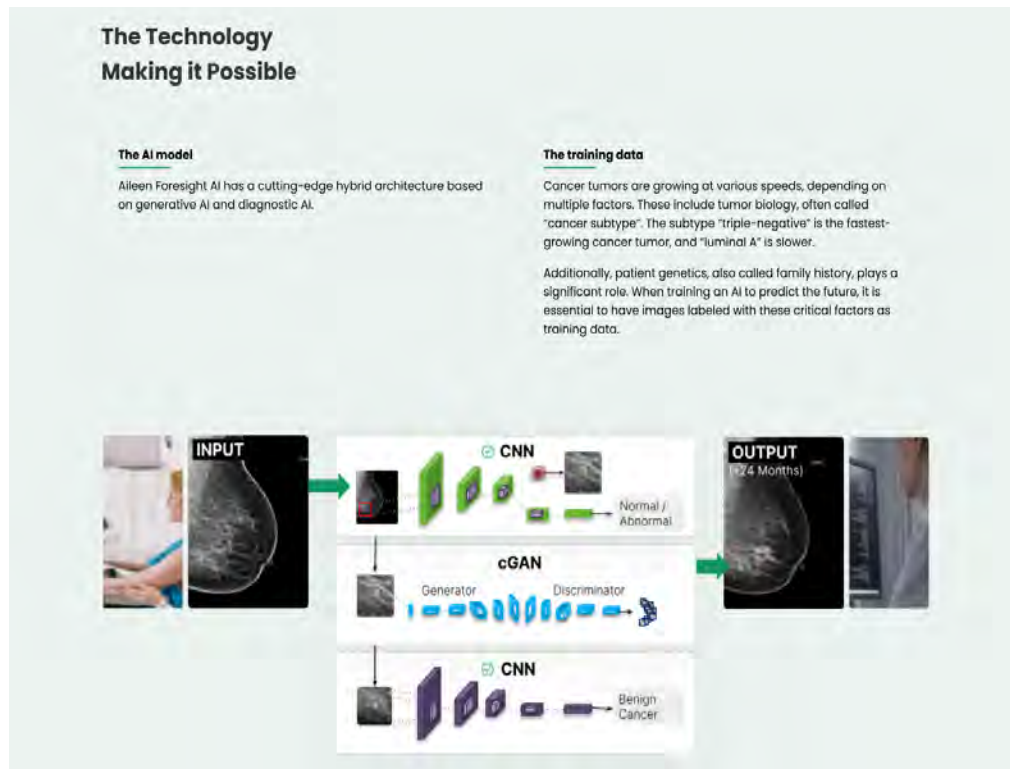


When establishing trust in AI, it is also crucial to consider the explainability of the algorithms’ decisions. Therefore, an explanatory visualization of how the AI is structured has been developed and added to the landing page to illustrate the process clearly (see

Figure 44). The image is intended to complement and clarify how the AI operates, how it has been trained, and the data it utilizes. Hopefully, this visualization can bridge the gap between the more complex and technical aspects of AI and the less AI-experienced radiologists. Furthermore, by ensuring transparency, users remain in the control loop, a key aspect for building trust (Rahwan, 2017).

Figure 44

A visual explanation of the AI model



Finally, the landing page also contains information about the company's founders and their previous experience (see Figure 45). Including such information adds credibility by highlighting their knowledge. It also establishes a personal connection with visitors, declaring the passion and expertise driving the brand.

Figure 45

The section presenting the company's founders

Management Team

66

From the beginning, patients drive my motivation. The founding team had one in situ case, two breast cancer cases and one double mastectomy.

Anna Tidstam, Founder and CEO
Sweden

Founder, CEO
Anna Tidstam, PhD, EMBA
- 15 years of AI experience
- Radiology AI vendor experience in EU+ USA

Founder, COO
Marzieh Zare, PhD
- AI neuroscientist who became entrepreneur
- Top 20 women in AI in North America

Supporters and Partners

LAND SALZBURG

INNOVATION SALZBURG
Shaping the Future

austria wirtschafts service

aws

AI SWEDEN

8. DISCUSSION

This chapter contains a discussion of the thesis's design process and findings, the validity of results, ethical considerations, and prospects for future research.

8.1 Result

This section includes a discussion and reflection of the thesis results, including the created requirements and guidelines and the final design proposal.

8.1.1 Requirements

The creation of requirements occurred in two distinct phases: the first part followed an extensive literature review, and the second part was based on insights gained from the usability tests. These requirements and guidelines were structured into four categories; usability, trust, motivation, and technical. The categorization aimed to provide an overview and understanding of the different areas covered by the specification. It also reduced the risk of important requirements and guidelines being overlooked or forgotten. While these requirements are based on a literature review and insights gathered from radiologists, it is important to acknowledge that diverse design solutions may address them. The requirements and guidelines, however, offer a valuable framework for future projects involving the integration of AI into radiology or similar medical domains to create trust for the users.

Furthermore, the formulation of the requirements and guidelines could have been influenced differently if the study had a larger pool of participants in the usability test. Only three participants, all from Sweden, participated, and the response did not achieve saturation since the answers were relatively broad. Therefore, including more participants could have altered the resulting requirements and guidelines.

8.1.2 Design Proposal

One of the thesis objectives was to develop a design proposal for Aileen Health that visually illustrates the creation of a design aligned with the specified requirements and guidelines. These designs aim to be tailored to the needs of radiologists, creating a sense of comprehension and trust in the system. One of the primary challenges was facilitating radiologists' understanding of AI-generated images and fostering confidence in their decision-making based on the insights provided. Therefore, the Future Multiview and the Future AI Overview aimed to address these challenges. Additionally, the landing page was designed to offer radiologists comprehensive information about the product, providing them with knowledge and insights before its implementation, thereby further encouraging trust in the system.

The usability tests with radiologists showed that the design varied in its alignment with the specified requirements and guidelines. Since guidelines are subjective and can be interpreted differently by various users, they are also challenging to quantify. While some aspects appeared to be fulfilled during testing, others still require additional design iterations and refinement to meet the specifications. For instance, guideline 3.1, which focuses on encouraging radiologists to act upon AI-generated information, requires further design iterations. Particularly with a larger sample of radiologists to identify effective strategies for motivating action based on AI-generated information.

Throughout the design process, several challenges emerged. One notable design constraint stemmed from the prototype's reliance on the DICOM format, which has a fixed format and lacks interactivity. Consequently, the scope for design exploration was considerably constrained. Several potential design suggestions were left unexplored due to the limitations of the format. Additionally, the initial evaluation was conducted using a demo with an existing design, potentially constraining the creative freedom and hindering the extent of innovation in the design. A design developed entirely from scratch might have led to different results. Starting from square one and not basing the design on the existing demo might have facilitated the exploration of alternative design directions and encouraged the emergence of new approaches. On the other hand, this approach provided a solid starting point for the evaluation process.

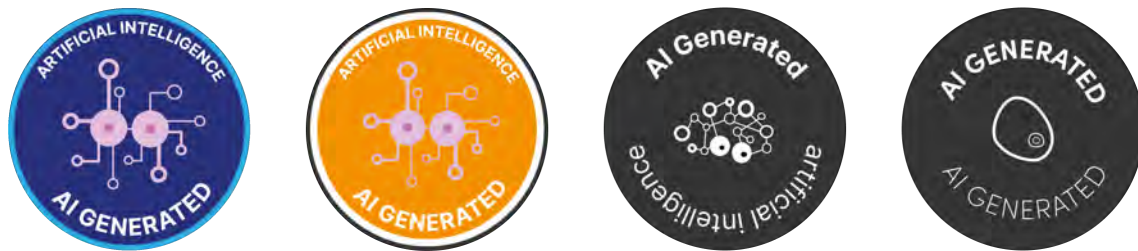
Throughout the thesis, the prototypes have been designed considering the computer system used at Unilabs, ensuring compatibility and resonance with its features. This decision aligns with the guideline 1.4 (see Table 4). However, it is crucial to acknowledge that these systems may vary across different countries. Therefore, the design might not seamlessly integrate with other systems or align with the mental models of radiologists who are used to another system. An additional challenge throughout the design phase revolved around finding a balance between introducing new elements to the interface and preserving its familiar essence. For instance, incorporating color coding to distinguish between different outcomes was one of the challenges, as the system employed by radiologists at Unilabs does not utilize color. This led to difficulties determining what elements could be effectively incorporated without disrupting the radiologist's workflow.

Moreover, another key objective was to align the system with the mental models of radiologists, presenting another challenge. The product needed to find a balance: while it should draw attention to alert radiologists when cancer risk is detected, it should avoid overwhelming or distracting them from the X-ray images. Furthermore, ensuring that the design was perceived as professional and avoiding any perception of being childish was crucial. This was particularly evident in the design of elements such as the AI stamp, where maintaining a professional aesthetic was challenging. For instance, initially, a colorful stamp was created to draw attention and inform radiologists that the image was AI-generated. However, the last evaluation indicated that a black-and-white stamp was enough to communicate this information clearly to the radiologists while also ensuring compatibility with their mental models. The design process of the stamp can be seen in Figure 46.

In addition, integrating additional elements without disturbing the X-ray images was another challenge. For instance, while elements such as labeling the image as AI-generated needed to be prominent enough to prevent confusion with authentic X-ray images, they should not obstruct the radiologists' diagnostic process.

Figure 46

The stamp's transformation from its initial state at the start of the design process to its final form, progressing from left to right



Furthermore, fostering trust in the system for the radiologists leads to an additional challenge. Based on the findings from the last usability test, radiologists' current level of trust toward the product appears to be relatively low. Trust is influenced by many factors, among which brand identity plays a significant role (Michler et al., 2019). Users tend to place greater trust in products associated with brands they believe in. Consequently, several design proposals incorporated the brand logotype, although these did not receive widespread approval during the evaluation. This could be due to the brand's current lack of recognition, as Aileen Health is presently in the process of being launched and has not yet established its brand identity. Nevertheless, it was decided to continue including the brand logo within the AI stamp, anticipating increased brand recognition post-launch and implementation. Furthermore, to build trust in Aileen Health, confidence in the company's future prognosis is essential. Maintaining this trust necessitates transparency about the creators' expertise, providing insight into their ability to deliver reliable predictions (Gönül et al., 2012). Additionally, understanding the driving forces and objectives behind the company's progress is crucial. Therefore, the landing page should incorporate information regarding the Aileen Health team's backgrounds and motivations for founding the company, with the aim of increasing the user's trust in the product.

In Chapter 2, a study by de Vries et al. (2022) concerning radiologists' attitudes toward implementing AI in screening programs revealed a generally positive perspective on AI integration. These findings differ from the results of this thesis, where participants had more negative attitudes toward AI implementation, as reflected in the relatively negative responses on the Hesselgren Emotional Scale (see Figure 34). Participants reported low expectations of the product, coupled with feelings of fear. Several factors could explain this. Firstly, de Vries et al.'s study (2022) had a larger sample size of radiologists, which may have influenced the outcomes. Additionally, the AI they presented acted as a support tool during the diagnostic process, while Aileen Health aims to use the AI to generate

future predictions. Radiologists might be less familiar with AI for future predictions than for diagnostic support, potentially affecting their attitudes toward Aileen Health's product. Moreover, since attitudes toward AI are highly subjective, the personal views of the radiologists might have influenced the evaluation outcomes. However, it is worth noting that these attitudes may not have stemmed directly from the product's design but rather from the broader integration of AI into their workflow, which can evoke concern and anxiety. These attitudes may lead to doubt regarding the safety and accuracy of AI, concerns about increased workload, and worries about the potential for AI to replace radiologists. In essence, many emotions associated with AI adoption could have impacted the evaluation results beyond the scope of the interface design itself.

Furthermore, the intended users of this design are radiologists, characterized as expert users within their current system. The objective of Aileen Health's product is to integrate into the daily workflow of radiologists, rather than being utilized occasionally. Therefore, immediate intuitiveness may not be as critical as ensuring good learnability, allowing users to learn the system efficiently without the need for constant re-learning. For instance, during the evaluation, the radiologists did not grasp the color coding for low and high-risk cancer. However, they might find the colors helpful after becoming more familiar with the system and understanding the purpose behind the colors. Nevertheless, to ensure good learnability, further testing of the system integrated into the radiologists workflow is essential. However, intuitive design remains important to foster user confidence and prevent users from being intimidated of the system. If the system is difficult to comprehend at first glance, it may discourage users from engaging with it.

8.2 Process and Execution

The project progressed mainly as anticipated and followed the planned process. The thesis started with a thorough literature review covering aspects relevant to the project. It later proceeded with dedicated time to gain an understanding of the product demo created by Aileen Health. These initial parts were conducted within the empathize phase, during which there was also an opportunity to visit Unilabs in Gothenburg, gaining valuable insights into the radiologist's operational workflow. These insights were vital since they gave a deeper understanding of the radiologists' work and the product. Moreover, the literature review continued during the project's later stages.

As radiology is an area of expertise that requires skills not previously possessed by the thesis team, considerable time was dedicated to familiarizing with radiology and AI at the beginning of the process. Understanding the subject matter was essential for conducting thorough evaluations with radiologists, ensuring effective communication by speaking their professional language, and understanding specific terminology for accurate assessment. Additionally, understanding the reasoning and workflow of radiologists was crucial for informed design decisions. Consequently, a substantial amount of time was invested in comprehending the workflow of radiologists and the process of mammography screenings. Additionally, there were several technical aspects to consider at the beginning of the project, such as configuring Aileen Health's product demo on the computers,

accessing datasets, and debugging code in the Python script. These challenges consumed considerable time during the empathize phase. In hindsight, using a Figma prototype could have produced equivalent results while saving considerable project time.

While still in the empathize phase, the initial assessment of the product demo was conducted in collaboration with radiologists. This phase presented several challenges. Firstly, accessing radiologists was difficult due to their demanding schedules and limited availability. This issue is due to the lack of breast radiologists, with only 71 in Sweden, all of whom are burdened with heavy workloads. Consequently, engaging radiologists to participate in usability testing posed a significant challenge. Furthermore, during the initial usability test, time constraints caused radiologists to participate in the evaluation collectively rather than individually. This group dynamic may have influenced their perceptions of the design, potentially leading to conscious or unconscious alignment of responses. Hence, the initial usability test outcomes may have been subject to bias and could have differed if evaluations had been conducted individually.

During the initial usability test, participants encountered challenges comprehending the assigned tasks for evaluating the product's design. Their struggle to overlook technical constraints such as screen resolutions and sizes hindered their ability to focus on assessing the design beyond the X-ray images. Additionally, radiologists expressed doubtfulness regarding the product's feasibility, which may have decreased their motivation to contribute to its improvement. Their hesitation might have partly come from apprehensions about AI implementation in the workplace, fearing loss of control and potential replacement. Consequently, preconceived attitudes toward AI likely influenced the quality of responses and engagement during the usability test. Nevertheless, despite these challenges, the usability test generated numerous valuable insights for refining Aileen Health's product demo and confirmed the findings from the literature review.

Furthermore, the later evaluation, mainly through the Hesselgren emotional scale, revealed some negative sentiments among participants regarding the product. This observation bears significance as the emotions a product evokes profoundly influence trust in the product and user motivation (Hekkert, 2006). Consequently, future iterations need to delve deeper into these emotional responses evoked by the product and create strategies to reduce feelings of sadness, fear, and anger, thereby enhancing the user-product interaction. The negative sentiments expressed by radiologists towards the product might be explained through the AIDUA model. This model consists of two appraisals and two consequent outcomes, differentiating the decision-making process regarding willingness or objection to use AI. Both appraisals must be satisfied to establish a willingness to utilize the AI. In the first appraisal, users evaluate the significance of the AI by considering factors such as social influence, pleasure, and human-like qualities (Kelly et al., 2023). Given the absence of prior experience with AI among the participants, alongside concerns about potential increases in workload, it is possible that the conditions of the first appraisal still need to be met. Consequently, the second appraisal, focusing on design and usability considerations, may have also remained unfulfilled. Thus, it becomes evident that more than solely focusing on design and usability may be needed to build acceptance among radiologists. Instead, a

holistic approach of factors within the initial appraisal is important to establish the user's trust in the product.

8.3 Validity and Generalization of the Results

The usability tests provided valuable insights incorporated into the requirements, guidelines, and design proposal. However, the validity of these results is subject to discussion due to the participation of only three radiologists in the evaluation. Considering the shortage of breast radiologists in Sweden, these three participants represent approximately 4.2 percent of the population. Nevertheless, due to the small sample size, a generalization of requirements or guidelines cannot be drawn without additional testing.

Furthermore, all three participants are employed at Unilabs in Gothenburg, potentially influencing each other's opinions or introducing bias. A more diverse participant pool from various workplaces with different routines and workflows would have likely resulted in a broader spectrum of responses. For example, including radiologists from a workplace where AI has been implemented might have led to other results. Furthermore, as all participants were from Sweden, this may have resulted in a design that is applicable to Sweden but less so to other countries. It is important to note that cultural differences in the perception of color symbolism, hierarchy, and other design elements could account for this variation. What is distinctly clear in one culture may be ambiguous or hold entirely different meanings in another.

Nevertheless, the objective was to craft requirements and guidelines that are broad enough to apply to similar projects in other medical domains implementing AI. When formulating the requirements and guidelines, a decision was made to keep them broad, with design elements such as colors and typography left unspecified. Instead, the focus remained on usability and fostering trust in a general sense, with the aim of appealing to a wider target audience where specific details could be tailored to individual target groups and cultural preferences. Thus, these requirements are inherently more adaptable and generalizable. However, since the design proposal is more closely aligned with Aileen Health and customized to the specific brand identity, its applicability to other products or brands may be limited.

8.4 Ethical Considerations

Throughout this thesis, conscientious efforts have been made to uphold ethical standards on multiple fronts. Firstly, anonymity protocols were enforced for all participants involved in the evaluation process. Furthermore, to protect patient privacy, all identifiable information, such as names and birth dates, was systematically replaced with randomly generated data using Python scripting, particularly those containing authentic breast X-ray images. This process ensured that patients from the dataset identities remained anonymous.

Moreover, there are ethical considerations concerning how radiologists perceive the product and its limitations in response to AI cancer risk predictions. Regulations governing healthcare staff vary across countries. In Sweden, for instance, radiologists are restricted from altering the screening program process. For instance, if the current mammography indicates no signs of cancer, but the AI predicts a risk of cancer within the subsequent 24 months, radiologists are not permitted to schedule an earlier mammography examination. Instead, they must adhere to the standard two-year interval. Patients are only eligible for examination upon detecting a breast lump or the arising of symptoms. This restriction can evoke profound frustration and stress among radiologists who recognize potential dangers to patients but cannot intervene. Additionally, this predicted cancer reply can also cause patients much stress and anxiety, especially if the healthcare system does not allow earlier exams.

8.5 Future Work

The thesis is part of a more extensive project of the medical startup Aileen Health. The startup is in the process of being launched, and the product is under development. There will be many future design iterations, and the thesis is part of the first. Currently, the thesis's design proposals are a part of the product demo at Aileen Health and will be used in further product development. For future work, the design needs to be tested with more radiologists. Additionally, it would be preferable to test with radiologists who work at various workplaces and have different experiences to get a more extensive spread of opinions. Furthermore, the design needs to be tested in practice. This evaluation is essential for future work, ensuring the design meets all requirements and guidelines. Testing is also necessary to confirm that the design does not increase mental workload or time consumption and to understand its impact on the workflow. Moreover, it would be interesting to explore whether radiologists' prior experience with AI influences their perspectives on the product and design in upcoming usability tests. It is essential to understand what specific information needs to be incorporated to create trust and reduce the gap between radiologists with different levels of AI experience. Furthermore, investigating whether opinions will shift after Aileen Health has launched and established itself could provide valuable insights by confirming the brand's impact on trust.

9. CONCLUSION

This thesis was conducted in collaboration with the TechMed startup Aileen Health. The aim has been to identify key factors influencing radiologists' trust in an AI system that will be implemented into their workflow. The thesis has been conducted using a Research through Design approach to create requirements and guidelines for designers of an AI system that aims to be applicable to both radiology and other similar medical fields. An additional goal was also to develop a design proposal, based on the requirements and guidelines for Aileen Health, of the AI-generated future prognosis to be understandable and trustworthy for breast radiologists. More specifically, the thesis has aimed to answer the following research question:

What are key factors for trusting and integrating AI-based predictions into the current mammography screening workflow to support radiologists?

The research question was examined using an iterative design process with multiple iterations. The process involved five distinct phases; empathize, define, ideate, prototype, and test. However, the process has not been done sequentially but repeated multiple times. The project started with an extensive literature review. Based on the review, the first part of the thesis's result, guidelines, and requirements were formulated. Afterward, an evaluation and usability test of the existing demo of the AI system was conducted. Based on the evaluation, additional requirements and guidelines were added to the list. In total, two requirements and twelve guidelines were developed. These were divided into four distinct categories:

Usability - To ensure that the design is intuitive and easy to understand for the users. Key factors here are ensuring ease of understanding, minimal workload, alignment with user expectations, consistency within the system, compatibility with various platforms, and a clear indication of AI-generated images. More specifically, as an example, the added elements should attract attention without disrupting the assessment of the image.

Trust - To establish trust in the design. Key factors here are providing users with sufficient control, access to comprehensive information about the AI system, and transparency by offering details about the product. More specifically, as an example, the motives and purpose of the company behind the AI system should be clearly stated.

Motivation - To generate motivation for using the product and taking action based on the provided information. Key factors here are to encourage radiologists to act on the AI-generated information, ensuring efficiency in its usage and avoiding any perception of competition with radiologists. More specifically, for instance, by clearly explaining the product's purpose and how it can positively impact both their workflow and patients.

Technical - To ensure that the design is technically feasible and capable of being implemented in the final product. Key factors here are the ability to generate the desired design using the company's program languages and ensuring that the resulting images have high resolution. More specifically, as an example, to ensure that the quality of the images does not compromise the quality of the radiologists' assessments.

Drawing from the requirements and guidelines created, a design proposal was created for Aileen Health. The objective of these proposals was to visualize examples demonstrating how AI-generated content can be designed to foster trust and enhance usability for radiologists. The design proposal included a Future Multiview, a Future AI Overview, and a landing page for Aileen Health. Subsequently, the design proposals were evaluated through usability testing with breast radiologists at Unilabs in Gothenburg. The usability test provided valuable insights for advancing the design process. Findings indicated a notable enhancement in product usability compared to the initial product demo of Aileen Health, with radiologists demonstrating a clear understanding of the displayed information. Despite this improvement, the results also showed that radiologists still had a quite negative attitude toward the product of Aileen Health.

Subsequently, the final design proposal was formed based on the insights gained from the usability tests. When designing the final design proposal, the aim was a strong focus on clarifying which part of the images were the future prognosis and AI-generated, what level of risk the AI estimated, and adding information about patient perspective to the landing page to establish more trust. Nevertheless, as discussed in Chapter 8, it is crucial to recognize that many factors beyond design influence trust in an AI system. This includes, for instance, pre-existing attitudes towards AI, adherence to national guidelines, and confidence in the product brand. Consequently, design requirements and guidelines should be considered together with external factors for comprehensive trust establishment. Moreover, conducting tests with a more significant number of radiologists is essential for future iterations to get a deeper understanding of key factors of trust-building and motivations for AI implementation.

In summary, this thesis contributes insights into implementing AI applications within healthcare systems by highlighting crucial factors when designing visual representations of AI-generated content. The thesis has focused on investigating interactions between radiologists and AI systems and conducting a literature review. Through this methodology, key factors that promote trust and facilitate the implementation of these systems have been identified. These factors have been transformed into requirements, guidelines, and a detailed design proposal to address the research question. The design proposal is now a part of Aileen Health's product demo and will be used in further product development. Hopefully, the insights can contribute to the transition from a reactive to a proactive healthcare system by establishing trust in the AI system for radiologists. The thesis can also inspire other areas within the healthcare system to enhance the comprehensibility of AI-generated information, promoting more extensive utilization and potentially saving resources and lives.

REFERENCES

1177.se. (2022, October 31). *Mammografi*. 1177.se. Retrieved 2024-01-16 from <https://www.1177.se/Vastra-Gotaland/undersokning-behandling/undersokningar-och-provtagning/upptack-och-forebygg-cancer/mammografi/>

1177.se. (2023, March 1). *Bröstcancer*. 1177.se. Retrieved 2024-01-16 from <https://www.1177.se/Vastra-Gotaland/sjukdomar--besvar/cancer/cancerformer/brostcancer/>

Andersson, J. (2010). *A conceptual model for analysis of automation usability problems in control room settings*. Chalmers Tekniska Högskola. <https://publications.lib.chalmers.se/records/fulltext/121920.pdf>

Arechar, A. Epstein, Z., & Rand, D. (2023). *What label should be applied to content produced by generative AI?* <https://doi.org/10.31234/osf.io/v4mfz>

Avison, D. E., Lau, F., Myers, M. D., & Nielsen, P. A. (1999). Action research. *Communications of the ACM*, 42(1), 94-97. <https://doi.org/10.1145/291469.291479>

Aydin Baytaş, M. (2022, July 08). *The Origin and Purpose of Research through Design*. Design Disciplin. Retrieved 2024-01-29 from <https://www.designdisciplin.com/rtd-origin/>

Bedué, P., & Fritzsche, A. (2022). Can we trust AI? An empirical investigation of trust requirements and guide to successful AI adoption. *Journal of Enterprise Information Management*, 35(2), 530-549. <https://doi.org/10.1108/JEIM-06-2020-0233>

Bligård, L. O. (2012). *Predicting mismatches in user-artefact interaction. Development of an analytical methodology to support design work*. Chalmers Tekniska Högskola.

Bligård, L. O. (2015). *Utvecklingsprocessen ur ett människa-maskinperspektiv*. Chalmers University of Technology. <https://doi.org/10.13140/RG.2.1.1954.4400>

Braun, C. C., Mine, P. B., & Clayton Silver, N. (1995). The influence of color on warning label perceptions. *International Journal of Industrial Ergonomics*, 15(3), 179-187. [https://doi.org/10.1016/0169-8141\(94\)00036-3](https://doi.org/10.1016/0169-8141(94)00036-3)

Cambridge Dictionary. (n.d). *ambiguity*. Cambridge Dictionary. <https://dictionary.cambridge.org/dictionary/english/ambiguity>

Cambridge Dictionary. (n.d). *symbol*. Cambridge Dictionary.
<https://dictionary.cambridge.org/dictionary/english/symbol>

Cardona Cervantes, G. (2024-03-19). Så upptäcks cancer med AI – svenska företaget först i världen. *Tidningen Näringslivet*.
<https://www.tn.se/entreprenorskap/36406/sa-upptacks-cancer-med-ai-svenska-foretaget-forst-i-varlden/>

Cooper, A., Reimann, R., & Cronin, D. (2014). *About Face: The Essentials of Interaction Design* (4th ed.). John Wiley And Sons.

de Vries, C. F., Colosimo, S. J., Boyle, M., Lip, G., Anderson, L. A., Staff, R. T., Harrison, D., Black, C., Murray, A., Wilde, K., Blackwood, J. D., Butterly, C., Zurowski, J., Eilbeck, J., & McSkimming, C. (2022). AI in breast screening mammography: breast screening readers' perspectives. *Insights into Imaging*, 13(1).
<https://doi.org/10.1186/s13244-022-01322-4>

Ganta, V. C. (2014). Motivation in the workplace to improve the employee performance. *International Journal of Engineering Technology, Management and Applied Sciences*, 2(6), 221-230.

Google. (n.d). *Crazy 8's*. Design Sprints.
<https://designsprintkit.withgoogle.com/methodology/phase3-sketch/>

Guimaraes e Equipe Aela, F. (2022, July 29). *Benchmarking: How To Conduct One?*. Aela.
<https://aelaschool.com/en/strategy/benchmarking-how-to-conduct-one/>

Gönül, M. S., Önköl, D., & Goodwin, P. (2012). Why should I trust your forecasts?. *Foresight: The International Journal of Applied Forecasting*, 27, 5-9.

Hanington, B., & Martin, B. (2019). *Universal methods of design expanded and revised: 125 Ways to research complex problems, develop innovative ideas, and design effective solutions*. Rockport publishers.

Hekkert, P. (2006). Design aesthetics: principles of pleasure in design. *Psychology science*, 48(2), 157.

Hellmark, C. (1994). *Typografisk handbok*. Ordfront & Ytterlids

Hisarciklilar, O., & Boujut, J. F. (2009). An annotation model to reduce ambiguity in design communication. *Research in Engineering Design*, 20(3), 171-184.
<https://doi.org/10.1007/s00163-009-0073-6>

Hovda, T. (2022). *Improved breast cancer screening for women and society: A radiological approach*. Pubmed.

iCAD. (n.d). *ProFound AI® Risk**. iCAD.
<https://www.icadmed.com/wp-content/uploads/2023/11/DMM272-ProFound-AI-Risk-Rev-7.pdf>

Interaction Design Foundation. (2016, May 25). *What is Design Thinking?*. The Interaction Design Foundation.
<https://www.interaction-design.org/literature/topics/design-thinking>

Interaction Design Foundation. (2023). *Design Iteration Brings Powerful Results. So, Do It Again Designer!*. The Interaction Design Foundation. Retrieved 2024-01-29 from
<https://www.interaction-design.org/literature/article/design-iteration-brings-powerful-results-so-do-it-again-designer>

International Organization for Standardization. (2018). *Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts* (ISO Standard No. 9241-11:2018).
<https://www.iso.org/obp/ui/en/#iso:std:iso:9241:-11:ed-2:v1:en>

Jatoi, I., Sung, H., & Jemal, A. (2022). The emergence of the racial disparity in US breast-cancer mortality. *New England Journal of Medicine*, 386(25), 2349-2352.
<https://doi.org/10.1056/NEJMp2200244>

Jordan, P. W. (1993). *Consistency and usability* (Doctoral dissertation, University of Glasgow). Archive.
<https://theses.gla.ac.uk/5577/>

KARMA. (n.d). *Varför är forskning inom bröstcancer viktig?* KARMA.
<https://karmaprojektet.se/vad-ar-karma-2/>

Kelly, S., Kaye, S. A., & Oviedo-Trespalacios, O. (2023). What factors contribute to the acceptance of artificial intelligence? A systematic review. *Telematics and Informatics*, 77, 101925.
<https://doi.org/10.1016/j.tele.2022.101925>

Kim, J., Giroux, M., & Lee, J. C. (2021). When do you trust AI? The effect of number presentation detail on consumer trust and acceptance of AI recommendations. *Psychology & Marketing*, 38(7), 1140-1155.
<https://doi.org/10.1002/mar.21498>

Kunskapsbanken. (n.d). *Primär prevention*. Cancercentrum
<https://kunskapsbanken.cancercentrum.se/diagnoser/brostcancer/vardprogram/primar-prevention/>

- Lunit. (n.d). *Lunit INSIGHT MMG*. Lunit.se. Retrieved 2024-01-20 from <https://www.lunit.io/en/products/mmg>
- McCormick, K. (July 13, 2023). *The Complete Guide to (Persuasive) Landing Pages – With Examples!* WordStream. <https://www.wordstream.com/blog/ws/great-landing-pages>
- Michler, O., Decker, R., & Stummer, C. (2019). To trust or not to trust smart consumer products: a literature review of trust-building factors. *Management Review Quarterly*, 70, 391-420. <https://doi.org/10.1007/s11301-019-00171-8>
- Mohamed, A. A., Luo, Y., Peng, H., Jankowitz, R. C., & Wu, S. (2018). Understanding clinical mammographic breast density assessment: a deep learning perspective. *Journal of digital imaging*, 31, 387-392.
- Munzner, T. (2014). *Visualization analysis and design*. CRC press.
- Norman, D. A. (2013). *The design of everyday things*. Basic Books.
- Norori, N., Hu, Q., Aellen, F. M., Faraci, F. D., & Tzovara, A. (2021). Addressing bias in big data and AI for health care: A call for open science. *Patterns*, 2(10). <https://doi.org/10.1016/j.patter.2021.100347>
- Osvaelder, A-L & Ulfvengren, P. (2015). *Arbete och teknik på människans villkor*. Prentent.
- Plutchik, R. (1984). Emotions: A general psychoevolutionary theory. *Approaches to emotion*, 1984(197-219), 2-4.
- Rahwan, I. (2017). Society-in-the-loop: programming the algorithmic social contract. *Ethics and Information Technology*, 20(1), 5–14. <https://doi.org/10.1007/s10676-017-9430-8>
- Rodríguez-Ruiz, A., Krupinski, E., Mordang, J.-J., Schilling, K., Heywang-Köbrunner, S. H., Sechopoulos, I., & Mann, R. M. (2019). Detection of Breast Cancer with Mammography: Effect of an Artificial Intelligence Support System. *Radiology*, 290(2), 305–314. <https://doi.org/10.1148/radiol.2018181371>
- Romice, O., & Uzzell, D. (2003). Analysing Environmental Experiences. *Espace de vie, Aspects de la relation homme-environnement*, 49-83.
- Screenpoint medical. (n.d). *Transpara*. Screenpoint-medical.com <https://screenpoint-medical.com/transpara-findings/>

Sharp, H., Rogers, Y., & Preece, J. (2019). *Interaction design : Beyond human-computer interaction* (5th ed.). Wiley.

Socialstyrelsen. (2022). *Nationell utvärdering – bröstcancerscreening med mammografi*. Socialstyrelsen.

<https://www.socialstyrelsen.se/globalassets/sharepoint-dokument/artikelkatalog/nationella-riktlinjer/2022-6-7958.pdf>

Teodorescu, I., & Vasile, V. (2015). Landing Pages Features to Attract Customers. *Ovidius University Annals, Economic Sciences Series*, 15(2), 360-363.

Vara. (n.d, a). *We won't rest until every deadly breast cancer is detected early*. Vara.se. Retrieved 2024-01-20 from <https://www.vara.ai/about-us>

Vara. (n.d, b). *Decision Referral: Combining strengths of radiologists and AI*. Vara.se. Retrieved 2024-01-20 from <https://www.vara.ai/decision-referra>

Vara. (n.d, c). *Vara and University of Lübeck kick off first-ever prospective AI study for breast cancer screening in Germany*. Vara.se. Retrieved 2024-01-20 from <https://www.vara.ai/praim>

Unilabs. (n.d). *Vårt företag*. Unilabs. <https://unilabs.se/vårt-företag>

Watzman, S., & Re, M. (2007). *The Human-computer Interaction Handbook*. CRC Press.

Yu, K. H., Beam, A. L., & Kohane, I. S. (2018). Artificial intelligence in healthcare. *Nature biomedical engineering*, 2(10), 719-731.

Zimmerman, J., Forlizzi, J., & Evenson, S. (2007, April). Research through design as a method for interaction design research in HCI. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 493-502). <https://doi.org/10.1145/1240624.1240704>

Zimmerman, J., & Forlizzi, J. (2014). Research through design in HCI. In *Ways of Knowing in HCI* (pp. 167-189). New York, NY: Springer New York.

APPENDIX A

Interview Template for First Visit at Unilabs

Beskrivning

En kort beskrivning av projektet

- Vi kommer spela in ljud, känns det okej?

Inledande frågor

- Hur länge har du jobbat med mammografi?
- Hur många bilder kollar du per dag? Hur länge kollar du på varje bild?
- Ser processen alltid likadan ut när ni kollar på röntgenbilderna?
- Vad har du för relation till och kunskap om AI idag?
- Upplever du att dina kollegor har samma uppfattning och relation?
- Har du jobbat med AI tidigare?

I vilket sammanhang?

Hur kändes det?

Varför?

- Hur ser du på att implementera AI i ditt arbetsflöde idag?

Usability test

Case 1

Du kommer nu få bilder från patienten Maria Johansson senaste mammografiundersökning, vilket innefattar höger och vänster bröst i CC och MLO view. Maria är 52 år gammal.

- Hur hade du diagnostiserat den här patienten och hur hade du gått vidare?
- Vilken åtgärd föreslår du?

Du kommer nu även få tillgång till den AI genererade informationen som kommer ge dig en 24 månaders framtidsprognos. Du kommer få en multiview som jämför vänster och höger bröst. En CC och en MLO på det bröst som förväntas utveckla cancer och en framtidsrapport som presenterar in-zoomade bilder av det specifika området i bröstet där cancer förväntas uppstå, från det första mammografibesöket till det senaste.

- När du navigerar bland bilderna, är det tydligt för dig vilka som är AI genererade och inte?

Varför? Varför inte?

- Vad tror du att den här fyrkanten visar?
- Vad är din spontana tanke när du ser dessa framtidsbilder?
- Påverkar framtidsbilderna ditt tidigare beslut eller diagnos?

Varför? Varför inte?

Case 2

Nu kommer du i stället få ta del av bilder från patienten Malin Eriksson senaste undersökning. Malin är 45 år.

- Hur hade du diagnostiserat den här patienten och hur hade du gått vidare?
- Vilken åtgärd föreslår du?
- Påverkar framtidsbilderna ditt tidigare beslut eller diagnos?

Varför? Varför inte?

Case 3

Du kommer du få se bilder från Marie Fransson senaste undersökning. Marie Fransson är 67 år.

- Hur hade du diagnostiserat den här patienten och hur hade du gått vidare?
- Vilken åtgärd föreslår du?

I det här fallet visade AI:n inga tecken på cancer inom 24 månader, men du fick fortfarande en genererad framtidsbild. Vad tyckte du om det?

- Hade du föredragit att inte få någon bild?

Varför?

Case 4

Nu kommer du få studera bilder från Gun Eklund senaste undersökning. Gun är 50 år. Men tillskillnad från tidigare ser framtidsrapporten lite annorlunda ut. Förutom de inzoomade bilderna tillkommer även en biomarkör och en AI-poäng som ska indikera hur stor sannolikheten är att något faktiskt växer.

- Vad är din spontana tanke?

- Hur tolkar du biomarkören?

Varför?

- Tror du att den hade hjälpt dig i ditt arbete?
- Vad föredrar du, den tidigare varianten med bara en framtidsbild eller den här mer detaljerade vyn?
- Tidigare presenterades informationen vertikalt, men den nu presenteras den istället horisontellt. Vad föredrar du?

Varför?

- Vad hade du tyckt om AI:n också hade ett score på hur sannolikt det är att något faktiskt växer?
- Så om det hade stått AI:10 här, hur hade du tolkat det? Vad tror du att det betyder?
- Vad hade du tyckt om att AI:n också hade föreslagit åtgärder som ultraljud eller biopsi?
- Om AI:n hittar tecken på cancer i ena bröstet, vill du fortfarande få en framtidsprognos och bilder på den andra bröstet?

Varför?

Avslutande frågor efter casen

Det var alla casen, så nu går vi över till lite avslutande frågor.

- Hur skulle du beskriva produkten för din kollega?
- Vad hade fått dig att vilja implementera ett sånt här system till ditt nuvarande arbetsflöde?

Varför?

- Hur mycket teknisk kunskap och bakgrund hade du behövt för att lita på den här AI:n?
- Spelar det någon roll vem som tagit fram den?
- Vad ser du som eventuella hinder?
- Något du saknade?

Varför?

APPENDIX B

Interview Template for Second Visit at Unilabs

En kort beskrivning

- Vi kommer spela in ljud och ta några bilder utan att era ansikten syns, känns det okej?

Del 1: AI genererade bilder

Vi kommer börja med de AI genererade framtidsbilderna och du kommer få svara på några frågor angående dessa. Vi kommer nu visa dig sju bilder och vill att du ska bedöma om bilden är AI genererad eller inte.

- Hur säker är du från skala 1–10?

Varför?

Det är dessa fem som är AI genererade och vi vill nu att du rangordnar efter vilken du tycker är tydligast AI genererad. Tydligast längst till höger och minst tydlig till vänster. Förklara gärna varför du lägger bilderna som du gör.

- Vilken föredrar du?

Varför?

- Vad tycker du om rubrikerna på bilderna?
- Tycker du att rubriken förklarar innehållet?
- I dessa bilder du har sett har orden "AI-genererad" använts, hur tolkar du detta? Vad innebär det för dig?
- Om ordet Artificiell hade använts i stället, hade du tolkat det på ett annat sätt?
- Om syftet är att ordet ska indikera att framtidsprognosen är skapad av en AI, vilket ord är tydligast för dig?

Del 2: Rapport

Tillsammans med en bild som liknar de du precis fått se, kommer du också få en sammanställning över AI:ns bedömning. Tanken är att du ska få en snabb överblick och se om det finns potentiella risker du behöver kolla djupare på. Du kommer nu så se en sådan sammanställning.

- Kan du beskriva hur du tolkar informationen du har framför dig? (Utan cancer)
- Är det något du tycker är otydligt?

Varför?

- Är det någon speciell del du tycker är bra?

Här får du en annan sammanställning. (Med cancer)

- Kan du notera någon skillnad?
- Vad tror du skillnaden betyder?
- När du tittar på sammanställningen, vilken del tolkar du är den AI genererade framtidsbilden?
- Och när du tittar på den här? (Ge sammanställning med rutan)
- Vilken föredrar du?

Sammanställningen innehåller även ett AI score som visas i en skala mellan 1 och 10.

- Föredrar du att AI scoret visas i heltal eller ett mer specifikt tal med två decimaler, till exempel 7.88?

Varför?

Du har nu sett ett alternativ på sammanställningen. Här är ett annat förslag.

- Kan du beskriva hur du tolkar informationen du har framför dig?
- Är det något du tycker är otydligt?

Varför?

- Vilken föredrar du?

Varför?

Del 3: Landing page

Du kommer nu få ta del av ett kort scenario och svara på några frågor. Programvaran Aileen har precis lanserats och Unilabs har bestämt att det ska implementeras in i ditt arbetsflöde. Inför lanseringen får du en länk till den här hemsidan och du ombeds läsa igenom information noggrant inför utbildningen av systemet. Du får nu gärna läsa igenom i lugn och ro.

- Vad är ditt första intryck av sidan?
- Påverkar den här informationen din attityd till systemet?

På vilket sätt?

- Hur ställer du dig till att få information på det här sättet?
- Hade du hellre velat ha informationen på ett annat sätt? Alternativt i kombination med annat format.
- Har du något förslag på vad?
- Var det någon del på sidan som du tyckte bättre om?
- Upplever du att någon information är otydlig?
- Upplever du att någon information saknas?

Del 4: Allmänna frågor

Du kommer nu få en lista med ord och vi vill att du svarar på en skala hur väl programvaran stämmer in på ordet.

Ger semantisk ordskala

Avslutningsvis vill vi även att du ska fylla i en liknande skala gällande dina känslor för programvaran.

Ger Hesselgren

APPENDIX C

Semantic Differential Scale

Credible 0 _____ Max

Informative 0 _____ Max

Reliable 0 _____ Max

Necessary 0 _____ Max

Understandable 0 _____ Max

Helpful 0 _____ Max

Consistent 0 _____ Max

APPENDIX D

Hesselgrens Emotional Scale

Joy

Do not agree Agree

Trust

Do not agree Agree

Fear

Do not agree Agree

Surprise

Do not agree Agree

Sadness

Do not agree Agree

Disgust

Do not agree Agree

Anger

Do not agree

Agree



Anticipation

Do not agree

Agree

