Line Silsand, Gro-Hilde Severinsen, Mari Kannelønning (2023): Preparing for Implementing Commercial Algorithms in Radiology; A Formative Evaluation Study. 9th International Conference on Infrastructures in Healthcare. DOI: 10.48340/ihc2023_p010

Preparing for Implementing Commercial Algorithms in Radiology; A Formative Evaluation Study.

Line Silsand, Gro-Hilde Severinsen, Mari Kannelønning
Norwegian Centre for E-health Research, Norway, Norwegian Metropolian university, Norway
Line.Silsand@ehealthresearch.no , *Gro-Hilde.Severinsen@ehealthresearch.no*, *Marika@Oslomet.no*

Abstract. The aim of the paper is to contribute with an understanding of the empirical preparation necessary for adopting and scaling AI solutions in clinical radiology. We do this by asking the following research questions; How is the organization preparing for implementing CE marked algorithms? How to establish acceptance among the radiologist when implementing commercially available AI solutions for radiology? Our empirical case is the procurement and preparation of implementing AI solutions for radiology in a Norwegian health trust. We have used a formative evaluation approach and followed the procurement and pre-implementation process from it started in autumn 2020.

Introduction

Artificial intelligence (AI) is developing rapidly and for radiology, it has the potential to handle an ever-increasing volume of imaging examinations and thus be a countermeasure against the increased lack of human radiological resources. AI is expected to cause large changes in clinical work practices and support radiologists by suggesting diagnoses and increasing the likelihood of earlier detection of problems (Silsand et al., 2023; Van Leeuwen et al., 2021). In high-tech radiology practices, AI can be used to plan examinations, in example to correct position of body parts to be examined; to prioritise tasks according to severity; or to perform time-consuming and repetitive tasks like processing images (Syed and Zoga, 2018).

Despite many studies showing how AI technologies for radiology can have certain benefits, very few commercially available AI solutions have made their way into actual use (Van Leeuwen et al., 2021; Strohm et al., 2020. Existing studies have largely blamed this on technology, while organizational aspects have been overlooked (Strohm et al., 2020). Among the most prominent hindering factors is the uncertain added value for clinical practice, which causes low acceptance of AI applications among adopters and complicates the mobilization of funds to acquire AI applications.

The market for AI software in radiology is growing and there are over 180 CE-marked products commercially available. CE-marking implies that AI is classified as a medical device and therefore must be CE-marked before it is allowed on the European market (Van Leeuwen et al., 2021). This means that it meets the essential requirements of the relevant directives and that it may legally be placed on the market freely throughout European member states (Harvey, 2019). The rapid increase in the number of commercially available algorithms and the variety of ways in which each algorithm can affect clinical workflows adds complexity to the AI procurement and implementation processes (Van Leeuwen et al., 2021, Strohm et al., 2020). This has led to a rise in marketplaces for AI software in radiology, which aim to offer multiple AI software on one platform (Ibid). Nevertheless, currently, CE-markings are granted without requiring proof of performance and added benefit for clinical practice. Previous studies have identified lacking acceptance as one of the most important causes for non-adoption and thus a barrier to the successful implementation of AI applications in radiology. A facilitator for adoption is the importance of clinician understanding of and trust in AI systems (Ibid). To increase the acceptance of commercial AI applications among adopters, more evidence of the added values of the applications in the clinical setting is needed. There is a need for more real-world research, which requires a shift from measuring technical accuracy to evaluating the impact on processes and people to overcome the resistance to implementing and adopting commercially available AI solutions in clinical operations (Ibid).

Based on this introduction, we ask the following research questions: How is the organization preparing for implementing CE-marked algorithms? How to establish acceptance among the radiologist when implementing commercially available AI solutions for radiology?

Our empirical case is the procurement and preparation of implementing AI solutions for radiology in a Norwegian health trust. The project procurement lasted from 2020 -2022 and the HealthTrust is now in the pre-implementation phase. The first algorithm is set to go-live in June 2023. We have used a formative evaluation approach and followed the procurement and pre-implementation process from it started in autumn 2020. We have used the Non-adoption, Abandonment, and challenges to Scale-up, Spread, and Sustainability (NASSS) framework to outline the complexities related to different aspects of the empirical process. In addition, we draw on the information infrastructure (II) literature (Hanseth and Monteiro, 1998; Hanseth and Lyytinen, 2010), where systems in an II are never seen as standalone entities. Algorithms must also be integrated in the existing II, comprising other information systems and embedded in conventions and practices (Hanseth and Monteiro, 1998). The aim of the paper is to contribute with an understanding of the empirical preparation necessary for adopting and scaling AI solutions in clinical radiology.

Background procuring AI solutions.

In Norway, almost non commercialized AI solutions have been implemented in clinical practice. This led to a national initiative coordinated by the Norwegian Directorate of Health (NDH), addressing the need for acquiring and implementing such solutions (Directorate of e-health, 2022). In relation to the national initiative, the HealthTrust's project gradually emerged as a national pilot for procuring and implementing commercial AI solutions for radiology. The motivation for the Health Trust is related to a steady increase in labour-intensive imaging examinations, estimated at 5–10% per year. The HealthTrust was given a budget of 1.7 MNK by the regional health authority and was expected to use internal human resources. Given its role as a national pilot, NDH commissioned the Norwegian Centre for e-health research (NCE), as an external research institution, to conduct a formative evaluation research study of the procurement and implementation process to identify and explore challenges and benefits of procuring and implementation project. The formative evaluation study is funded by NCE.

The HealthTrust started their process in early 2020. First, by 'scanning' the AI marked, mapping the needs of the radiologists, and then starting the procurement process. The procurement was aligned with EU regulations for public tender acquisitions conforming to the principles of competitive dialogues. During the procurement process, the HealthTrust project changed their plan of procuring 4-5 single algorithms to procuring a platform resembling an App-store. In August 2022 the procurement process was finalized, and a contract was signed with one of four vendors.

The HealthTrust defined three broad requirements for the AI solutions. First, they wanted to procure commercially developed CE-marked solutions already used in European clinical practices to limit the need for local validation of the algorithms. Second, they were going to acquire static CE-marked AI solutions that cannot be trained on local data – a of the shelf product easy to implement. Third, the AI solution should not operate autonomously but assist the radiologists. In addition, three modalities were highlighted as important for improving the radiologists' workflow I) CT thorax for lung nodules, pulmonary embolism, and lung metastases, II) MR caput for multiple sclerosis (MS) follow-up, and III) Conventional X-ray for skeletal X-ray and chest X-ray.

Theory

We use the EU's expert group's definition of AI: "Artificially intelligent systems perform actions, physically or digitally, based on interpretation and processing structured or unstructured data, for the purpose of achieving a given goal. As part of an AI system, specialized algorithms operate on specific datasets, i.e., an algorithm takes the dataset as input and produces some output, specifically, a classification". The algorithms may also have machine learning (ML) abilities that enable them to learn from previous actions, which means that they can change behavior (Directorate of e-health, 2022).

AI and ML are associated with an uncertainty that requires engagement from users to ensure that the algorithms adapt in a useful direction and conform seamlessly to the sociotechnical configuration in radiology departments (Silsand et al., 2023). It is established evidence that differences in the context in which technology is embedded are associated with changes in performance. The impact of technology will thus change with time, as the way it "fits", an organizational network changes (Strohm et al., 2020). Conceptually, we draw on the information infrastructure (II) literature to account for the size and scope of the socio-technical system, see for instance (Hanseth and Monteiro, 1998; Hanseth and Lyytinen, 2010). This concept has been frequently used for analyzing the implementation and use of large-scale socio-technical information systems [(Ibid). In healthcare, an II consists of a range of systems, health professionals, institutions, and established practices, i.e., the installed base that evolves gradually over time, and possibly in many directions (Hanseth and Monteiro, 1998; Strohm, 2020).

IIs develop through extending and improving the installed base, meaning that the existing infrastructure influences the design of new components and systems (Hanset and Lyytinen, 2010). However, CE-marked AI solutions cannot be changed, in terms of adding new data to improve the algorithm. Therefore, implementing such AI solutions requires changes in the installed base, which not influence the design of the AI component. When changing an installed base, it is essential to design the new version for usefulness and build upon the existing infrastructure (Ibid). According to Design Theories and Principles, new systems should deliver immediate benefits to users and gain momentum by being adopted by as many users as possible while incorporating new functionality only when required and supported by a sufficiently large user-base (Ibid).

Previous research has demonstrated that implementing healthcare technology is complex, and successful implementations for one organization can be a total failure in another setting (Greenhalgh, et al., 2017). However, the technology is only responsible for 20% of AI implementation failure, the rest of the complications are directly linked to the lack of socio-technical considerations (Lebcir et al., 2021). To map the socio technical complexities of the installed base and the evolving II, we lean on the NASSS framework (Greenhalgh et al, 2017). The framework is designed to guide and evaluate success and failure in technology implementations in healthcare organisations in addition to evaluating the complexity in a project. It includes seven categories that we use to define the socio-technical concept; the condition, the technology, different aspects of values, user adaption, organizational relations, the wider context the solutions interact in and adaption and embedding over time (Ibid). The categories provide an overview of the socio technical areas influencing the implementation process. By mapping these categories in different stages of the procurement and implementation, it is possible to understand how the socio-technical interplay influences the installed base and the II as a whole.

Method

The HealthTrust is one of the largest health trusts in Norway, located in south-east of Norway with about 10 000 employees and the responsibility of providing specialist healthcare services for about half a million people. It includes four different hospitals where the imaging department is organized as one department with branches across all of them. In 2019, the department performed approximately 50,000 CT examinations, 17 500 MRI examinations, and 155 000 X-ray exams.

This research project is inspired by the formative evaluation approach, which can help navigate the associated complexities by capturing the perceptions of the stakeholders involved and feeding findings back to program management (Cresswell et al., 2020). The research team has monitored the HealthTrust's empirical process from it started and to date (the research project continues throughout 2023). This has given us a thorough understanding of the facilitating and hindering factors of the process. The research team from NCE represents a neutral and independent party without business interests in the HealthTrust's project. The **data collection** has taken place in collaboration with the HealthTrust over a two and a half-year period. It consists of 27 semi-structured interviews of stakeholders and clinicians collected at two points, 26 meetings with the project group, 15 observations in meetings with vendors (50 hours), minutes, and document studies. The interview guide is based on the NASSS-cat tool (Greenhalgh et al., 2017) and tailored to our empirical case. Each interview lasted for 30-60 minutes and were recorded and transcribed verbatim.

By using the NASSS framework, it helped us to establish an overview of the interactions between the socio-technical domains over time and to untangle the complexities. We also used NASSS to categorize data from the interviews. The further **analysis** followed the hermeneutic approach informed by II theory, in where all the collected data was included to get an overall understand of the complex process (Hanseth and Monteiro, 1998; Hanseth and Lyytinen, 2010; Klein and Myers,1999).

Results

After starting out relatively modestly – with the requirements for a few algorithms, the project gradually scaled in ambition, purpose, and outlook. The organization needed to make several strategic choices in preparation for adopting the CE-marked algorithms.

A platform solution - Increasing flexibility and complexity.

Today the IT infrastructure at the radiology department consists of the electronic health record (EHR) system, the Picture Archiving and Communication System (PACS), the Radiology Information System (RIS), in addition to the medical imaging modalities e.g., Magnetic Resonance Imaging (MRI). The IT infrastructure is complex as of today and adding new technologies demanding smooth integrations can be challenging. Procuring a marketplace solution is expected to generate flexibility for the HealthTrust since it provides access to a number of algorithms through a single point of integration (see figure I.)

If the HealthTrust does not get the desired benefits of using an algorithm, then it can easily be replaced with another algorithm without having to undergo new procurement processes. The marketplace solution with its algorithms is supposed to work in the background without any user interaction. It is a cloud-based solution that will receive images from examinations, check these for findings, and report the findings back to the work list in RIS - where the algorithm can prioritize the various findings and mark the findings in the images in PACS.

However, it is not clear if the findings in the images will be presented within PACS or presented in an extended view – a widget. Then the question is, will each algorithm need to have its own widget? (See figure I).

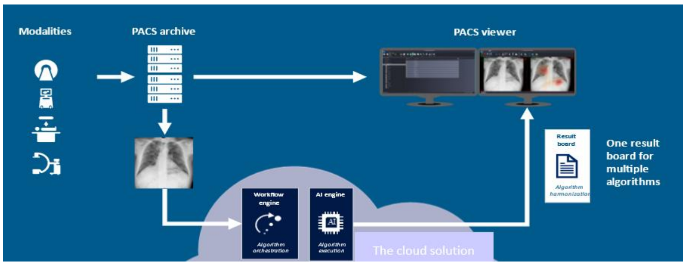

Figure 1. The integrated solution

The radiologists explained their concern:

> "We have to test the algorithm to find out how many extra clicks it will take to use it, how much time do we use or save by using it" (Radiologist II).

Even if the marketplace approach addresses flexibility related to the number of algorithms to access, it also addresses questions of complexity when it comes to integrating the findings generated from different algorithms into the clinical work processes. A radiologist stated:

> "No application can cover the whole patient, so if you really want an effect you may have to use three or four applications in addition to RIS/PACS. If they all have different interfaces, do you have to switch between them or does the marketplace solution solve that problem" (Radiologist VI)

CE-marked algorithms need local validation.

Another expected goal of procuring an AI platform offering CE-marked algorithms was that these solutions are ready for clinical use and should be easy to adopt in local clinical work processes. However, the CE-marking is generic and does not say anything about the overall quality of the product and the performance cannot be fully predicted in advance. Some of the algorithms have a well-documented clinical effect and provided benefits for the organization but most of the algorithms lack documentation about whether it works as intended in different contexts. The HealthTrust's project group analysed the documentation of the algorithms they wanted to procure. The documentation was more or less insufficient, in terms of which data was used to develop and test the algorithms, no description of the use cases or context of where the algorithms were deployed and how they were integrated into the local workflows (Van Leuween et al., 2020).

So even if they had bought CE-marked algorithms, the HealthTrust plans for an extensive local validation process (prospective and retrospective) to document how the algorithms function on the images from their patient population before using them in clinical operations. From the interviews, the radiologists emphasised that it was difficult to trust algorithms that had not been validated on data from their own patient population. The need for local validation is necessary to build trust in AI among clinicians.

> "If an algorithm is trained on one specific population, it is challenging to use it on other populations with a good result. This we need to know before implementing and trusting it". (Radiologist VI).

Moving the algorithm from one contextual infrastructure to another is challenging due to the risk of extensive differences in the patient populations, but also differences in the medical imaging equipment that provided the data when developing the algorithm compared to the equipment where the algorithm will be used. This also includes differences in how the images are taken (the protocol), which is the work of radiographers. Hence trusting the algorithms address the need for both validating the algorithm on local data and also in local work processes.

Choosing the first AI algorithm to implement

When preparing for implementation, the HealthTrust's project management decided to start with an application called "Bone View". This app was not one of the most wanted solutions pointed out by the radiologists in the beginning of the project. However, there were several arguments for starting with this application, First, the application is one of the few apps that is taken into use in large-scale in other countries (200 hospitals in 12 countries). Second, the algorithm is documented to be clinical accurate in sorting the images in positive, doubt and negative findings. The application can be used as a preliminary tool meaning that the final approval and confirmation still need to be done by the radiologists. However, the radiologists were sceptical to start with this application because bone fractures were not one of the areas they had pointed out as important for reducing their workload. In the interviews, a radiologist said:

> "The area where the first AI algorithm will assist us is not where we really have problems. Assessing x-ray fractures is a very small part of our workday and it does not take much time."

However, the clinicians are not negative to integrate new technology into their work practice but are unsure if this algorithm is the right means to solve their challenges. The scepticism is underlined by a statement from the interviews:

> "This is very new and the scepsis is high. Many radiologists don't really see how a few algorithms can be used to improve the clinical work processes." (Radiologist VII.)

Overall, the radiologists defined themselves as cautious optimists and one said; "I don't believe the effect of the algorithms will be that huge. Still, every small step ahead helps, and small pieces turn into a puzzle by the end."

Concluding Discussion

Implementing CE-marked algorithms – preparing the use case

In II theory the challenges of implementing new technology to an existing II is a dilemma that often is addressed (Hanseth and Monteiro, 1998; Hanseth and Lyytinen, 2010). For such implementation to succeed, it usually is necessary to change and adjust the technology to fit the existing installed base (workflow) in healthcare. Another alternative is that the installed base and the technology need to be adapted together. When it comes to implementing CE-marked algorithms, the preconditions are slightly different. CE-marking implicitly means that the algorithm cannot be changed. If changes to the algorithm are made, the CE-marking is no longer valid. Therefore, tailoring CE-marked AI applications to the II must be done at the organizational level, the installed base needs to change. This is contrary to the principle of a successful socio-technical implementation.

In addition, different organizations might decide to use an AI application in different ways. Some organizations may choose to run the "Bone View app" in the emergency rooms as an autonomous solution. Then the AI makes decisions without radiologists involved. If the AI-result is negative, then the patients can be sent home without further examination of a physician. In other organizations, e.g., remote radiology departments, the AI solution may be implemented as an assistant to radiologists or to physicians. While others organization may choose that the "Bone

View app" makes the first assessment of an image and then a radiologist do the second and final assessment before the finding is presented in PACS/RIS.

Accordingly, there are numerous ways to implement the same algorithm in an existing installed base, and each organization need to define its use case. Therefore, preparing the organization for implementing CE-marked algorithms is about describing the use case and how an app should function in a workflow. Formative evaluation approaches can support organizations in such change processes (Greenhalgh et al., 2017).

Indeed, there is a risk that the use cases are so different that it is difficult to identify overall values and effects for the HealthTrust. However, formative evaluations make it possible to outline the socio-technical values, and facilitating/hindering factors (Greenhalgh et al, 2017). Previous studies shows that formative evaluations make it possible to identify facilitating factors that makes the II evolve and identify values for different actors affected by the implementation. Formative evaluations can also reveal unexpected values, not defined in advance of the implementation but still of importance to the organization (Severinsen et al., 2022). In this case, the target group is the radiologists, but the implementation of the "Bone View app" is described to support a smaller part of their diagnostic work.

> "It is not the fracturs that breaks us. So, the workload comes from other heavy examinations, like cancer investigations" (Radiologist VIII).

However, when looking at the evolving II in total, there are several actors to include when defining values of a technology implementation. "I think that starting with the bone fracture app is sensible to do, but then you have changed the objective from what you wanted early in the project" (Radiologist I). In this case, the HealthTrust project's management starts using a 'simple' algorithm that is not too complicated to integrate with the installed base, is expected to generate value for the HealthTrust, however, not really solve the main issues for radiologists (Hanseth and Lyytinen, 2010). Overall, the implementation will generate important knowledge and competence for the next and more complex applications to integrate.

One way forward; start using AI algorithms.

To ensure that a technology like an algorithm fits with the installed base, there is only one way to actually do it; start using it. The discussions about promises of AI solutions has been ongoing for years. Healthcare organizations are evolving infrastructures and it is important to define and redefine the values alongside the growing II (Greenhalgh et al., 2017; Hanseth and Lyytinen, 2010). Procuring and implementing CE-marked algorithms is a learning and maturity process. In this case it several hindering factors to address and solve along the way to rich the starting point of implementation. This process has improved the HelathTrust's digital maturity. In addition, the hindering factors must be solved in the context where they belong, it is hard work demanding for collaboration among the actors and stakeholders – and is not to be solved in academia.

However, the HealthTrust took the chance of starting the procurement and implementing CE-marked algorithms for radiology to be able to make such maturity improvements. By implementing one algorithm at first, the HealthTrust learns and stays in the "drivers-seat" for making CE-marked algorithms part of their II. This provides important knowledge to enable scaling the use of AI solutions in the future. Sorting out the facilitating and hindering factors from the first implementation process, will support the next deployment of an AI app into the II (Greenhalgh et al., 2017; Hanseth and Lyytinen, 2010).

This empirical study describes the preparations necessary to start using CE-marked algorithms. The study shows the importance of validation to gain radiologists' trust in AI solutions. Moreover, this also addresses the need for sharing the results from local validation processes, as well as sharing the socio-technical considerations and changes made when implementing and adopting AI solutions. Validation and sharing of results from local validation processes are key factors to establish acceptance and trust among the radiologist when implementing commercially available AI solutions for radiology. In addition, then scaling and sustainability of the clinical use of AI becomes possible.

"When looking 10 years ahead I am pretty sure that a lot of radiology examinations will be more or less fully automated without radiologist involvement unless there are diagnostic doubts" (Radiologist VIII).

References

Cresswell K, Williams R, Sheikh A. (2020):'Developing and applying a formative evaluation framework for health information technology implementations: a qualitative investigation'. *Journal of medical Internet research*, vol. 22, no. 6, 2020, e15068.

Directorate of eHealth. (2022): 'Exploring the use of artificial intelligence in the healthcare sector' 2022, retrieved May 2th 2023 from: https://www.ehelse.no/publikasjoner/utredning-om-bruk-av-kunstig-intelligens-i-helsesektoren.

Greenhalgh, T., et al., (2017): 'Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies'. *Journal of medical Internet research*, vol. 19, no. 11, 2017, pp. e8775.

Hanseth, O., and Monteiro, E. (1998): 'Understanding Information Infrastructure'. Unpublished manuscript, accessed September 6, 2015 from http://heim.ifi.uio.no/~oleha/Publications/bok.pdf.

Hanseth, O. and Lyytinen K. (2010): 'Design theory for dynamic complexity in information infrastructures: the case of building internet', *Journal of Information technology*, vol. 25, no. 1, 2010, pp. 1-19.

Harvey, H. 2019, 'How to get clinical AI tech approved by regulators' retrieved May 4th 2023, from:https://towardsdatascience.com/how-to-get-clinical-ai-tech-approved-by-regulators-fa16dfa1983b.

Lebcir, R., etl al., (2021): 'Stakeholders' views on the organisational factors affecting application of artificial intelligence in healthcare: a scoping review protocol'. *BMJ open*, vol. 11, no. 3, 2020 e044074. 3804.

Severinsen, G. H., Silsand, L., Malm-Nicolaisen, K., Pedersen, R., Sørslett, B, & Ellingsen, G. (2022). 'Intended and Unintended Consequences of Implementing a Nursing App' eTELEMED 2022 : The Fourteenth International Conference on eHealth, Telemedicine, and Social Medicine, Porto, Portugal.

Silsand, L. et al. (2023): 'Procurement of artificial intelligence for radiology practice', *Procedia Computer Science*, 2023, pp. 1388-1395.

Strohm, L. et al., Implementation of artificial intelligence (AI) applications in radiology: hindering and facilitating factors'. *European radiology*, vol. 30, 2020, pp. 5525-5532.

Syed, A. and Zoga, Adam C. (2018): «Artificial Intelligence in Radiology: Current Technology and Future Directions», *Semin Musculoskelet Radiol* 22(5):540–545, doi: 10.1055/s-0038- 1673383.

Van Leeuwen, K. et al. (2021): 'Artificial intelligence in radiology: 100 commercially available products and their scientific evidence'. *European Radiology*, vol. 31, no. 6, 2021, pp. 3797–