THE ASSOCIATIONS OF FEDERATED MACHINE LEARNING ALGORITHM'S PERCEIVED TRUSTWORTHINESS AND ROBUSTNESS ON USER'S ADOPTION INTENT OF AI-BASED TOOLS IN RADIOLOGY MEDICAL IMAGING

by

Pavan Kumar Bhagavan Malwade, MBA

DISSERTATION

Presented to the Swiss School of Business and Management Geneva

In Partial Fulfillment

Of the Requirements

For the Degree

DOCTOR OF BUSINESS ADMINISTRATION

SWISS SCHOOL OF BUSINESS AND MANAGEMENT GENEVA

January 2023

THE ASSOCIATIONS OF FEDERATED MACHINE LEARNING ALGORITHM'S PERCEIVED TRUSTWORTHINESS AND ROBUSTNESS ON USER'S ADOPTION INTENT OF AI-BASED TOOLS IN RADIOLOGY MEDICAL IMAGING

by

Pavan Kumar Bhagavan Malwade, MBA

APPROVED BY

____
Sasa Petar, PhD, Chair

____
Ljiljana Kukec, PhD, Committee Member

Anna Provodnikova
____
Anna Provodnikova, PhD, Committee Member

RECEIVED/APPROVED BY:

____
Dino Kolak, Admission Director

Dedication

This study is dedicated to my loving wife Sarnali Mitra Malwade who have always encouraged me to pursue the highest level of education and my dearest son Darsh Malwade who was born during the mid of the study.

Acknowledgements

I would like to express my thanks to my supervisor, Dr. Anna L. Provodnikova for her support and guidance throughout my DBA studies.

Furthermore, I wish to thank Dr. Lidija Preglej, for her guidance and feedback on earlier drafts of this thesis work.

I would like to thank my fellow DBA students, who shared with me a great learning journey, full of mutual support and open debate.

Also, I would like to thank" Upgrad" for providing the platform, and its associated employees for DBA program who connected with the students, organized webinars, and interactive sessions to support in the journey.

Finally, I would like to thank my wife for her extraordinary support. Without her help and understanding, I would not have been able to meet the demands of doctoral studies in parallel to my full-time professional activities and personal commitments.

ABSTRACT

THE ASSOCIATIONS OF FEDERATED MACHINE LEARNING ALGORITHM'S PERCEIVED TRUSTWORTHINESS AND ROBUSTNESS ON USER'S ADOPTION INTENT OF AI-BASED TOOLS IN RADIOLOGY MEDICAL IMAGING

Pavan Kumar Bhagavan Malwade, **MBA**

2023

Dissertation Chair: Mario Silic, PhD, Chair
Co-Chair: Anna Provodnikova, PhD, Head of DBA Program

Purpose: To test the hypothesis whether user perception of trustworthiness and robustness of a machine learning algorithm using Federal machine learning approach of AI-based tools in radiology medical imaging has positive co-relation with user adoption intent.

Methodology: The quantitative study used perceived trustworthiness and robustness as the two independent variables and user adoption intent of AI-based tool in radiology medical imaging as the dependent variable. An online survey was planned as data collection instrument. The online survey consists of 12 research questionnaires, few of them on

Likert scale. To improve the quality and efficiency of the study, the Survey questionnaire was reviewed by experts and minor changes adopted. Pearson correlation was calculated on each independent variable versus the dependent variable and a linear regression was performed to test the correlation between both independent variables and the dependent variable.

Results: The survey link was sent to 256 recipients out of which 53 responded, giving the survey a 20.71% response rate, with 51 fully completing the surveys. Most of the respondents were Radiologists (45.10%), predominantly in the range of 10-15 years of experience (39.22%) and majority (35.29%) lived in India. Of 51 respondents, 44 (86.3%) had knowledge of federated learning and 7 (13.7%) had no knowledge of federated learning. Out of the 44 respondents who had knowledge of federated learning, 23 (52.27%) had moderate level of knowledge, 8 respondents had high level of knowledge, 1 respondent had very high and 5 (9.8%) had very low. Out of 51 respondents, 21 (41.2%) were currently a user of AI-based tools for radiology workflows and 22 (43.1%) either particpated or contributed in research or experiments related to federated learning or they intent to do. Due to the effect size (r=.549 for perceived trustworthiness and r=.303 for perceived robustness), it can also be stated that there is a moderately positive effect (medium) for trustworthiness and positively small effect for robustness individually in correlation with adoption intentions. The results from linear regression showed that model had adjusted r-squared of .302 indicating positive relationship. The average of the trustworthiness

variable had a positive unstandardized beta of .586 and the average of Robustness had a positive unstandardized beta of .020.

In addition, an analysis of the experience ranges showed a potential difference in perceived adoption intentions in respondents in range of 10-15 years. The respondents with moderate and high level of knowledge had same mean for adoption intent (4.2). The respondents who were radiologists have relatively lower level of perception when compared with other respondents.

Conclusion: The independent variables - User perception of federated learning's trustworthiness and robustness are statistically significant and influence positive correlation with dependent factor user adoption intent of AI-based tools in medical imaging in radiology. Also, it was concluded that difference in experience, levels of knowledge in federated learning and type of role of respondents had difference in answering the questionnaire and their perception. Further research could examine and provide valuable insights.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I:

INTRODUCTION

1.1 Definitions of Terms

Algorithms: Algorithm refers to a set of rules/instructions that step-by-step define how a work is to be executed upon in order to get the expected results.

Artificial Intelligence (AI): The simulation of human intelligence processes by machines, especially computer systems. These processes include learning (the acquisition of information and rules for using the information), reasoning (using rules to reach approximate or definite conclusions) and self-correction.

Bias: systematic difference in treatment of certain objects, people, or groups in comparison to others (DRAFT ISO/IEC DIS 22989, 3.4.4. p 10)

Deep Learning (DL): Deep learning is an artificial intelligence (AI) function that imitates the workings of the human brain in processing data and creating patterns for use in decision making.

Homomorphic Encryption: Homomorphic Encryptions is a technique which allows computational encryption on data enabling AI functions without the need to transfer personal information. It includes key generation, encryption, decryption and evaluation algorithms.

Robustness: ability of a system to maintain its level of performance under any circumstances (DRAFT ISO/IEC DIS 22989, 3.4.11. p 11)

Trustworthiness: ability to meet stakeholders' expectations in a verifiable way

(DRAFT ISO/IEC DIS 22989, 3.4.11. p 11)

Federated learning (FL): A learning paradigm seeking to address the problem of data governance and privacy by training algorithms collaboratively without exchanging the data itself (Reike et. al., 2020).

Machine Learning (ML): Machine learning is an application of Artificial Intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed

1.2 Introduction

The global market size for Artificial Intelligence (AI) in healthcare is estimated roughly $28 billion in 2025 (Benjamens et al. (2020). Research on artificial intelligence (AI), and particularly the advances in machine learning (ML) and deep learning (DL) have led to disruptive innovations in radiology, pathology, genomics and other fields. Presently, AI technology has been used to support patient-specific diagnosis, treatment decisions and perform population-based risk prediction analytics (Romero-Brufau et al., 2020). Choy et al. (2018) emphasized that machine learning has potential to improve different steps including clinical decision support systems, treatment planning, triaging in radiology. Healthcare Companies and start-ups are now focusing on strategizing to promote AI-based services and AI marketplace (Coombs et al., 2020; (Kumar et al., 2020).

Radiology is currently facing multiple challenges. With increasing population, and an increasing demand for imaging and diagnostic services, there is a global shortage of radiologists. Recent trends suggest that the volume of images is growing faster than the number of radiologists. The high workload may lead to errors in diagnosis due to human fatigue, unacceptable delays in reporting, and stress and burnouts in radiologists. These challenges have reduced the available time of the radiologists to evaluate a single scan and worldwide shortage of qualified radiologists to read, interpret, and report these images (Rimmer, 2017).

This pressure has been shown to influence the well-being of radiologists, pushing many significantly and negatively to burnout, and consequently reducing the department's productivity and quality of medical care given to patients (Hosny et al., 2018). The use of artificial intelligence (AI) might be one of the solutions to relieve the radiologists' workload, but what are some of the challenges for clinical implementation and adoption.

The development of artificial intelligence (AI) in healthcare has increased productivity and improving efficiency for specific tasks for Clinicians by enabling time-saving benefits. Hospitals are deploying artificial intelligence medical software suits as a complete package for the usage or taking up one program at a time which is used the most in the industry. The diagnostic imaging center's significant revenue generation is through imaging procedures, and they are primarily involved in implementing advanced products, which will attract customers (Markets, 2022). AI algorithms have proven that they can automate some of the tedious tasks in clinical practice (Barzescu, 2020). For instance, AI in medical imaging, along with clinical data, is helping physicians to predict heart attacks in patients accurately. To adopt AI solutions, buying decisions for Hospitals increasingly rely on considerations of their concrete added value to the health system.

The implementation of AI-based tools in radiology is expected to improve workflow efficiency without sacrificing accuracy, thus keeping radiology sustainable and accessible. AI-based tools can be applied to time-consuming and labor-intensive tasks, such as volumetric measurement or structural segmentation, which can accelerate the diagnostic pathway while maintaining quality continuous care. In addition, it can give radiologists more time for complex cases, speed up simple cases, and increase standardization (Hosny et al., 2018).

Such increased standardization, when it comes to tasks such as structured reporting, can have a positive impact on diagnosis and treatment, by reducing communication barriers and optimizing treatment decision-making process. In the case of prostate cancer, for example, AI-assisted MRI evaluation early in the diagnosis could help physicians later in the diagnosis process, helping urologists perform MRI-guided biopsies (Bjurlin et al., 2020), pathologists with tumor staging, and radiation oncologists with MRI treatment (Xie et al., 2022).

A recent paper calls AI regulators, cross-disciplinary organization and key stakeholder collaboration along the AI care roadmap to facilitate successful clinical deployment of AI resources (Daye et al., 2022). Such a regulator could create a road map for what kind of tools to deploy, how to assess them for their populations, how to implement and, equally important, how to monitor and maintain those developments. that declaration (Daye et al., 2022).

This can be important because while AI has many clinical benefits, hospital- and patient-level differences can lead to different priorities when it comes to implementation. For example, centers with high patient volumes and long grueling patient care routes can prioritize time savings with streamlined workflows. Other financially constrained centers may aim to improve accuracy, reduce overtreatment and redundant second-line diagnostic work, reduce overworked clinicians, and improve quality of life of patients (Beck, 2022).

Benjamens, Dhunnoo and Meskó (2020) reported in the study that the two top medical specialties leading innovation with AI-based algorithms are radiology (72.4%) and cardiology (13, 8%) based on agency regulatory license volume trends - USA. Food and Drug Administration (FDA). There are also other specialties that focus on internal medicine/endocrinology, neurology, ophthalmology, emergency medicine and oncology. This study covers approvals through February 2020. The number of regulatory approvals has increased significantly (521 entries as of January 17, 2023), which can now be tracked officially on the page (US Food & Drug Administration web site, 2023). The clear upward trend means that AI algorithms are being integrated and commercialized into clinical workflows and more customers are adopting AI-based tools.

Generally, machine learning techniques are developed by using a train-test system. Ideally, three primary sets of data for training, testing, and validation are needed. The training data set is used to fit the model. During training process, the algorithm learns from examples. The validation set is used to evaluate different model fits on a separate data and to tune the model parameters. Most training approaches tend to overfit training data, meaning that they find relationships that fit the training data set well but do not hold in general. Therefore, in order to avoid overfitting and optimize the algorithm, successive iterations of training and validation may be performed. In the testing set, after a machine learning algorithm is initially developed, the final model fit may then be applied to an independent testing data set to assess the performance, accuracy, and generalizability of the algorithm (Choy et al., 2018).

In radiology, developing machine learning models involves several challenges. High quality training data is vital for good model performance but is difficult to obtain. The available data may lack volume or diversity. It may be dispersed across multiple hospitals. Even if image data is available, it may not be labelled. Radiographic scans have a high degree of inter-reader variability when two or more radiologists label data inconsistently; this can lead to noise or uncertainty in the underlying truth label. The distribution of target classes can be very skewed, especially for rare diseases. This imbalance in class representation is often accompanied by the cost of unequal misclassification across classes. Care must be taken when dealing with unbalanced datasets, which sometimes require the use of special performance measures. A model that performs well on data from one hospital may perform poorly on data from another. Similarly, a model implemented in practice in one hospital may experience gradual deterioration in performance within the same hospital. Machine learning models have proven to be vulnerable to exploits and malicious attacks. To support adoption by radiologists, the models deployed must be able to explain their decisions and they must not discriminate against patients on the basis of gender, ethnicity, age, income, etc.

The major barriers in adaption the AI-based tools by customers were generalizability, lack of trust due to bias and safety mechanisms and complying with regulations through transparent and explainable algorithm while preserving data privacy and security (Kelly et al., 2019, Meskó and Görög, 2020). Bias in AI models may be inherited when datasets are not representative of the target population, or incomplete and inaccurate data are used by AI systems for decision-making (Vayena et al., 2018).

Data plays an important role in machine learning systems due to its impact on model performance. Although widely deployed remote devices (e.g. mobile/IoT devices) generate huge amounts of data, the thirst for data remains a challenge due to growing concerns about rights. data privacy (e.g. General Data Protection Regulation (GDPR). Protecting patient privacy in the provision of healthcare is an essential ethical principle, for the sake of privacy, patients' well-being and identity (Cath, 2018). If a patient's privacy needs are not met, the patient can suffer psychological and reputational damage (Dawson et al. work, 2019).

Another challenge that AI software faces in clinical implementation is the lack or lack of clinical evidence, i.e. studies that evaluate the software in clinical settings using actual use cases. While software manufacturers provide endorsements to regulators to demonstrate the use of AI required for their approval (e.g. FDA license or CE marking), conducting Larger clinical trials can be difficult to prove difficult to do. Joint research between clinics (and/or universities) and established companies to assess the (clinical) usefulness of real-world AI, encouraging better software and, therefore, filling fill these gaps in clinical evidence. In addition, peer-reviewed reviews that follow newly created guidelines, such as DECIDE-AI, are available (Beck, 2022).

For effective healthcare delivery, trust between the public and the health system is an important factor (Char, Shah, & Magnus, 2018). Trust in AI-based tools is considered an important factor influencing adoption decisions in clinical workflows (Hengstler, Enkel, & Duelli, 2016). Gaining confidence in the use of AI is considered a significant challenge for successful AI implementation in medical practice (Whittlestone et al., 2019). If users don't understand how AI devices work, they may have less confidence in their functioning and how they create therapeutic solutions.

As customer use of AI-based tools and services in X-ray medical imaging is expected to increase, the need to implement a robust and reliable machine learning algorithm is anticipated is an important requirement (Venugopal et al., 2022). These are essential for a successful AI solution implementation. Previous studies have shown that in clinical workflows, not all individuals are willing to accept the use of AI-based tools (Laï, Brian & Mamzer, 2020). Chew and Achananuparp (2022) took a strong position in research assessing the extent that the perception and need of AI in the use of AI in healthcare are key to improving adoption intention.

Building good machine learning models with limited data sets at individual locations is difficult because traditional machine learning centralizes training data on a single machine or in a data center. As suggested by Ng et al. (2021), an alternative is federated machine learning. The concept of Federated learning was first proposed by Google in 2016. In federated learning, client devices perform model training locally and generate a global model collaboratively. The data is stored locally and is never transferred to the central server or other clients. Instead, only model updates are communicated for formulating the global model (Qian et al., 2019).

Federated learning is a technique for training machine learning models with the knowledge that we do not have access to (Kaissis et al., 2020). The data is collected, processed into a dataset and transmitted to a central server to train the dataset in any model, and we get the predicted output. This helps us to feed the algorithm into the data instead of doing this federated learning and then passing the results to a central server. This implies that users will not be prompted to upload their personal information. Predictive maintenance is powered by Federated Learning. Based on the results from the central server, predictive maintenance predicts when the system will need maintenance.

In the healthcare domain, federated learning use cases for devices would allow the user to learn a model of machine learning that will help patients improve certain aspects of their health without having to upload their data to a central cloud. Federated learning entails using a wide corpus of high-quality decentralized data distributed through several client devices for instruction. Since the model is trained on client computers, no data from the user is expected to be submitted. Keeping the client's personal data on their computer gives them clear and physical control of their information (Shah et al., 2021).

The recent study across healthcare systems supports the hypothesis that federated learning trained models are generalizable and robust (Dayan et al., 2021). Federated learning promises to address above concerns with respect to trust, bias, privacy, generalizability by offering easy scalability, flexible training scheduling, and large training datasets through multi-site collaborations. However, there are still challenges remain and must be addressed before federated learning is optimally able to build trustworthy AI models. Additionally, as federated learning in medical imaging AI is novel topic, this has the potential to attract investment by vendors and inspire researchers, whose work will be necessary to advance the field forward. Based on my verification of the database and there were no specific federated learning based algorithms have obtained regulatory clearance till date (US Food & Drug Administration web site, 2023). It can be safely assumed from my opinion that there are no federated learning based algorithms that is being currently used in any of the institutions in their clinical workflows.

1.3 Research Problem

Users of traditional machine learning in medical imaging are concerned about processing the data over cloud and do not trust the algorithm due to concerns over bias, generalizability and data privacy issues resulting in slower adoption rate of AI-based tools in radiology medical imaging (Turja et al., 2020; Masud et al., 2019; Sun and Medaglia, 2019).

Several experimental studies conducted in radiology medical imaging has positive results showing federated learning performs better than traditional machine learning. The outcome of the preliminary literature study indicates theoretically that federated learning could have positive correlation in customers perception of trustworthiness and robustness of machine learning algorithms, however there is still gap to understand whether there is positive correlation of customer's perception in federated machine learning algorithm which may impact customers adoption intent of AI-based tools (Stripelis et al., 2021; Yang et al., 2021; Linardos et al., 2021; Sarma et al., 2021; Dayan et al., 2021; Sheller et al., 2020).

Currently, there are no user surveys available to evaluate the association of user perception about federated learning approach with their adoption intent which is important factor for vendors to consider investing in research and development in this emerging technology.

1.4 Purpose of Research

In this machine learning, as the paradigm is shifting towards using federated machine learning over traditional machine learning, based on outcome of literature research and in my opinion, there is a gap in understanding of perceived trustworthiness and robustness of federated machine learning algorithm and correlation with user adoption intent. Masud et al., (2019) has concluded that in general, the perceptions of radiologists have not been considered and details of implementation approaches for adoption of machine learning tools have not been reported.

This research is aimed to test hypothesis to understand the association of user's perception on trust and robustness in federated machine learning tools over traditional machine learning that could influence the user adoption intent of AI-based tools in radiology medical imaging.

1.5 Significance of the Study

To advance further with use of AI-based tools in medical imaging, machine learning vendors and users must adapt federated learning approach. To assess the business value of investing and deploying federated learning approach, it is critical to understand the influence of perceived trustworthiness and robustness of federated learning algorithm with user adoption intent. Ultimately, the study aimed to contribute to the development of a value proposition for AI-based tools by evaluating stakeholders' perceptions of the adoption of AI in radiology.

1.6 Research Purpose and Questions

The intent of this research is to test hypotheses whether there is positive correlation of trustworthiness and robustness of federated machine learning models which influences the user adoption intent of AI-based tools in radiology medical imaging.

For this study, the research question is:

What are the associations of federated machine learning approach over traditional machine learning towards perceived trustworthiness and robustness and its prediction with user's adoption intent of AI-based tools in radiology medical imaging?

Below is the hypothesis:

H1: User perception of federated learning's trustworthiness has a positive correlation with user adoption intent of AI-based tools in medical imaging in radiology

H2: User perception of federated learning’s robustness has a positive correlation with user adoption intent of AI-based tools in medical imaging in radiology

H3: User perception of federated learning’s trustworthiness and robustness have a positive correlation with user adoption intent of AI-based tools in medical imaging in radiology.

CHAPTER II:

REVIEW OF LITERATURE

The term “Artifical intelligence” dates lower back to as a minimum the mid-twentieth century. Artificial Intelligence has improved signifcantly and begins since the term was first coined in 1956 by American scientist John McCarthy. In the Eighties and 1990s, AI become focussed on rule-based ‘expert systems’ that carried out predefined logical rules. Later, considerable development was withnessed with hype, expectancies and disappointment. Data-driven AI has emerged as the principle enabler of technology in 2000’s. AI is now advanced at level wherein it could have interaction and speak with people via analytics and automation. AI can carry out human-like cognitive functions (e.g. recognising styles from statistics and expertise images, textual content or speech), in addition to make predictions, tips and decisions. AI can permit computer systems to imitate human intelligence so we can learn, sense, assume and act, with a purpose to gain automation and benefit analytic insights.

Machines have the ability to "feel" and "act" in relation to people and the environment through user interfaces, sensors and robotics, the AI is equipped with a "thinking" component - e.g. Support making predictions, recommendations or decisions. AI applications use two computational approaches: rule-based and rule-based to mimic the human mind.

Rule-based AI systems became popular in the 1960s and dominated from the 1970s to the 1990s, especially for industrial robots performing repetitive tasks. However, it has limited applications, adding new rules or knowledge to these systems is very time-consuming and expensive so that they can respond to the changing business environment. Also, it was very difficult to clearly define the rules programmatically or declaratively. This led to the development of illegal AI. AI without reasoning can make decisions according to a set of pre-defined rules and can also gain meaningful insights by automatically "learning" from its input.

Machine Learning (ML) enables training capabilities in AI. Machine learning is a data science technique that allows computers to learn without being programmed with explicit rules. Machine learning makes it possible to develop algorithms that can learn and make predictions. Unlike rule-based algorithms, machine learning uses access to large and fresh data sets and has the ability to improve and learn from experience.

Instead of being explicitly programmed to perform certain tasks, machine learning involves using a set of learning algorithms driven by mathematical techniques which allow machines to learn from data. The training process uses the learning algorithm to derive relationships between data points from training data, which is commonly a subset of historical data. The outputs of the training are trained machine learning models, which can perform predictions or make decisions according to the data patterns observed from the input data, or from queries provided by users.

Machine learning (ML) is a tool used in the context of artificial intelligence (AI). ML includes pattern recognition, artificial neural networks, computational statistics, data mining, image processing, and adaptive algorithms. ML algorithms are designed and developed to recognize specific patterns in data and gradually improve their predictive ability.

In other words, ML can be seen as a natural extension of traditional statistical methods. ML is a model that learns from case examples rather than rules. The examples act as the ML input (classification) and the calculated result as the output (label). For example, a biopsy sample read by a pathologist can be scanned and converted into layers, i.e. a set of pixels that make up the test, and into labels, i.e. disease classification information (Kashani et al., 2020).

ML is capable of identifying subtle data patterns that cannot be easily described by humans and extracting insights from less structured data. As a result, ML has become the main technique driving the current wave of AI applications – from call center voice analysis to autonomous vehicles.

However, ML is generally heavy on data and computation. To find data patterns or relationships and make accurate predictions, ML has to process huge volumes of data. They must perform computationally intensive statistical techniques and mathematical algorithms to train the model as well as to tune and test model selection. Even for rule-based AI, experts often perform data analysis and statistical hypothesis on historical data sample sets using statistical techniques. Also known as "data mining," this method finds statistically significant data patterns when it comes to predefined rules to program into AI's knowledge base.

In fact, machine learning and data mining involve many statistical models and algorithms. The former differs from the latter in that data mining is performed by a knowledge worker on a particular data set with a specific goal i.e. to discover patterns and derive insights. color from the data set. In contrast, machine learning algorithms are performed automatically or semi-automatically by computers. The quality and quantity of training data directly affects the prediction accuracy of most AI models. Proper data preprocessing, including data collection, data profiling, cleaning, transformation, and labeling of training data, is critical to successful AI development.

Raw training data can come in many formats: including structured data (such as historical stock trading records stored in a database), semi-structured data (such as like social media feeds) and unstructured data - like audio, video, and image. ML algorithms require this training data to be cleaned and converted into a deterministic format that can be easily read by the machine. For example, speech recognition technologies do this by converting voice recordings into machine-readable text that can be used to train Chatbots. To derive insights about relationships between data, data scientists may also need to assign meaning to training data based on domain knowledge, i.e. data labeling. For example, if the task is to train the AI to classify scanned image documents, then data scientists will need to label the document type (i.e. output or labeled data) for each scanned image (ie input data).

Training ML models requires millions of data. The preprocessing and labeling of this data is labor intensive and takes up most of the work in the ML model training process. As a result, the development of machine learning in previous waves of AI was slow. But with recent advances in big data technology, large amounts of data can now be collected and processed.

Machine learning tasks are typically classified into three broad categories, depending on the type of task: supervised, unsupervised, and reinforcement learning (Figure 1).

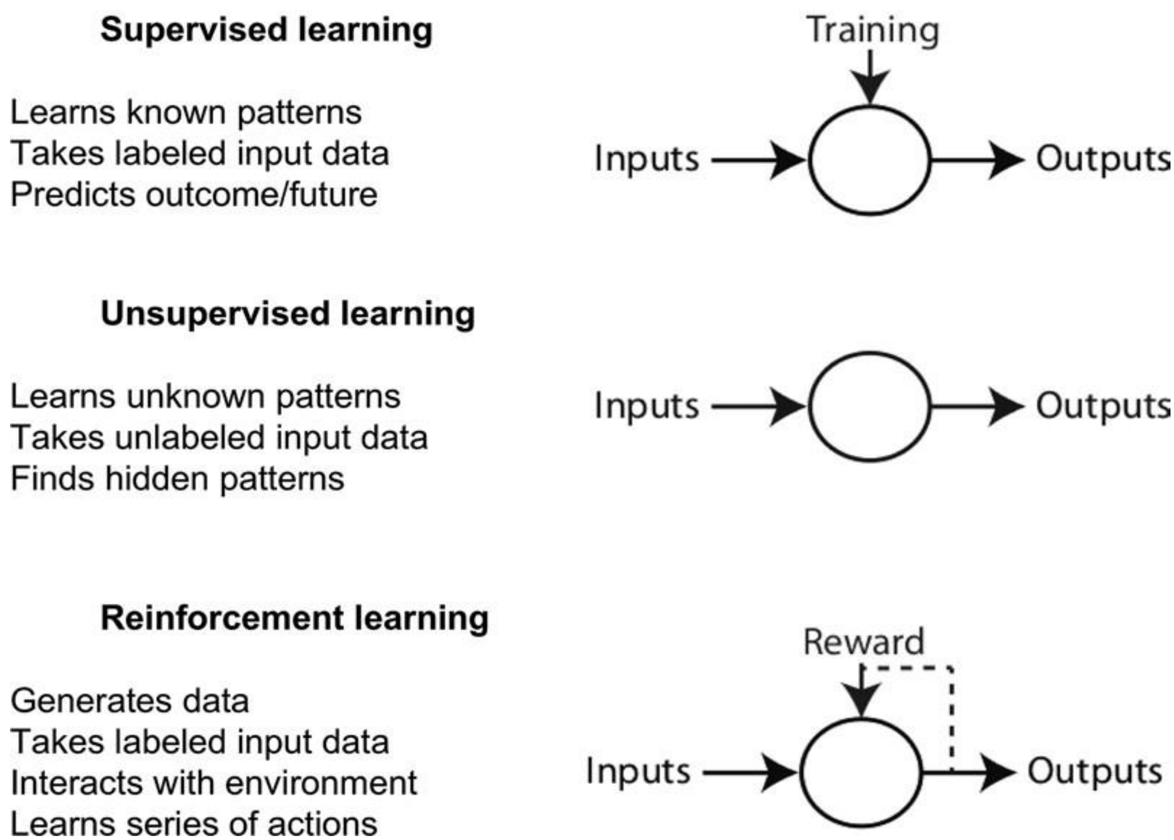

Figure 1 Image shows different categories of machine learning (Choy et al., 2018)

In supervised learning, labeling information is provided to the algorithm during the training phase with supervision in training. Expected results are usually noted by human experts and serve as ground truth for the algorithm. The goal of an algorithm is usually to learn a general rule that maps inputs to outputs. In machine learning, data that is true is called "ground truth". In unsupervised learning, no data labels are assigned to the learning rate. The goal of the machine learning task is to find the hidden structure in the data and separate the data into clusters or groups. In reinforcement learning, an algorithm performs a specific task in a dynamic environment where it receives feedback in the form of positive and negative reinforcement (for example, playing a game against an opponent). Active learning is learning the consequences of interacting with the environment without explicit instruction. Examples of supervised and unsupervised learning methods are shown in Figure 2. A machine learning paradigm can use a combination of supervised and unsupervised techniques with a robust feedback loop.

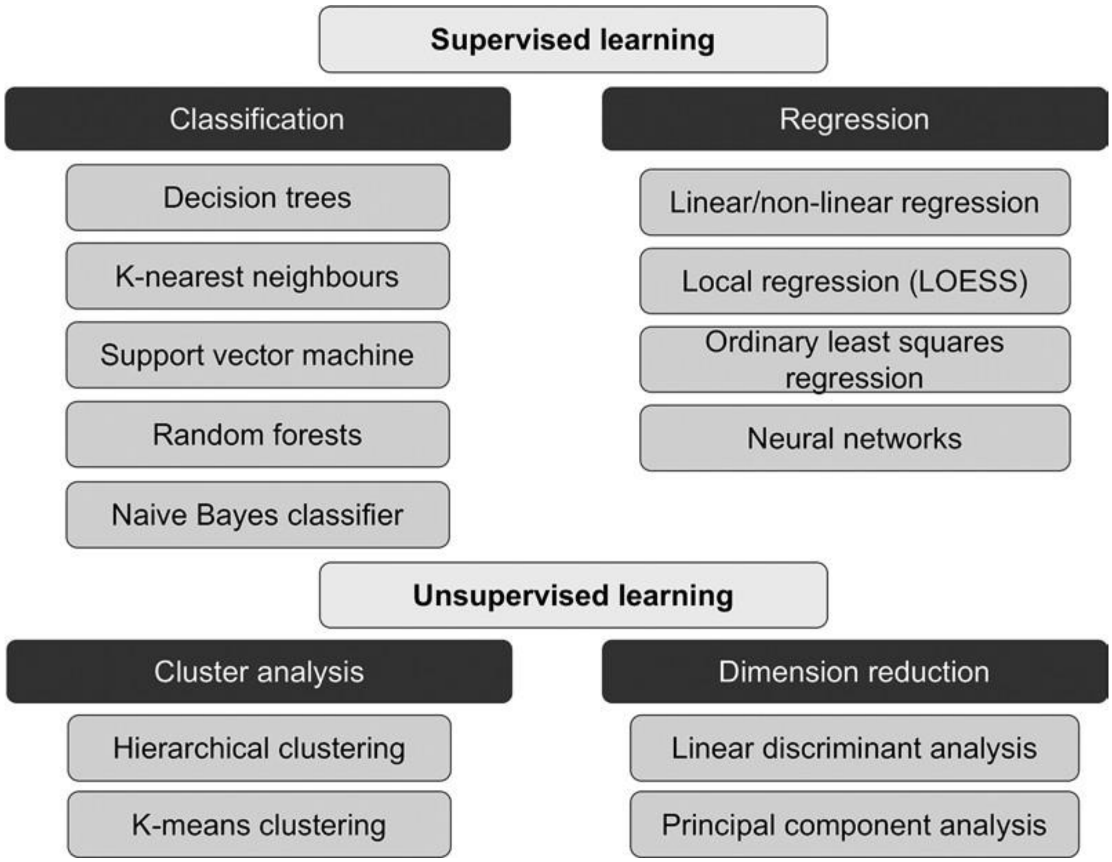

Figure 2 Supervised and unsupervised learning paradigms (Choy et al., 2018)

Artificial intelligence, in contrast to machine learning, encompasses a broader range of intelligent tasks performed by computers, such as problem solving, planning, knowledge representation, language processing, or "learning." Therefore, machine learning is a form of artificial intelligence. For example, rule-based algorithms such as computer-aided diagnosis, which have been used in mammography for several years, represent a type of artificial intelligence, but not a type of machine learning. However, computer-aided diagnosis is a broader term and can include machine learning approaches. By definition, machine learning algorithms improve automatically with experience and are not rule-based. Machine learning is gaining popularity in a variety of use cases, and in fact many artificial intelligence applications are currently using machine learning approaches.

A suite of new machine learning algorithms has increased computing power, and the explosion in the availability of very large datasets (“big data”) (Samek, Wiegand, and Müller, 2018) due to the digitization of health information has led to remarkable recent advances with evidence of machines achieving human-level competence in many fields (e.g., breast cancer prevention) for solving well-defined tasks (McKinney et al., 2020).

Medical Imaging

There have been several publications on the use of AI in medical imaging over the past few years. Several areas have received considerable interest - radiology, pathology, ophthalmology and dermatology - due to the intuitive nature of diagnostic tasks in these specialties and the increasing availability of highly structured images. A number of companies have been approved by the US FDA and CE Europe for AI in medical imaging, and the commercial market has begun to take shape as sustainable business models are created. For example, high-throughput healthcare regions, such as India and Thailand, have welcomed the deployment of technologies such as diabetic retinopathy screening systems. This rapid growth has now reached the point of having a direct impact on patient outcomes – US CMS recently approved reimbursement for radiographic stroke management, which reduces time required for the patient to be treated. In radiology, the field of AI has seen tremendous growth in research, spanning all modalities centered on X-rays, CT scans, and magnetic resonance imaging. An example of this is X-ray analysis - a major clinical area. Brain imaging analysis (especially for time-sensitive cases such as stroke) and abdominal imaging are also of great interest. Disease classification, node detection, and partitioning (e.g., ventricular) models have been developed for most diseases for which data can be collected. This has enabled the industry to respond quickly in times of crisis by applying AI-based algorithms for classification, such as the development and deployment of diagnostic models for COVID-19. The field continues with work in image translation (e.g., converting noisy ultrasound images to MRI), reconstruction, and image enhancement (e.g., converting low-frequency CT images and imaging). low-resolution to high-resolution imaging), automating reporting and time tracking (e.g., recording images to monitor tumor growth over time). In the following sections, we explore applications in different fields.

Cardiology

Cardiac imaging is increasingly used in various diagnostic and clinical procedures. The main clinical applications of deep learning include diagnosis and screening. The most common imaging method in cardiovascular medicine is echocardiography or echocardiography. As an inexpensive, radiation-free modality, echocardiography is suitable for LD due to ease of data collection and interpretation - it is commonly used in most acute care settings, outpatient centers and emergency room. In addition, 3D imaging techniques such as CT and MRI are used to understand cardiac anatomy and better characterize the mismatch between supply and demand. CT segmentation algorithms have even been approved by the FDA for coronary angiography.

Pathology

Pathologists play an important role in cancer detection and treatment. Pathological analysis - based on visual examination of tissue samples under a microscope - is inherently subjective. Differences in visual perception and clinical training can lead to differences in diagnostic opinion and prognosis. Here, DL can support critical medical tasks including diagnosis, predicting treatment outcome and success, pathology segmentation, disease monitoring, and more.

In recent years, whole gigapixel (WSI) submicron resolution tissue scanners have been introduced. This development, coupled with advances in AI-based digital histopathology, has led to research and commercial activities. This field has the potential to (i) overcome the limitations of human visual perception and thinking by increasing the efficiency and accuracy of routine tasks, (ii) develop signs of and new methods of treating diseases from morphological structures invisible to the human eye, and (iii) combining pathology with radiometric measuremens, genomics and proteomics to improve diagnosis and prognosis.

A research topic focused on automating the frequent and time-consuming task of locating and quantifying morphological features. Examples include the detection and classification of cells, nuclei, and mitosis, as well as the location and segmentation of histological primitives such as nucleus, gland, duct, and tumor. These methods often require manual annotation of tissue components by pathologists as training data.

Another line of research focuses on direct diagnosis and prognosis, from WSI or tissue microscopic (TMA) for many types of cancer – breast, prostate, lung, etc. The use of digital archives of pathology images and easily accessible annotations from electronic health records has the potential to transform the fields of pathology and oncology.

Dermatology

The main clinical tasks of DL in dermatology include differential diagnosis of specific lesions, finding lesions among many benign lesions, and helping to monitor lesion growth over time. A series of studies demonstrated that CNN could match the performance of board-certified dermatologists in classifying malignant skin lesions from benign lesions. These studies have sequentially examined an increasing number of dermatologists, consistently demonstrating classification sensitivity and specificity that matches or even exceeds physician levels. These studies were mainly limited to the task of binary classification to distinguish benign skin lesions from malignant skin lesions, classifying melanoma from nevi or carcinoma from seborrheic keratosis.

Recently, this line of work has been expanded to include differential diagnoses for dozens of skin conditions, including non-cancerous lesions such as rashes and genetic conditions, and incorporates non-visual metadata (e.g. patient demographics) as categorical input. This work is catalyzed by an accessible archive of images and AI challenges that encourage teams to measure themselves against predefined standards. Integrating these algorithms into clinical workflows allows their services to support other critical tasks, including large-scale melanoma detection in patients with multiple lesions and Monitor lesions on multiple images to capture transient features such as growth and change. color This area remains largely unexplored, with early CNN studies working together to detect and monitor lesions.

Ophthalmology

The field of ophthalmology in recent years has seen a significant increase in AI efforts, with dozens of papers demonstrating clinical diagnostic and analytical capabilities far beyond current human capabilities. The potential clinical impact is significant, the portability of the machines used to examine the eyes means that temporary clinics and telemedicine can be used to distribute testing sites to underserved areas. The field relies heavily on fundus imaging and optical coherence tomography (OCT) for patient diagnosis and management. CNN can accurately diagnose a number of conditions. Diabetic retinopathy - a condition in which blood vessels in a diabetic's eyes "leak" and can lead to blindness - has been studied extensively. CNN consistently demonstrates physician-grade classification from fundus photographs, leading to a system recently approved by the US FDA. Similarly, they can diagnose or predict the progression of diabetic central macular edema, age-related macular degeneration, glaucoma, excessive visual field loss, blindness in children and other diseases.

The eyes contain a number of human-incomprehensible features that represent important medical information that CNN can gather. Notably, CNN has been shown to categorize several cardiovascular and diabetes risk factors from fundus photographs, including age, sex, smoking status, hemoglobin A1c, index body mass, systolic and diastolic blood pressure. CNN can also detect signs of anemia and chronic kidney disease from fundus pictures. This presents an exciting opportunity for future AI studies that predict non-visual information from visual images. This could lead to a paradigm shift in healthcare, where an eye exam allows you to detect the presence of ocular and extraocular diseases, currently limited to human doctors.

Surgical applications

In procedural areas such as surgery and endoscopy, AI algorithms can provide significant utility. Some of the key clinical applications include improving surgeon performance through real-time contextual awareness, skill assessment, and training. Early studies have begun to pursue these goals, mainly in laparoscopic surgery and video-based robotics - some of which propose methods for detecting surgical instruments and gestures. Another use case is to recognize distinct surgical stages during surgery, in order to develop contextual computer-aided systems.

Integration of AI in radiology workflow

Although most of the literature is focused on the role of machine learning in detection of radiology findings, machine learning also has the potential to improve different steps of radiology workflow.

Order scheduling and patient screening
Automated clinical decision support and examination protocoling
Image acquisition
Automated detection of findings and features
Automated interpretation of findings
Image management, display and archiving (eg, picture archiving and communication systems)
Postprocessing: image segmentation, registration, and quantification
Image quality analytics
Automated dose estimation
Radiology reporting and analytics
Automated correlation and integration of medical imaging data with other data sources

Figure 3 Clinical Applications of Machine Learning in Radiology (Choy et al., 2018)

The current cycle is largely driven by the impressive progress of deep learning, a branch of machine learning that effectively uses artificial neural networks to solve previously difficult problems. Deep learning applications have achieved human or superhuman performance in many areas such as image recognition and natural language processing (Esteva et al., 2019). An important feature of deep learning is that the neural network parameters are tuned in a complex multi-level iterative automatic training process. In many cases, no expert level knowledge is used during training, except for the direct input and output parameters (e.g. set of pixels and their associated labels), which leads to this learning being called "end-to-end" learning. (Esteva et al., 2019). In other words, networks learn to jump directly from one end - input - to the other - output - without requiring domain-specific expertise in between. The resulting network structures are often so large, often with billions of parameters, and so complex that their behavior cannot be described in simple terms, which has led to new explanatory and their interpretability.

The recent digitization of all kinds of health data and the fact that computers are increasingly capable of interpreting some non-medical images and texts almost as accurately as humans (He et al., 2015; Wu et al., 2016) enables a multitude of applications of AI in healthcare. Much of the recent work on AI for health has focused on applications that revolve around image interpretation and natural language understanding.

In the field of medical image interpretation, one of the most widely published studies is by Esteva et al. (2017). The authors demonstrated accurate classification of skin lesions using a deep neural network trained in clinical images and evaluated performance by comparing the classifications with those performed Presented by board-certified dermatologists. This revealed that the network had reached the level of human accuracy; however, that validation in a broader geographic area is required, where the shape and color of the lesion may vary based on the variation in skin color of each area. Lijens et al. (2017) reviewed more than 300 articles using deep learning in medical image interpretation. These papers focus on detection, segmentation, or classification tasks. They include X-ray, CT, MRI, digital, cardiac, abdominal, musculoskeletal, fetal, dermatologic, and retinal pathology imaging analysis. For natural language understanding, the fields of biomedical text mining, electronic health record analysis, sentiment analysis on internet-derived data, and medical decision support systems have shown results. positive results (Ching et al., 2018). Furthermore, AI methods can automatically interpret laboratory results (from standard blood tests to recent advances in genomics and high-throughput proteins; e.g. Gunčar et al., 2018) and time series e.g., electrocardiogram, temperature, oxygen saturation, and blood pressure (Attia et al., 2019). AI can also be used beyond the specialist hospital level. For example, it can be used from primary care centers to different levels of hospital specialization, including national health institutes or national reference laboratories. The role of AI varies depending on the requirements and feasibility of the context.

Figure 4 illustrates the implementation of AI in radiology. Continuous learning functions that work normally can be disrupted by sudden workflow changes or big data errors resulting in continuous learning AI models exhibiting completely unintended behavior after training.

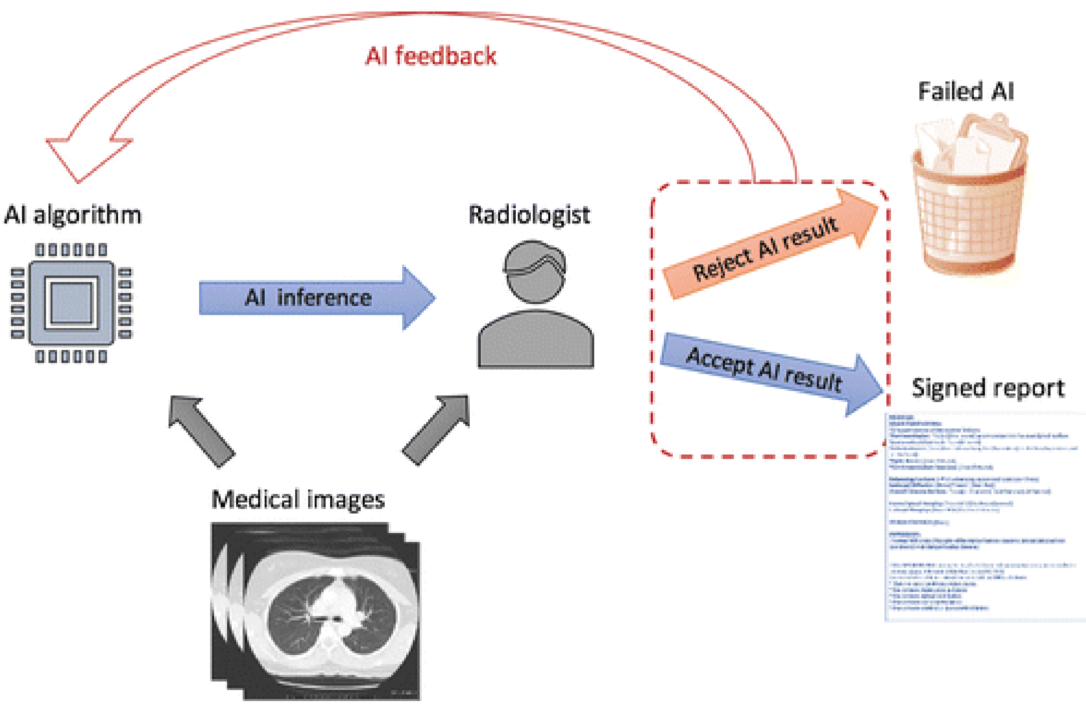

Figure 4 Illustration of implementation of an artificial intelligence in radiology. (Pianykh et al., 2020)

The applicability of these technologies has global potential. A large portion of the world's population has access to devices that can use AI-powered supercomputing applications i.e., smartphones, and other devices that store models locally or are connected via the Internet to powerful computing clusters (Albertini, 2019). Given the speed at which AI-based algorithms can be developed, improved, and deployed, this technology has the potential to ensure that first-class medical decision-making is accessible and affordable across the globe (Bell et al., 2018). This can help reach people faster and easier, conditions can be diagnosed at an earlier stage, which can lead to better health outcomes and lower costs. However, this will also require internet connectivity infrastructure and facilities, especially in remote and offsite environments.

While these advances are promising, AI for health also faces several challenges. As noted earlier, deep learning models are notoriously difficult to read and interpret, which can significantly hinder their acceptance when faced with critical, even life-saving decisions. Therefore, interpretability, explanability, and proven robustness (e.g. resistance to outliers and adversarial attacks) are important aspects that must be considered for reliability. trust. Trust. Although accuracy is reported for many healthcare AI models, there is currently a lack of data on efficacy (especially comparative efficacy), cost-effectiveness, or safety in clinical settings.

The data underlying the AI model must also include all relevant regional, gender, and age variations to be robust enough for the model to perform well in public health settings without errors. In addition, access to health data may be impeded by strategic issues of the data owner or custodian, and because the data is sensitive and subject to privacy laws depending on the country and region as well as ethical considerations related to their collection and use. Thus, access to adequate experimental data is a major limiting factor for the predictive performance of models on unpublished data, especially due to local regulatory limitations. direction. to access health data.

This problem is further complicated because most modern AI applications are based on labeled data and supervised learning. In the medical field, labels can usually only be given by trained professionals. For example, this contrasts with simple object recognition, where photos can be labeled by countless ordinary people. In addition, machine learning methods need to account for biases that data (e.g., text and image-based medical data) may contain (Caliskan, Bryson, & Narayanan, 2017). In machine learning, models and training data must be considered in combination. Models cannot be extrapolated. Instead, they can only learn the patterns contained in the training data. This data must be of high quality, of sufficient quantity to understand the multitude of parameters of a “data-intensive” algorithm, and should theoretically cover all possible scenarios, including exceptions (Hägele et al., 2020).

A thorough understanding of user attitudes and perceptions is necessary for the successful implementation of the AI-based system (Romero-Brufau et al., 2020). Due to concerns about the reliability and robustness of AI-based tools, healthcare professionals still express fundamental concerns about implementation leading to adoption challenges (Kaissis et al. Events, 2020).

The results of a survey conducted by Esmaeilzadeh (2020) on consumers' attitudes towards the use of AI-based tools in healthcare indicate that three concerns (technology, ethics, ethical, regulatory, and ethical) that directly affect the risks commonly associated with AI tools.

2.1 Technological Concerns

Availability of datasets for training and validating algorithms is limited due to lack of standardized electronic medical records and strict legal and ethical requirements to protect privacy patient privacy, which leads to a slow adoption rate of AI-based tools (Kaissis et al., 2020). The available data is divided into two parts. One part is used for training and validating the model. The second part, called the test set or the retainer, is used to estimate the final performance of the trained model after it is deployed. The basic premise is that the data used to train the model is representative of the data that the model will encounter in clinical use. This assumption is often violated in practice, making performance on storage an unreliable indicator of future clinical implementation performance.

To develop robust machine learning models, researchers need access to large health data sets that fully represent the diversity of data on population characteristics such as age, sex, ethnicity, ethnic, ethnic origin, health status, etc. and imaging characteristics such as device manufacturer, imaging parameters, patient position. and so on Most of the datasets available in medical imaging do not meet these requirements. Government policies such as the General Data Protection Regulation (GDPR), the Health Insurance Information Portability and Accountability Act (HIPAA), the Personal Data Protection Act of Singapore, etc

Multiple medical data sets are distributed natively across multiple networked storage devices owned by different organizations. In a traditional machine learning environment, these datasets must be merged into a single repository before training the models. Moving large volumes of data over the network poses a number of logistical and legal challenges, addressing potential biases.

Generalization refers to the ability of a predictive model to perform well on unseen data. To test the generalizability of a model, the data is usually divided into a training set, used to tune the model's parameters, and a test set used to evaluate whether the model generalization or not and whether or not to be considered representative of the reference population. Generalizability can then be assessed by assessing whether the test is truly representative of the reference population. The data used to train the model must be representative of the data that the model will encounter in clinical use. This assumption is often violated in practice and makes storage performance an unreliable indicator of future clinical deployment performance. The poor generalizability of models to diverse patient populations is one of the biggest barriers to the adoption of artificial intelligence and machine learning in healthcare. One of the reasons for poor generalization is the difference in image characteristics between images from training sites and deployment sites. This variation, also known as dataset change, can occur due to differences in hospital procedures, device manufacturers, image acquisition settings, disease manifestations, patient populations, etc from another hospital. The generalizability of a model, i.e. the ability to accurately predict the occurrence of events when exposed to a new set of data, depends on the balance between the bias and variance of the model. Machine learning bias, if unchecked, increases training and generalization errors by oversimplifying model assumptions. On the other hand, variance occurs when small variations in the training set lead to a significant increase in the generalization error. Several important problems arise when trying to balance bias and variance.

Machine learning is a real-world field that requires simulations to verify generalization. It is increasingly possible to understand how process automation can facilitate the daily lives of healthcare professionals, from the simplest to the most complex (Obermeyer & Lee, 2017; Rajkomar et al., 2019). The goal of most machine learning research is to one day apply models in a real-world context, which means that algorithms must be able to generalize to new data sets.

Types of generalization:

a. Local generalization: when the goal is to predict the outcome using data from the same location as the training set. Also known as internal authentication or reproducibility (Steyerberg, 2019).

Limitations: New data may not follow the same pattern over time as the data used in model training: new interventions may be introduced after the training set is collected and the New diseases may appear.

b. Extrapolation: Challenges increase when the goal is extrapolation, i.e. applying the model to a different domain than the domain used to train the algorithm. It is also known as external authentication or portability (time, geography, method, and spectrum) (Steyerberg, 2019).

Limitations: Algorithms should be applied to reasonably related populations, that is, populations in which there are similar relationships between predictors and outcomes. However, this is difficult to test empirically and may increase prediction error. Another concern is that machine learning models can break easily when put into actual use, which affects performance. As an example, researchers have shown that the performance of a deep learning model used to diagnose pneumonia on chest X-ray films is significantly lower when used to evaluate chest X-rays- optics from different hospitals (Zech et al., 2018).

The performance of the deployed machine learning model degrades over time. This phenomenon is known as model degradation. This happens due to changes in the underlying data. Detecting model degradation requires continuous monitoring of deployment-time performance against human-labeled subsamples of data. If performance drops below a predefined threshold, an alarm is triggered and the model is retrained or adjusted on the most recent data. This recycling can also be performed periodically as part of routine maintenance.

Measuring robustness can be done from different perspectives or a combination of them:

1. In the input space, use both training and test data samples. Distribution distance/dissimilarity or mismatch can be measured as illustrated in. The measurement of training data and bias can suggest how difficult a scenario is for a model in terms of maintaining accuracy or explaining its diminishing accuracy. Furthermore, the representativeness of the data (laziness or heterogeneity) can be measured in both the training and test datasets. One particular powerful scenario is noise in data labels. A model can be considered robust if its accuracy does not change much when trained with a high level of label noise.

2. In the feature space: Modern AI solutions make extensive use of deep learning architectures. These architectures typically learn a feature space from the data, making it more feasible to use to measure the similarity of training and test data.

3. Model Output: Shuffle in the model output leads to a significant mismatch between the training and test data. Model sensitivity to increasing training to test divergent data/features is a common approach to determine model robustness.

Therefore, we can define certainty as the ratio between the noise of the model output and the dissimilarity of the training test data. Endurance assessment is of prime importance in medical applications, as input test data disruption is a frequent challenge in real-world medical AI systems.

2.2 Ethical Concerns

Health data is one of the most sensitive (Vayena, Blasimme and Cohen, 2018). Patient confidentiality is of paramount importance, as failure can affect patient psychology and damage their reputation (Dawson et al., 2019). Trust can only be created by explainable AI with a deep understanding of the algorithmic decision-making process (Scherer et al., 2020).

Concerns may arise due to unrepresentative data and AI bias due to social discrimination (such as poor access to health care) and patterns small groups (such as minorities) (Reddy, Fox and Purohit 2021). Essentially, the training data should contain a meaningful representation of diverse populations. One study indicates that incomplete and non-representative data sets in AI models can lead to inaccurate predictions and medical errors (Reddy et al., 2019). Algorithmic systems play an important role in guiding decisions that affect the delivery of healthcare to patients. It is therefore desirable that these systems be free from social bias and that their decisions be fair and equal. Unfortunately, many existing datasets reflect the biases of the companies they represent, and it is difficult to detect and remove biases inherent in the training data.

The topic of Trusted AI has been debated many times in the scientific community, from journals to conferences (Pupic et al., 2022; Lockey et al., 2021; Santomartino and Yi, 2022). Distrust in AI is thought to stem from uncertainty about the value it can actually bring to clinical practice (Rylands-Monk, 2022). This uncertainty leads to a long list of common myths, perceived lack of transparency, and lack of formal education about AI. In principle, a predictive model is considered fair if it does not discriminate against patients on the basis of sensitive variables such as gender, ethnicity, disability, income, etc. However, putting this seemingly simple principle into practice is difficult.

Learning hidden features (also known as confounding factors) can lead to algorithmic bias that can produce unreliable predictions when the algorithm is applied on an external test set from a population with different hidden feature distributions. In addition, algorithmic bias can occur when an algorithm has been trained with data that only represents a subset of the real-world data that it is expected to. This has the potential to lead to predictive outcomes that are detrimental to everyone and are undesirable by the model creators (Chen et al. 2019). For example, an AI system may perpetuate a racial bias because the bias already exists in historical data. This may reflect differences in biological vulnerability to disease as well as differences in social resources. Identifying algorithmic biases is no small task, requiring specialist domain knowledge in the targeted use case scenario as well as expert knowledge of identification methods. and minimize algorithmic errors. Not only in healthcare but also in other AI application areas, the identification of learned latent characteristics, especially sensitive social factors, is of particular importance, to ensure fairness, avoid discrimination and unreliable predictions (Holstein et al. 2019). The use of machine learning algorithms for clinical decision-making should focus on demonstrating a clinically important improvement in patient outcomes rather than relying solely on performance measurements such as surface area. area under the curve and accuracy. It is extremely important to ensure that all genders, ethnicities and age groups are correctly represented, if the AI-based product is then applied to multiple patients. Statistical accuracy is not necessarily equal to clinical accuracy. To address these challenges, technologists must address the limitations of machine learning algorithms and ensure the quality control of their application in diverse clinical environments and patient populations. and document and state their limitations.

Bias definition

Bias can be thought of as the systematic deviation of the result from the true estimate. This can happen in AI models if the training database is significantly different from the target population (defined by use) and can happen in the accuracy assessment of model if the database check is incomplete. As a result, algorithms may not be beneficial, such as for people whose data is not represented in the dataset.

Potential sources of bias

Knowledge of the intended context (field expertise) and the use of a model will inform the identification of sources of bias. Potential sources of bias in healthcare algorithms can arise at preprocessing (data collection, data preparation) and post-processing (model deployment and evaluation).

Pre-processing stage:

● Representation bias: If the targeted user group and the targeted patient group differ between the data used for the development process and the data is intended for.

● Learning hospital-specific features: Hospital departments often have a well-defined area of responsibility and rarely deal with cases outside of this area. Different departments also often use different medical devices. Learning characteristics such as hospital-specific parameters can bias the predictions.

● A 'case-control' database design could over-inflate measures of accuracy- i.e. if the data from a group of patients known to have a given condition is combined with the data from a group who do not, cases where there may be uncertainty and in which the model may not perform as well are excluded. The test database should ideally comprise a non-selected group of individuals reflecting the intended use population as closely as possible.

● Measurement bias: If variables were measured with different methods of different accuracies i.e., a positive result of a disease is more likely to be truly positive when it was measured with test A, rather than test B.

● Label bias: Annotation and label bias can arise when data was labelled by different practitioners with different levels of experience

● Assigning ground truth- if the ground truth (reference standard) in the test set is established by raters who have knowledge of the outcome of the AI model in the test group, this could inflate measures of accuracy.

● 'Over-curation' of the data- e.g. if poor-quality MRI scans are excluded, measures of accuracy may not reflect the real-world application where noise or artefacts may be common. Similarly, if cases with missing data are excluded from the population, the accuracy of the model in the real-world setting may be lower than in the test setting.

● Issues related to data integrity & data quality: Improper procedures on data inclusion and exclusion, input and output variable selection, pre-processing methods (data encoding-decoding formats, data compression and encryption, outlier and missing value treatment).

● Lack of standardized protocols and tools for data reproducibility (Who, When, Where, How, etc.), lack of interoperable data interfaces to collect and integrate diverse data types

Post-processing stage:

● Historical bias: An algorithm might be biased by social factors especially when training data was collected through services, surveys, or social media that are predominantly used by a certain social group (defined by ethnicity, religion, gender, …).

● Representation bias: An underrepresentation of minority or marginalized social groups in the training data can lead to unreliable predictions on underrepresented social groups. In this case, algorithmic fairness is not guaranteed.

● Algorithmic tuning: When business heuristics are applied to model outputs e.g. differential tuning of performance parameters in order to optimize for chosen business logic (e.g. differential diagnosis based on age, gender, ethnicity, etc.)

● Aggregation bias: arises during model building. If there are two or more distinct populations that are inappropriately combined. In that case, the population of interest is heterogeneous and a single model is unlikely to suit all minority groups.

● Evaluation bias: occurs during model iteration and evaluation. It can arise when the testing or external benchmark populations do not equally represent the various parts of the population it is applied on. Evaluation bias can also arise from the use of performance metrics that are not appropriate for the way in which the model will be used.

● Deployment bias: occurs after model deployment, when a system is used or interpreted in inappropriate ways.

Federated learning as Bias error mitigation method

To avoid data bias and thus model bias, the training dataset must contain many features, which is difficult in healthcare, as it is often not possible to merge datasets from different organizations. Federated learning is a machine learning method that allows to train models on decentralized pools of data. The configuration consists of several local nodes and one global node, where the gradients are calculated locally on the local nodes and they are combined into a global model. Therefore, there is no need to merge data sets for medical applications. This allows for the collection and use of training datasets, which are often not shareable due to data privacy concerns. Larger, more comprehensive datasets can be used to train models, such as geographically unbiased. Some of the most well-known myths about AI in healthcare are that it can replace the work of radiologists, or even that AI will dehumanize interactions with patients. These misconceptions, especially seen among medical students, may stem from the educational gap, as observed by a recent systematic review showing a lack of formal training on AI in general, while demonstrating a generally positive attitude towards AI in radiology (Santomartino and Yi, 2022). However, universities and healthcare associations are gradually rolling out formal training in addition to the already extensive portfolio of resources available for self-education, such as placement reports, seminars, etc. conferences and blogs dedicated to explaining AI.

2.3 Regulatory Concerns

Perceptions of a lack of transparency may stem from radiologists' doubts about aspects such as AI being used, how AI is trained and validated, or what steps the company takes regarding safety side (Rylands-Monk, 2022). However, there are many systems in place to ensure that AI software is well developed, such as IEC and ISO standards for medical software; regulatory agencies, such as the FDA, ensure that products can be marketed and therefore used in clinical settings; or specific regulations and laws related to the protection of patient information, such as HIPAA, that manufacturers must comply with.

The main regulatory concerns are the governance of autonomous AI systems, clear accountability and lack of responsible rules for the use of AI, and the lack of formal industry standards for the use of AI and performance evaluation (Dwivedi et al., 2019). According to the guidance published by the EU High-Level Expert Group on AI, ethical guidance presented in April 2019 for trustworthy artificial intelligence, AI must be legal, ethical and robust (European Commission, 2019).

A well-developed medical AI product has the potential to have tremendous clinical impact. In order to be open to the market, development must first follow clear legal roadmaps and practices.

The regulation of AI in radiology is increasingly associated with the medical device concept as reported in studies (Muehlematter, Daniore & Vokinger, 2021). AI-based tools in medical imaging is recognized as "Software as a Medical Device (SaMD)" by regulators.

The management principles of AI-based medical devices are comparable to managed software such as medical devices. However, there are specific additional considerations such as continuous learning capabilities, level of human intervention, model training, retraining, etc. for AI-based medical devices should be carefully considered and handled. All activities related to the design, development, training, validation, recycling and deployment of AI-based medical devices must be performed and managed in accordance with a standards-based quality management system (QMS) ISO 13485 standard.

The below Figure 5 illustrates the process of developing and deployment of the AI-based medical devices.

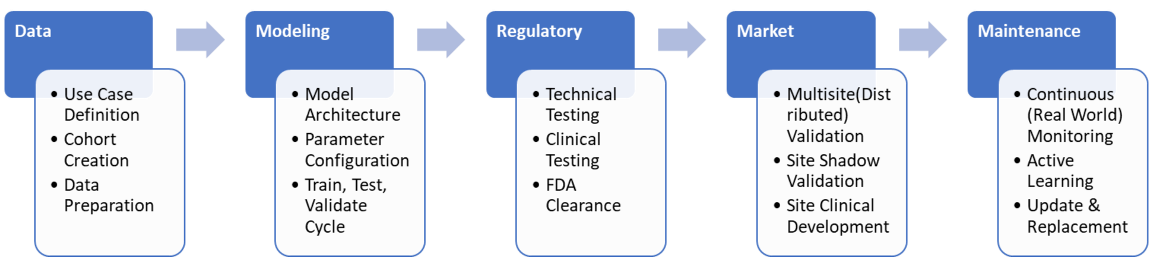

Figure 5 The AI Lifecycle (Dreyer and Coombs, 2020)

The dataset consists of different inputs and features/attributes selected for the AI-based medical device to produce the corresponding outputs. This can be in the form of diagnostic images, patient history records, physiological signals, medication records, healthcare professional handwritten texts, literature reviews, and more. The specifications or acceptance criteria for the selection of input data and features/attributes should be defined. Where pre-processing (e.g. signal pre-processing, image scaling) is required, this procedure should be clearly defined and included in the submission. An explanation should be provided for the preprocessing steps applied to the input data.

The source and size of the training, validation, and test datasets must be provided. Information about data set labeling, management, annotation, or other steps should be clearly presented. A description of the cleaning of missing datasets and data induction should be provided. The developer must also ensure that there is no duplication in the training and validation datasets. Rationale for the relevance and completeness of the selected data set and the factors that could potentially influence the outputs must be provided. In addition, any potential trends in the selection of training and validation datasets must be addressed and managed appropriately.

A description of the machine learning model (e.g., convolutional neural network) used in an AI-based medical device, including any underlying models (e.g., Inception V3 model) must be provided. The suitability of the model for the intended use of the AI-based medical device should be demonstrated. Any limitations of the model and, if possible, mitigation measures to address the shortcomings should also be explained. Model evaluation should be performed using a separate test dataset from the training dataset. The parameters (e.g. classification accuracy, confusion matrix, log loss, area under the curve (AUC)) selected to evaluate the performance of the selected machine learning model must be provided, including model evaluation results.

Based on the performance specifications of the AI-based medical device, the test protocol and test report should be provided. Information on control measures should be provided to detect discrepancies/abnormalities. Any limitations of the medical device and AI-based operating system should be clearly evaluated and also communicated, if any, to the user in the instructions for use or product labeling. Performance specifications such as instrument accuracy, specificity, and sensitivity must be provided (e.g. 90% accuracy, 91-93% sensitivity, 95% specificity). Verification and validation report(s) must be provided to support this performance claim. The presence of a valid clinical association between the output of an AI-based medical device and its target clinical status must be demonstrated by appropriately designed clinical studies (Health Sciences Authority, 2019). Device workflow, including how the outputs are used, the proposed or planned workflow should be presented and explained when deploying the equipment. When there is human intervention in the human system in the loop, the workflow must specify the level of intervention and the workflow step(s) for that intervention. In the event that data is collected after the implementation of an AI-based medical device (fixed version) and these data sets are used to recycle subsequent samples of the AI-based medical device, information about the training data update cycle interval will be provided. If a new set of data is collected that changes the original specifications and performance of your device, you must notify the HSA of the change. As with other software, change notifications will be required for changes to registered AI-based medical devices. This includes any changes to performance specifications, input data types, device workflow, level of human intervention, AI model selection, and more. The decision stream also applies to AI-based medical devices. For post-market traceability purposes, the exact version of the AI-based medical device should be provided and an explanation of how the version number is assigned and retrieved (Health Sciences Authority, 2019). The AI-based medical device with continuous learning has the potential to change behavior after deployment. The manufacturer shall define the learning process and shall have appropriate process controls in place to effectively control and manage the learning process. For example, appropriate quality checks are needed to ensure that the quality of the training dataset is equivalent to the quality of the original training dataset. Authentication procedures must be integrated into the system to closely monitor the overall learning and development of the AI-based medical device performance after the learning process. This is important to ensure that learning does not affect the defined specifications or output of the AI-based medical device. Because AI-based medical devices with continuous learning can automatically change behavior after deployment, manufacturers need to ensure that robust process controls are in place. This can ensure that the performance of the AI-MD does not degrade over time (Health Sciences Authority, 2019).

Once AI-based medical devices are deployed in the real-world environment, active monitoring, review and tuning are necessary. Developers and distributors should establish a process in collaboration with the implementers and users to ensure traceability and also implement mechanisms to monitor and review the performance of the AI-based medical device deployed in clinical setting. Such monitoring could also be in the form of autonomous monitoring embedded in the system. A robust surveillance model to ensure that the AI-based medical device especially those with continuous learning algorithms remain accurate and to prevent any concept drifts (Health Sciences Authority, 2019).

Ebrahimian et al. (2022) focuses on FDA-regulated AI algorithms. They reviewed 127 managed software in an effort to rank the information available and reported. They record (if any) the number of studies included along with other parameters, for example, specificity, specific receptor activity area under the curve, and sensitivity. They report the number of rejections. FDA-regulated physicians increasing from 2008 to 2021. Their critical review concluded that incomplete public data on validation/testing datasets in various algorithms cannot justify applications in healthcare because it cannot eliminate generalizations and/or bias rates.

The review identified three important areas for intervention: ethical issues, international regulatory frameworks, and bottlenecks in regulatory development. Ethically, areas of intervention have been identified that, in addition to those using traditional medical technology, include new areas due to the specificity of AI related to the production, datasets and avoid bias in them. Furthermore, regulatory studies have shown that these emerging regulatory approaches are inconsistent and different in the case of the United States, Europe, Canada or other countries. Different approaches have been used to address emerging issues, such as cybersecurity in medical devices. Studies have revealed a number of important issues and in particular the need to ensure a well-defined and rigorous roadmap for the approval and maintenance of AI-based medical devices. Among the recommendations for these issues, greater transparency of the approval and post-approval processes as well as the design of a full and open access database specifically for these MDs were proposed. Bottlenecks have also been identified specifically with regard to workload, suggesting that regulating it without scientific principles can be more dangerous than no regulation at all. References are also made to incomplete public data on validation/testing of datasets in various algorithms that are not guaranteed for healthcare applications as they cannot be excluded apart from generalizability and/or bias rate. More importantly, limitations and vulnerabilities emerge from these assessments. It is clear that they limit themselves to the review of regulatory tests that do not include an international approach. For example, some significant experience not considered, such as the NMPA experience, can act as a legal mediator between certain positions. Also related to ethical aspects, it is desirable to better share important experiences in the production of documents, such as those available in Europe.

When it comes to AI-based medical devices, the issue of bias is even more concerning (Giansanti, 2022). Research by Allen et al. (2021) goes this way. They report that trading algorithms are influenced by gender, ethnic, and social biases. This shows significant and dramatic implications for the design of healthcare algorithms. They also report that it is important to prevent bias in healthcare through strong stakeholder engagement to ensure robust and unbiased algorithms and datasets (Belenguer et al. , 2022).

Although AI technologies are attracting considerable attention in medical research, their implementation in real life still faces obstacles. The first obstacle comes from regulations. Current regulations lack standards for assessing the safety and effectiveness of AI systems. To overcome the difficulty, the US FDA tried to come up with the first guideline for evaluating AI systems. The first directive classifies AI systems as "general healthcare products", which are loosely regulated as long as the devices are intended for general healthcare purposes only and pose little risk to users. The second orientation justifies the use of real-world evidence to access the performance of AI systems. Finally, the guide clarifies adaptive design rules in clinical trials that will be widely used to evaluate the performance characteristics of AI systems. Soon after these guidelines were released, Arterys' medical imaging platform became the first FDA-approved deep learning clinical platform that can help cardiologists diagnose heart disease.

The second obstacle is data exchange. To function properly, AI systems need to be trained (ongoing) using data from clinical studies. However, once an AI system is deployed after initial training with historical data, continuous data feeding becomes an important issue for further system development and improvement.

The current healthcare environment discourages sharing of system data. However, a healthcare revolution is underway to promote data sharing in the United States. Reform begins with the revision of the payment system for medical services. Many payers, mainly insurance companies, no longer reward doctors by shifting treatment volume to treatment outcomes. In addition, the payer also reimburses a drug or treatment based on its effectiveness. In this new environment, all participants in the health system, doctors, pharmaceutical companies and patients, are encouraged to collect and exchange more information. Similar approaches are being explored in China.

2.4 Federated Learning Approach

The term “federated learning” was put proposed by McMahan et al. in 2016: "We call our method FL because the learning tasks are solved by a loose association of participating devices (what we call clients) coordinated by a central server." FL was originally defined as a distributed ML method that uses a lot of user data to train a central model (Hu et al., 2021). Federated learning was first introduced by Google as a decentralized distributed machine learning model (Chowdhury et al., 2022).

The goal of FL is to implement efficient distributed ML among many participants or many computing nodes on the principle of ensuring information security when exchanging big data, protecting mobile data and privacy as well as security ensuring compliance with the law. FL uses classic distributed ML framework and adopts distributed ML technology, but central server control is different from distributed ML's control. Researchers can extract and use data without breaking laws and regulations. In a broad sense, FL refers to a method by which data owners can achieve model training without downloading the data locally. The FL model is based on a local model uploaded by each participant, then the generic training model is sent back to each participant to achieve the same results as traditional ML without breaking any laws giving FL a privacy advantage.

Compared with traditional machine learning, FL algorithm consists of three main parts, namely learning algorithm and training method, data privacy protection mechanism and user incentive mechanism. The learning algorithm and training method refer to the iterative process of a set of servers after each client has trained the local model. It uses a privacy protection mechanism to protect the security of data and uses an incentive mechanism to incentivize customers to participate in generic model training. Therefore, the quality of of the FL algorithm is closely related to three aspects: (1) the quality of the learning and training model, and (2) the quality of the privacy protection mechanism. links and (3) quality of incentive mechanisms.

The design and training quality of the learning model

The FL algorithm is designed in two parts. One is for client-side local learning and the other is for server-side synthesis. The quality of learning, as measured by accuracy, precision, etc., is highly dependent on the design of the model and the training between the client side and the server side.

Quality of federated privacy

FL training can maintain local data, which improves privacy quality compared to centralized machine learning. In order to protect data privacy, the FL formation process must also ensure the following two facts.

(1) The trained model passed during training contains information about the original data. Therefore, we must avoid inferring the original data from the trained model.

(2) The server side only retrieves the agreed information from the client side and does not take any other redundant information. Therefore, it must be ensured that only intermediate results are submitted without additional information

Quality of federated privacy protection

FL's training process can maintain local data, which improves security compared to centralized machine learning. In order to protect data privacy, the FL formation process must also ensure the following two facts.

(1) The trained model passed during training contains information about the original data. Therefore, we must avoid inferring the original data from the trained model.

(2) The server side only retrieves the agreed information from the client side and does not take any other redundant information. Therefore, it must be ensured that only intermediate results are sent without additional information.

Quality Driven by an Insensitive Mechanism

The effectiveness of machine learning also depends on the quality and quantity of data used for training (Liu et al., 2020). While FL ensures that the data remains local, the customer must provide their resources, such as computing power, data samples, communication costs, etc. These factors may prevent a client from participating in FL without compensation. In FL, customers with large volumes of high-quality data often cannot achieve higher returns than training alone, and customers with small amounts of data are more interested in participating. Therefore, it is necessary to use an incentive mechanism to maximize the common good, ensure that private interests are not compromised, and to encourage more customers with high-quality data to participate.

The standard federated learning model is:

i) multiple client sites, each containing a local dataset that remained at the client site during training, connected to a global server;

ii) A global model is initialized in the global server and the weights of this global model are transferred to each local client location;

iii) Each customer site trains a local version of the global model on its respective dataset, then sends the updated model weights to the global server;

iv) The global server updates the global model by aggregating the weights it receives from local clients and then transmitting a copy of the updated global model to each client. The process that occurs between steps i to iv is called a loop, and during link formation, steps i to iv are repeated for several rounds until the population sample converges to a minimum. The most important aspect of this process is step iii. During this phase, all data used for training is kept strictly on local customers. The only information that is passed between the client and the server is the weight update. This allows multiple sites that aggregate their data to train synthetic models while maintaining data privacy. In step iv, the algorithm is used by averaging the association to sum the weights. In this algorithm, each update weight is weighted according to the size of the client dataset it comes from, relative to the size of the other client datasets. The aforementioned client-server topology is known as centralized federated learning. Another topology found in the study is decentralized federated learning, in which clients communicate between peers without a central server.

Federated learning can be divided into three main subtypes: horizontal federated learning, vertical federated learning, and transfer federated learning. These three subtypes follow the basic model of federated learning, which is a decentralized collection of data through the use of weight sharing and aggregation across multiple global clients and servers. They differ in how different their data sources are. In Horizontal federated learning, each client's website has different users in their data, but these users all share the same functionality exploited by the network. In Vertical Federated Learning, users are the same across all client sites, but each customer site's data functions differently, so the same user will be analyzed by different ways depending on customer's site. In Transfer Federated Learning, customer sites don't have common users or features, but their dataset tasks are still small, so grouping them together often leads to network training. Popular federated learning platforms are: OpenFL, PySyft, Tensorflow-Federated, FedML, Flower, NVIDIA Clara, Personal Health Train (PHT).

Federated Learning establishes a learning model based on distributed data sets. Unlike traditional machine learning, the algorithm will be implemented in the hospital's servers, which keep patient data in the facility. The raw data is distributed across client devices and forms a shared model on the server by aggregating locally computed updates, as depicted in Figure 6. Therefore, the algorithms Machine learning, such as deep neural networks, is trained on some local dataset contained in the local edge intersection (Ng et al., 2021). In turn, the algorithm will be continuously trained on local data. Such configurations will be deployed in multiple hospitals across regions using a centrally deployed cloud-based algorithm. The vendor then collects the algorithm changes and performs retraining to improve performance (Ng et al., 2021; Rieke et al., 2020).

The comparison between traditional and federated machine learning can be explained in Figure 6.

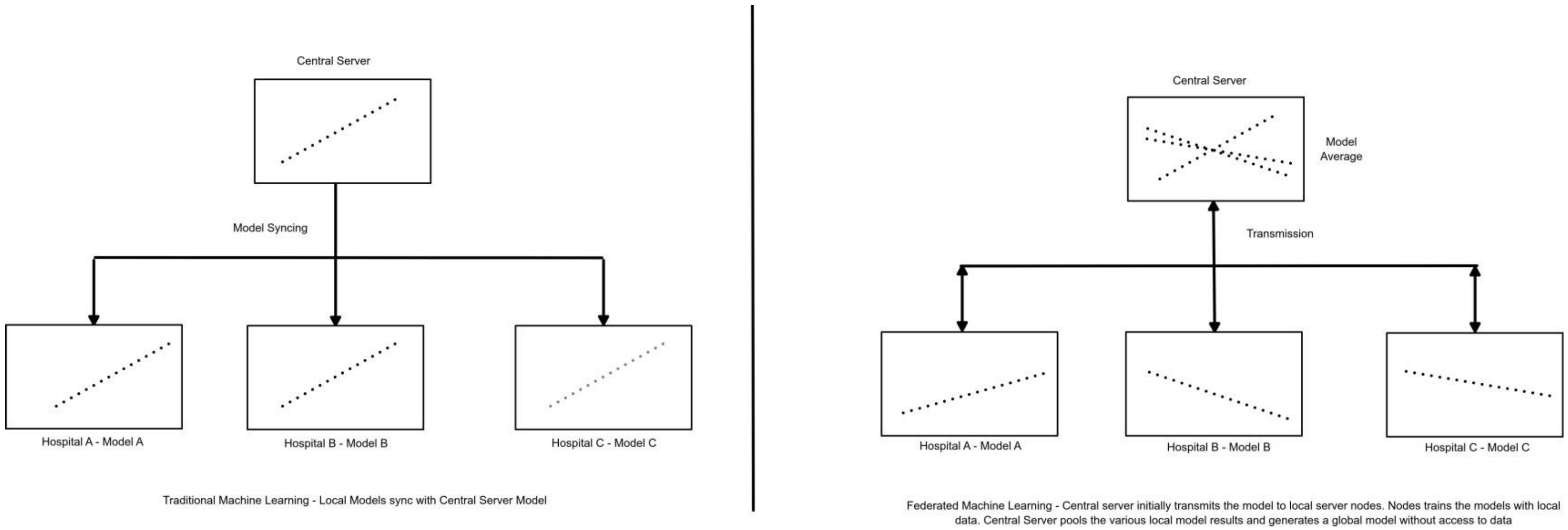

Figure 6 Comparison of Traditional Vs Federated Machine Learning (Ng et al., 2021; Rieke et al., 2020)

The benefit of federated machine learning is that the data resides on the hospital campus. Installation at multiple sites will provide the necessary data delivery. Federated learning will address concerns about data privacy protection, distributed data collection. Vendor-hosted central algorithms will ensure that all algorithms installed at multiple locations are trained and updated with distributed data (Naseriet al., 2021; Liu et al., 2021; Sohn and Kwon, 2020).

Data privacy and effective communication are the two main drivers of federated learning adoption. Data privacy is preserved during federated learning because no raw local data leaves the device. In addition, federated learning achieves higher communication efficiency by only exchanging model parameters or slopes. High data privacy and communication efficiency also promote scalability. As a result, many clients are motivated to participate in the training process.

Advances in genetics and biology have led to an increase in medical knowledge, which in turn has resulted in tremendous advances in diagnostic and therapeutic options. Due to the increasing age of the population and the patient's comorbidities, the volume and depth of information required to provide appropriate medical management has increased dramatically. As a result, electronic health records (EHRs), as a repository of this information, continue to evolve. EHR's high speed, authenticity, volume, and diversity qualifies as big data. The EHR contains fragmented data elements related to each patient's medical history. Linking this information is necessary to enable clinicians to use it for clinical purposes, including but not limited to predicting the risk of different diseases or the likelihood of adverse outcomes. EHRs have become an important source of real-world healthcare data that has been used to unify important biomedical research, including machine learning research. While providing a large amount of patient data for analysis, the EHR contains systemic and randomized biases across the globe as well as per hospital that limit the generalizability of the results. Federated learning is a viable method for connecting healthcare organizations' EHR data, allowing them to share their experiences, not their data, with guaranteed privacy. In these cases, the performance of the ML model will be greatly improved by iterative improvements in learning from large and diverse medical data sets. Several tasks have been investigated in fedearated learning in healthcare, e.g. patient similarity learning, patient representative learning, phenotypes and predictive models.

Federated learning has also activated the model. Predictions are based on a variety of sources, which can provide clinicians with additional information about the risks and benefits of treating patients earlier (Kashani et al., 2020).

Recently, a study on federated learning conducted by Linardos et al. (2021) showed that privacy and robustness have increased compared to traditional centralized learning.

Studies performed by different researchers (Stripelis et al., 2021; Yang et al., 2021; Linardos et al. 2021; Sarma et al., 2021; Dayan et al., 2021; Sheller et al., 2020) demonstrate a federated semi-supervised learning framework that can capture valuable insights from customers with only unlabeled data. The privacy of all patients is preserved because they do not need to share their own databases to collaborate on the joint training of models that improve the generalizability (Yang et al., 2021).

Since patient data sharing is not required to achieve full learning in federated learning, this overcomes technical and data ownership issues and helps meet requirements of privacy regulations (e.g. the European General Data Protection Regulation (GDPR) and the US Health Insurance Information and Accountability Accountability (HIPAA)) therefore facilitate collaboration between multiple organizations (Sheller et al., 2020).

Salam et al. (2021) demonstrated that the federated machine learning model has better predictive accuracy and attenuation, but higher execution time than the traditional machine learning model. The study compared the effectiveness of federated learning with traditional learning by developing two machine learning models to detect COVID-19 using chest X-ray (CXR) images of patients. patients infected with COVID-19.

In another study validating the federated learning model using actual clinical prostate imaging data (Sarma et al., 2021), successfully demonstrated in 3 institutions, reported that federated learning improved performance on both organizationally and externally maintained test data sets, allowing for greater generalizability in clinical use 2021). Compared with conventional federated learning (FL), clustered federated learning (CFL) is an emerging concept and is expected to better cope with differences in the distribution of data from different sources (Qayyum et al., 2021).

Federated learning investigated in the field of neuroimaging by collaborative learning a global brain age prediction model (Stripelis et al., 2021). Based on our literature search, we have determined that federated learning has been explored in many cancer studies, where the goal is to compare federated learning with data analysis methods, whether to focus normally on performance or develop new methods to address the various challenges encountered using federated learning. Learning (e.g. domain name change, missing label etc.,). In the most common training scenario, researchers simulate a federated learning environment by taking an existing dataset and dividing it into subsets using a partitioning scheme, where each subset represent a client in an federated learning group.

Federated learning has been applied to brain tumor detection in several studies. In the study by Chowdhury et al., (2022), the authors used Kaggle's "Brain MRI Segmentation" dataset to segment low-grade gliomas (Muehlematter, Daniore, & Vokinger, 2021), dividing the whole datasets at 5 “customer” locations. The authors designed a network that achieves state-of-the-art results in the glioma segmentation task, and these results remain consistent when applied to the federated learning framework. In the paper by Sheller et al., (2019), two distinct federated learning environments for brain tumor segmentation were simulated using the BraTS dataset (Menze et al., 2015). In both environments, the federated learning model was compared with two other cooperative learning techniques and outperformed both. It also achieves almost 99% of the DICE scores obtained from a model trained on the entire dataset without hierarchies. Similarly, Sheller et al., (2020) demonstrated comparable performance between the linked and shared mean for brain tumor segmentation on the BraTS dataset (Menze et al., 2015). Sheller et al., (2020) also shows how federated learning improves the learning of each participating organization both in terms of performance on local data and performance on data from unknown domains. In the paper by Sheller et al., (2019a), the authors presented a comparison between federated learning and individual training of 3D-Unet models for glioblastoma segmentation in 165 images multiparameter structural magnetic resonance imaging (mpMRI). Federated learning has been proven to deliver superior quantitative results.

Additional studies have explored federated learning on many other types of cancer, including those that are less common. Some of the types mentioned in the use cases we looked at include: skin cancer, breast cancer, prostate cancer, lung cancer, pancreatic cancer, anal cancer, and thyroid cancer.

Federated learning for COVID-19

Since the outbreak of COVID-19, the global pandemic disease in 2020, it has attracted attention as a new research topic in the field of medical AI. Common symptoms in patients with COVID-19 are lung tissue damage, leading to cell destruction and lung fibrosis. The results of chest X-rays or CT scans are regularly reviewed to determine the stage of the disease, and various research trials are underway to classify medical images collected by machine algorithms. Through analysis of chest images, COVID-19 symptoms are effectively differentiated from difficult to classify pneumonia, and at the same time, protection of personal information by association learning technique is also observed. Zhang introduced a dynamic association-based learning algorithm for diagnosing COVID-19 infections, using medical imaging datasets collected from Kaggle and GitHub (W. Zhang et al., 2021). A test was conducted with a framework in which three clients participated in training and tested accuracy and convergence time were evaluated using three different training methods: GhostNet, ResNet50 and RestNet101. Save et al. (2020b) recently applied a federated learning-based Covidnet algorithm to distinguish between chest X-ray images of pneumonia patients and COVID-19 patients. Although inferior to ResNet models, it exhibits similar classification performance as MobileNet, a lightweight model, and shows the ability to analyze medical images through federated learning. Unlike the previous two studies, Dayan's team demonstrated the benefits of federated learning through the involvement of multiple medical institutions to analyze COVID-19 patient data (Dayan et al., 2021). A total of 20 institutions participated in the creation of the federated learning health data classifier model, which has the same architecture as the concept shown on the right in Figure 6. Compared to the classifier case given created with a local organization, using 20 organizations showed an average performance improvement of 13.9%. Kumar and associates. (2021) also developed a framework that unifies capsule networks and blockchain-based federated learning for the diagnosis of lung CT images of patients collected from different hospitals. Data sets were collected from three hospitals and performance in terms of sensitivity and specificity was improved by more than 7% compared to existing benchmark machine learning models. Based on the above test results, the performance degradation is not significant compared to centralized machine learning when federated learning algorithms consider PHI protection for COVID-19 diagnosis. Since centralized machine learning performs better when using the same amount of data for training, federated machine learning can use more data to achieve higher performance with the benefit of training through Customer data is distributed. Various mutated COVID viruses are emerging in countries around the world, and it is important to form a topology based on understanding the data collected by each organization. While many studies in the medical field point to the benefits of adopting federated learning, there are still research issues that need to be addressed for better use.

2.5 Challenges in Federated Learning

Although federated learning has its advantages, it does not solve all the problems inherent in learning about medical data (Rieke et al., 2020). In the study published by Xu et al., (2021), it was concluded that successful model training always depends on factors such as data quality, bias, and normalization. Xu et al., (2021) also suggest that these issues should be addressed for both linked and unconnected learning efforts through appropriate measures by careful study design, common protocols for data collection, structured reporting, and implementation of sophisticated methods for uncovering potential biases and stratification.

Federated learning is more vulnerable to security threats due to advanced computer networks. Attacks against federated learning (FL) techniques have highlighted weaknesses in both robustness and privacy (Naseri et al., 2020).

Another challenge is the return on investment (ROI) in both time and money. When it's unclear how long preparation, installation, training, and possibly workflow tuning will take, this naturally worries potential users (Beck, 2022).

Federated Learning has the ability to connect any medical institution, hospital or isolated device to share their experiences with guaranteed confidentiality. Currently, however, the healthcare system is experiencing data clutter and efficiency problems. There is no uniform data standard for the data collected, and the quality of data from multiple sources is uneven. The key to improving machine learning models is to ensure quality data with clean, accurate, and complete data if we are dealing with a federated learning scenario (Xu et al., 2020).

Xu et al., (2020) list some directions or open questions that may be encountered when applying federated learning in healthcare, data quality, expertise integration, mechanisms incentive, custom, precision of the model.

Assessing whether artificial intelligence should be used

There are risks in overestimating what AI can achieve, unrealistic estimates of what can be achieved as AI evolves, and unproven products and services that remain untested rigorously tested for safety and effectiveness. This is partly due to the lingering appeal of "technological solutionism", in which technologies such as AI are used as a "silver bullet" to break down deeper social barriers, structural, economic, and institutional. The allure of technology solutions and the promise of technology can lead to overemphasizing the benefits and ignoring the challenges and problems that new technologies such as AI can bring. This can lead to unbalanced health policies and poor investment by under-resourced countries, and pressure on PREs to reduce public spending on health. It can also divert attention and resources from proven but underfunded interventions that can reduce morbidity and mortality in LMICs.

First, the AI technology itself may not meet the standards of scientific validity and accuracy currently applied to medical technologies. For example, digital technologies developed in the early stages of the COVID-19 pandemic do not necessarily meet an objective standard of effectiveness to justify their use. AI technologies have been introduced as part of the pandemic response without sufficient evidence, such as randomized clinical trials or protective measures. Emergencies do not justify deploying unproven technologies; In fact, efforts to ensure that resources are allocated where they are most needed should increase the vigilance of businesses and governments (such as regulators and ministries of health) to ensure that the technology is accurate and effective. Second, the benefits of AI can be magnified when incorrect or overly optimistic assumptions are made about the infrastructure and institutional context in which the technologies will be used when needed. Intrinsic needs of technology use cannot be met. In some low-income countries, the financial resources and infrastructure for information and communication technology are inferior to that of HICs and the large investments required may discourage its use. The quality and availability of data may not be suitable for the use of AI, especially in LMICs. There is a risk that poor quality data will be collected for AI training, which could lead to false prediction patterns in the data rather than actual clinical outcomes. It is also possible that the lack of data as well as poor quality data distort the performance of the algorithm, resulting in incorrect performance or the AI technology being unavailable. In addition, significant investment may be required to make the heterogeneous datasets collected in PRITI usable. Collecting aggregate data in under-resourced settings is difficult and time-consuming and must consider the additional burden on public health workers. There is no data on the most vulnerable or disadvantaged populations, including those lacking health services, or the data may be inaccurate. Data can also be difficult to collect because language barriers and mistrust can lead people to provide inaccurate or incomplete information. Often, unlinked data is collected, which can degrade the overall quality of the data set. Wider concerns about data collection and use, as well as biases in data, which will be discussed below.

There may not be appropriate or enforceable regulations, stakeholder involvement or oversight, all of which are necessary to ensure that ethical and legal issues can be resolved. decide. establishment and human rights are not violated. For example, AI technologies may be introduced into countries that do not have up-to-date privacy and data protection laws (especially for health-related data) or without oversight by regulatory agencies. data protection agency to strictly protect the security and privacy of individuals and communities. In addition, LITI regulators may not have the capacity or expertise to evaluate AI technologies to ensure that system failures do not affect diagnosis, monitoring, and treatment. Third, there may be enough ethical concerns for a particular use case or AI technology, even if it provides accurate and useful information and insights, to preclude a particular use case. AI technology that can predict which individuals are more likely to develop type 2 diabetes or HIV infection can benefit individuals or communities at risk but can also lead to unnecessary stigma. against individuals or communities, whose choices and behavior are questionable or even criminalized for the over-medication of normally healthy individuals, creating unnecessary stress and anxiety, and leading to aggressive marketing by individuals by pharmaceutical companies and health services for other profits. In addition, some AI technologies, if not carefully implemented, can exacerbate disparities in healthcare, including those related to ethnicity, socioeconomic status, or gender.

Fourth, like all new medical technologies, even if AI technology does not come with ethical caveats, its benefits may not be justified by additional costs or expenses (in addition to AI, information, and communication technology infrastructure) related to procurement, training and required technology investments. Robotic surgery may yield better results, but the opportunity cost associated with the investment must also be considered. Fifth, it may not be enough to determine whether AI technology is appropriate and appropriate for the context of LMICs, such as linguistic and canonical diversity within a country or across countries. For example, a lack of investment in translation could mean that some apps don't work well or are simply unusable by users. Such a lack of foresight leads to a larger problem, which is that many AI technologies are designed by and for high earners as well as individuals or companies that do not fully understand the specifics, body of technology, scores of the target population.

However, unrealistic expectations about what AI can achieve can prevent its use unnecessarily. So, machines and algorithms (and the data used for algorithms) are said to be perfect in the public imagination, while humans can make mistakes. Healthcare professionals may overestimate their ability to perform their duties and overlook or underestimate the value of algorithmic decision-making tools, the challenges that can be managed, and evidence that benefits are measurable. Failure to use this technology can lead to preventable illness and death, making the failure to use a certain AI technology to blame, especially if the standard of care has shifted to use it. For healthcare professionals to conduct such assessments, they require greater transparency about the performance and usefulness of AI technologies, as well as effective regulatory oversight. The regulator's role in ensuring rigorous testing, transparent reporting of results and monitoring performance. Even after the introduction of AI technology in the healthcare system, its impact must be continuously assessed when used in the real world, as well as the performance of the algorithm if it learns from the data different from its training data. Impact assessment can also guide decisions to use AI in the medical field before and after its introduction. An assessment of whether AI technology should be introduced in a low-income country, or a resource-poor environment may lead to a different conclusion than that in a high-income country. Risk-benefit calculations that do not favor the specific use of AI could be interpreted differently for a low-income country, such as a lack of sufficient medical staff to do the work. Certain tasks may forgo the use of more accurate diagnostics tools, so that individuals receive incorrect diagnoses and mistreatment. However, the use of AI in resource-constrained environments must be carefully scaled up to avoid situations where many people receive an accurate diagnosis of their health condition but do not have access to appropriate treatment. Healthcare professionals are tasked with providing treatment after testing and confirming the disease, and the relatively low cost of performing AI diagnostics must be accompanied by careful planning to ensure people are not left behind. Predictive tools to anticipate outbreaks will need to be complemented by robust monitoring systems and other effective measures.

Liability for use of artificial intelligence in clinical care

Using AI to support or enhance clinical decision-making raises several questions. Are doctors liable if they follow an AI suggestion that leads to a medical error or if they ignore a suggestion that could avoid illness or death? The answers to these questions largely depend on other options, such as the types of behavior that are encouraged or discouraged by the legal system and the standard of care when the use of AI in clinical practice is concerned. Another option is whether liability rules should encourage clinicians to rely on AI to inform and confirm their clinical judgment, or to deviate from their own judgment if the calculation makes an unexpected conclusion. If liability rules penalize healthcare providers for relying on findings from AI technology that turn out to be inaccurate, they can only use the technology to identify take your own judgment. While this could protect them from liability, it would discourage the use of AI to its full potential, which is to enhance rather than just confirm human judgment. If physicians are not penalized for relying on AI technology, even if its recommendations go against their own clinical judgment, they may be encouraged to use these technologies more widely to improve their health. improve patient care or at least consider using them to challenge their own assumptions and conclusions. Whether or not a doctor uses AI also depends on the prevailing standard of care. If AI technologies are deemed to deviate from the standard of care or are not recognized as meeting the standard of care, physicians will not be encouraged to use them, as failure to meet the standard of care will result in disincentives and prohibits (but does not completely) medical errors. If the standard of care calls for the use of AI technologies, then physicians will essentially be required to integrate their use into clinical practice.

A separate but related issue is the responsibility of hospitals and healthcare systems to select a particular technology. Hospitals may be liable if they are not diligent in selecting technology or introducing, using, or maintaining it. In general, a hospital can be indirectly responsible for the errors of clinicians working in the hospital. Therefore, hospitals are encouraged to be both cautious in their choice of technology and to ensure that clinicians have clear instructions on how to use them to care for their patients and how to avoid the errors that lead to them. Liability for clinicians and hospitals. One possibility is to attribute the hospital's responsibility to a "sloppy certification". As a rule, hospitals are liable if they do not fully consider the credentials and work history of medical staff and physicians, they may have similar obligations when referring WHO. To do this, hospitals and healthcare systems need the information and tools to identify the right AI technologies for clinical use. Hospitals should also have a duty to restore control of a process or system that has been automated and is now presenting actual or potential risks that were previously unforeseeable.

Challenges in commercialization of artificial intelligence for health care

The practices of the biggest tech companies in the health AI space present a variety of ethical challenges, although some concerns apply equally to small and medium-sized companies. The use of AI for health has been driven by businesses – from small startups to large tech companies – primarily through significant advocacy and investment. Supporters of the growing role for these companies expect them to be able to raise capital, in-house expertise, IT resources, and data to define and create new applications that support service providers. and health system. During the COVID-19 pandemic, many companies have sought to provide services and products to respond, many of which involve forms of public health surveillance. This raises a number of ethical and legal concerns, discussed throughout this document. A number of services are already widely used in the health sector for "logistics" and health system management functions. Some companies involved in technology development, such as the pharmaceutical and medical device industries, are integrating AI into their processes and products, and insurers are using AI to evaluate risk pricing or even automating insurance provision, which can present ethical issues. related decision making by algorithm. An important application of AI in healthcare is to aid in diagnosis, treatment, monitoring and adherence. Such applications can benefit healthcare systems; however, concerns have arisen in the past as more tech companies, especially larger ones, have moved into the healthcare sector.

A common problem is the lack of transparency. While many companies know a lot about their users, their users, civil society, and regulators know very little about what the company does, including how they (and the government) operate, which has a significant impact on the public interest.

Their activities remain partially concealed due to trade-secret agreements or a general lack of transparency practice obligations, including the role these companies play in the healthcare and data is selected, data is collected, used to form and validate AI algorithms. Without transparency (and accountability), these companies have little incentive to act in ways that don't cross certain ethical boundaries or reveal deeper issues in technology, data or their model. Many companies prefer to keep their algorithmic models confidential and proprietary, as complete transparency can lead to criticism for both the technology and the company.

The second major concern is the overall business model of the largest technology companies including the active collection and use of data to make their technology efficient and the use of redundant data surplus for commercial purposes. As a result, over the past decade, there have been several examples of big tech companies using large data sets of sensitive health information to develop AI technology for healthcare. While this health data may have been collected and used to develop useful health AI technologies, the data was not collected with the express consent of those providing the benefits of data to these companies may go beyond what is required to provide the product, and the companies may not offer the same benefits to the people who created the data in the first place. The collection of such sensitive health information may give rise to legal concerns. First, even if the data is anonymized by the company purchasing it, the company will be able to combine the data and anonymize relevant datasets from the amount of information it already has from other sources. Second, some big tech companies have been charged and even fined for data mishandling, and this concern could be heightened for companies that often collect sensitive health data.

Thirdly, as firms continue to accumulate large amounts of data, this can introduce anti-trust concerns (although it may not lead to regulatory enforcement), related to the growing market power of such companies, including barriers to smaller companies that may wish to enter an AI market. An additional concern is the growing power that some companies may exert over the development, deployment and use of AI for health (including drug development) and the extent to which corporations exert power and influence over individuals and governments and over both AI technology and the health-care market. Data, computing power, human resources and technology can be concentrated within a few companies, and technology can be owned either legally (IP protection) or because the size of a company`s platform results in a monopoly. Monopoly power can concentrate decision-making in the hands of a few individuals and companies, which can act as gatekeepers of certain products and services and reduce competition, which could eventually translate into higher prices for goods and services, less consumer protection or less innovation. While the growing role of large companies in the USA, such as Google, Facebook and Amazon, in the development and provision of AI for health care has been under scrutiny, large technology companies in China and other Asian countries are playing a similar role in health through such services and technologies. These include Ping An, Tencent, Baidu and Alibaba, both of which are building their own technology platforms and collaborating with user platforms like WeChat to reach millions of people in China. For example, Tencent is investing in at least three key areas of healthcare: AI-based technology to support diagnosis and treatment, "smart hospitals" to provide online service networks, and online services. connection. data through a smart health card (primarily raising privacy and data usage concerns) and a "vehicle" for providing health information to users online. Alibaba is working with hospitals to forecast patient needs to allocate medical staff and develop AI-powered radiology diagnostic tools.

The market power and control of such large corporations may be part of the "first-mover" advantage that some large corporations can enjoy through their access to AI for health. Even if the data used by a firm (for example, data from a public health system) could be used by others, other firms might be discouraged or unable to replicate use of such data for a similar purpose, especially if another company has already done so.

Such power also means that the rules set by certain companies can force even the largest and wealthiest governments to change course. For example, during the COVID-19 pandemic, Google and Apple introduced a technical standard for where and how data should be stored in proximity-tracking applications that differed from the approach preferred by the governments of several HIC, which resulted in at least one government changing the technical design of its proximity-tracking application to comply with the technical standards of these two companies.

Although the approach of these companies may have been consistent with privacy considerations, the wider concern is that these firms, by controlling the infrastructure with which such applications operate, can force governments to adopt a technical standard that is inconsistent with its own public policy and public health objectives. When most data, health analytics and algorithms are managed by large technology companies, it will be increasingly likely that those companies will govern decisions that should be taken by individuals, societies and governments, because of their control and power over the resources and information that underpins the digital economy.

This imbalance of power also affects those who should be treated fairly by their government or at least, if treated unfairly, their government could be held accountable for the injustice. Equality. Without a strong government role, corporations can ignore the needs of people, especially those on the margins of society and the global economy. Close government oversight and good governance are essential in this area. Monitoring mechanisms can be built. If these partnerships are not carefully designed, they may lead to misuse of resources (usually patient data) or conflicts of interest in decision-making in these partnerships, or may prevent or limit the use of the regulation to protect the public interest if necessary.

The Data Protection Act is a “rights-based approach” that provides standards for regulating data processing that both protect the rights of individuals and establish obligations for controllers and processors. Data protection law also increasingly recognizes that individuals have the right not to follow decisions guided only by automated processes. More than 100 countries have passed data protection laws. A well-known part of data protection law is the General Data Protection Regulation (GDPR) of the European Union (EU); In the United States, the Health Insurance Portability and Accountability Act, enacted in 1996, applies to the confidentiality and security of health data.

Regulations on AI technologies are likely to be developed and implemented by health regulators tasked with ensuring the safety, effectiveness, and rational use of technologies for the development of healthcare. strong. and therapy. A group of WHO experts is preparing to review AI regulation for the discussed health areas that stakeholders, including developers and regulators, should take into account when reviewing public companies. New technology in healthcare. These include documentation and transparency, risk management and lifecycle approaches, data quality, clinical analysis and validation, engagement and collaboration, as well as privacy and data protection. Many regulators are preparing considerations and frameworks for the use of AI and these need to be reviewed, possibly with the relevant regulator. Governance of AI through legal frameworks and ethical principles needs to be taken into account.

Governance and oversight of large technology companies

Large tech companies, especially those from China and the United States, should play a central role in the development and implementation of medical AI, through partnerships, in-house AI development. or acquire other businesses. The role and involvement of these companies raise additional considerations for private sector oversight. Large technology companies, few, hold significant power in the field of AI through human, economic and technical resources, accumulated data about their products and services, values. The main influence they can exert through their relationships and partnerships with governments and organizations, employees, and the ability to use their platform to recommend products and services to many users who regularly connect to their platform. Over time, big tech companies can develop even more diverse products and services. Google is developing a range of diagnostic apps whose safety and effectiveness are still under scrutiny, and its parent company, Alphabet, has launched a new health insurance service that will partner with Swisse.

Companies may also introduce products and services that can compete with, replace, or introduce a function or process normally operated by the government. Tencent introduced an app that uses information voluntarily provided by individuals to determine the type of healthcare provider a patient should see, in part to address the reality in China where the patient is ill. Individuals use their research or intuition to seek medical advice from experts in fields unrelated to them. The growth of telemedicine provides platforms through which companies can bring patients to their platforms and they recruit doctors to provide telemedicine on the platform. this platform. For example, Tencent WeDoctor, in partnership with the government, has registered at least 240,000 providers on its platform, 2,700 hospitals and 15,000 pharmacies. At least 27 million monthly users access the "healthcare collaboration platform" for remote or AI-guided consultations. The user is then matched with the appropriate healthcare system specialist. This may mean that in the long run, governments may not so much regulate the companies that provide these services but depend on them to fill gaps and manage parts of the system. system. health care. Technology companies can provide the necessary infrastructure to operate health services, which also creates government reliance on the services and companies' capabilities, rather than regulating the industry to satisfy the needs of government and the public. As noted above, tech companies have started publishing tutorials on how to use AI; however, they are sometimes seen as "ethical cleansing", which can create a lack of accountability (retrospective responsibility for harm), not involving the public in the development process. their development and may be administered in a manner that is not transparent to the public or government, without public participation or by an independent body responsible for monitoring compliance with the principles.

2.6 Consumer Perception

Various models have been developed to explain the user adoption of new technology and these models introduce factors that can affect user acceptance (Taherdoost, 2018). Technology Adoption Model (TAM) introduced by Davis (1989) explains the adoption and usage of technology by individuals in an organization. Kim et al. (2007) argued that TAM is limited, however, in explaining the adoption of new ITs or behaviors, such as mobile commerce. The model suggests that several factors influence their decision about how and when they will use it, notably perceived usefulness, and ease-of-use. User's attitude influences the behavioral intention in adoption of new technology. While external variables can influence the user's attitude, however the perception may depend on age and gender.

The Value-based Adoption Model (VAM) is more suitable for explaining the dual role of technology users and service consumers (Taherdoost, 2018). The value-based Adoption Model (VAM) based on the mental accounting theory thoroughly explains the costs and benefits associated with choices made by users. A comparative study also supports the theory that the value-based adoption model could best explain the consumer acceptance of AI-based services compared to other widely used technology acceptance theory (Sohn et al., 2020).

Safi, Thiessen and Schmailzl (2018) conducted study to determine and evaluate the factors that influence acceptance and resistance to achieve a successful implementation of new technologies. confirmed that adoption of new technologies in health care depended on individual opinions of the factors relating to them. The acceptance of digital solutions and innovative medical technology by patients and professionals relies on understanding their anxieties and feelings of insecurity.

The acceptance of use of AI-based tools in healthcare services is challenging as consumers express fundamental concerns about the technology (Turja et al., 2020). Generally, the perceptions of radiologists have not been considered and details of datasets used for training and implementation approaches for adoption of machine learning tools have not been reported (Masud et al., 2019). A study reports that, in general, there is a lack of trust in the features of AI systems (Sun and Medaglia, 2019). For instance, individuals may not trust AI 's performance and diagnostic ability for treatment purposes. indicated that the consumer's perception of trustworthiness is affected by the level of autonomy of AI systems.

Understanding the barriers and customer perception in adoption of these services are critical to invest in R&D and strategize sustainable business models (Esmaeilzadeh, 2020).

In study conducted by Esmaeilzadeh, (2020) on use of AI-based tools for healthcare purposes, the results indicate that individuals' positive perceptions toward AI-based devices can lead to a higher intention to use AI. This study focused on general AI-based tools and did not distinguish traditional and federated learning-based tools.

As general conclusion, further studies are essential to examine the acceptance of consumers using federated machine learning approach in AI based tools in radiology medical imaging. Federated learning promises to bring trust since it addresses the privacy concerns and build robust generalizable algorithm with multisite collaborations. The significance of understanding the influence of federated learning approach towards user adoption intent of AI-based tools is critical for AI developers and vendors to invest and build sustainable business models.

2.7 Summary of Literature Review

AI algorithms are likely to suffer from a range of shortcomings, including inability to apply outside the field of training, bias, and fragility (which tends to be easily fooled). Important factors to consider include changing the data set, randomly adjusting for confounding factors instead of the actual signal, the propagation of unintended biases in clinical practice, and the provision of algorithms. interpretability, develop reliable measures of model reliability, and challenge generalization to different populations.

Particularly important for the EHR algorithm, it is easy to overlook the fact that all inputs are generated in non-permanent environments with variable patient numbers, where clinical and operational behavior changes over time. The introduction of a new prediction algorithm can lead to changes in practice, resulting in a new distribution compared to the distribution used to train the algorithm. Therefore, methods of identifying deviations and updating models to cope with the reduced performance are essential. Mitigation measures to manage this effect include careful quantification of performance over time to proactively identify problems, in addition to the need for periodic retraining. Data-driven testing procedures have been proposed to recommend the most appropriate update method, from simple recalibration to full model retraining, to maintain performance over time (Kelly et al. associates, 2019).

Machine learning algorithms will use all available signals to achieve the best possible performance in the data set being used. This may include exploiting unknown confounding factors that may be unreliable, weakening the algorithm's ability to generalize to new data sets. For example, in a classic example, the machine learning model did not learn the intrinsic difference between dogs and wolves, but instead learned that wolves are often depicted standing in the snow, while dogs are often depicted as standing in the snow.

Similar concerns exist in the medical field. In one study, an algorithm was more likely to classify a skin lesion as malignant if the image contained a ruler, because the presence of a ruler correlated with an increased likelihood of a cancerous lesion. The presence of surgical marks on the skin has also been shown to falsely increase the melanoma probability of the deep learning model and, consequently, the false-positive rate. In another study, hip fracture detection was supported by confounding factors, including CT scan pattern and scans marked as 'urgent'. Another algorithm for detecting pneumonia on chest x-rays can accurately identify hospital equipment and services, learning the link between portable X-ray machines and pneumonia. Ongoing work is needed to understand the specific features learned by neural networks and will be essential for generalization in many healthcare settings.

Most AI systems lack reliable generalization, let alone clinical applicability, to most types of medical data. A fragile model can have blind spots that can make particularly bad decisions. Generalization can be difficult due to technical differences between sites (including differences in equipment, coding definitions, EHR systems, laboratories, and testing equipment) as well as differences in clinical practice and local management.

The problem of generalizability is closely related to the problem of discriminatory tendencies. Machine learning's blind spots may reflect societal biases at worst, with the risk of unintended errors or unknown accuracy in small groups and concerns about the potential for accuracy. ability to amplify biases present in historical data. Studies show that in some current contexts, the weaknesses of AI systems disproportionately affect groups that are already disadvantaged by factors such as race, gender, and socioeconomic background.

Inequity in the algorithm can be distilled into three components, namely

(1) model bias (i.e., models chosen to best represent the majority and not necessarily groups). underrepresented),

(2) model variance (due to incomplete minority data) and

(3) outcome noise (the effect of a set of unobserved variables potentially interacting

with the model's predictions, which can be avoided by defining subpopulations to measure additional variables)

The researchers ensured that the steps were taken correctly to quantify accuracy. bias before deploying the models. Algorithms should be designed with the global community in mind and clinical validation should be performed using a cross-section of the intended implementation team. Careful analysis of performance by subgroups of the population should be conducted, including age, ethnicity, sex, socioeconomic class, and location. Analysis to understand the impact of a new algorithm is especially important, i.e., if the disease spectrum detected by the AI system differs from current clinical practice, the pros and cons of the development must be assessed of that algorithm have this different spectrum of diseases. In mammography, it is possible to detect less severe ductal carcinoma in situ, potentially leading to increased treatment with little outcome benefit. Potential pilots in health systems should be undertaken to understand product features and identify potential pitfalls during actual implementation.

Algorithms have been shown to be sensitive to the risk of adversary attack. Although somewhat theoretical at this point, an adversarial attack describes an alternative efficiency model that is susceptible to manipulation by inputs explicitly designed to fool them. For example, in one study, an image of a benign mole was misdiagnosed as a malignant mole by adding conflicting noise or even just rotation.

A fundamental element for the safe and efficient implementation of AI algorithms is the development of the necessary regulatory frameworks. This poses a particular challenge given the current pace of innovation, the significant risks involved, and the potentially flexible nature of machine learning models. Proactive regulation will bring confidence to clinicians and health systems. Recent guidance from the US Food and Drug

Administration has begun to develop a state-of-the-art regulatory framework to ensure safe and effective artificial intelligence devices can work for patients data is not available for machine learning. Data is often stored in countless medical imaging systems, pathology systems, EHRs, electronic prescribing tools, and insurance databases, which are difficult to piece together. Through unified data formats, such as Rapid Healthcare Interoperability.

It is also important to consider the regulatory impact of innovations and upgrades that AI product vendors are likely to develop over the life of the product. Some AI systems will be designed to improve over time, challenging traditional evaluation processes. As AI learning is continuous, periodic system-wide updates after a full assessment of clinical significance are better than continuous updates that can lead to bias. Developing continuous performance monitoring guidelines for continuously calibrating models using human feedback will help identify performance deviations over time.

Even with a highly efficient algorithm that overcomes all the above challenges, the human barriers to adoption are substantial. To ensure that this technology is accessible and beneficial to patients, it is important to focus on clinical applicability and patient outcomes, methods of algorithmic interpretation, and improvements and better understand human-computer interaction.

Several studies (Turja et al., 2020; Masud et al., 2019; (Sun and Medaglia, 2019) identified that users of traditional machine learning in medical imaging express concerns like mistrust, bias, generalizability, explainability and conformance with data privacy regulations resulting in very slow adoption rate. There is fundamental concern expressed by healthcare professionals with trustworthiness and robustness of traditional machine learning algorithms (Kaissis et al., 2020).

The experimental studies (Stripelis et al., 2021); (Yang et al., 2021); Linardos et al. (2021) (Sarma et al., 2021)Dayan et al., 2021; (Sheller et al., 2020)) conducted in various modalities of radiology medical imaging have positive results showing federated learning performs better than traditional machine learning and theoretically improve the robustness of algorithm while preserving the patient data. The analysis of the literature provides sufficient evidence that federated learning offers easy scalability, flexible training scheduling, and large training datasets through multi-site collaborations, fulfilling the essential conditions to the successful deployment of an AI solution. Federated learning approach is expected to reduce the bias, allows generalizability and explainability due to distributed data and increase the performance of the algorithm.

Federated learning (FL) is a learning model that seeks to address data governance and privacy by training algorithms collaboratively without exchanging data. Originally developed for various fields, such as mobile and peripheral use cases, it has recently gained popularity for healthcare applications. FL allows information to be obtained in a collaborative way, for example in the form of a consensus model, without moving patient data outside the firewalls of the organizations in which they reside. Instead, the ML process takes place locally at each participating organization, and only model features (e.g., parameters, gradients) are transferred as shown in Figure 6. Recent research has shown that FL-trained models can achieve comparable performance levels to centralized host-trained models, datasets and outperform models that view only separate data from a single organization. While federated learning has some advantages, significant challenges such as security issues, regulatory compliance, model performance monitoring, etc. must be solved before federated learning can optimally generate acceptable AI models (Xu et al., 2021 and Rieke et al., 2020). In my opinion, some other technical challenges would be maintainability of distributed models, internet connectivity, possible bias in received response, infrastructure, and algorithm optimization.

The successful implementation of FL could therefore have significant potential to enable large-scale precision medicine, leading to unbiased decision-making models that optimally reflect an individual's physiology and susceptible to rare diseases. However, FL still requires rigorous technical review to ensure that the algorithm runs optimally without compromising patient safety or privacy. However, it has the potential to overcome the limitations of methods that require a single centralized data set (Rieke et al., 2020). The FL's promise is simple: address privacy and data governance challenges by enabling ML from non-localized data. In the FL environment, each data controller not only establishes their own governance processes and associated privacy policies, but also controls access to data and potentially data revocation. This includes both the training and validation phase. In this way, FL can create new opportunities, such as by enabling large-scale institutional validation or enabling new research in rare diseases, where incident rates are low and data sets are low. data of each organization is too small. Migrating from model to data rather than the other way around has another big advantage: large, archival medical data doesn't need to be copied from local institutions into a centralized group and copied to new data by each user using this data for the local model. Once the model is delivered to local organizations, it can scale naturally with a potentially growing global dataset without disproportionately increasing data storage requirements.

Health data is very sensitive and should be protected appropriately, following appropriate security procedures. Therefore, some of the key considerations are the trade-offs, strategies, and residual risks of the regarding FL's potential for privacy protection.

It is important to note that FL does not address all potential privacy issues, and like ML algorithms in general, there will always be some risk. Privacy protection techniques for FL provide levels of protection that exceed the ML models currently on the market. However, there are performance trade-offs, and these techniques can affect, for example, the accuracy of the final model. In addition, future ancillary techniques and/or data may be used to compromise a model previously considered low risk (Rieke et al., 2020).

In addition to the available literature, based on my verification of the US FDA official published database on the website fda.gov (Artificial Intelligence and Machine Learning (AI/ML) -Medical Devices Economic Support, 2021), no federated learning-based algorithms have been cleared.

In my opinion, it is safe to assume that no federated learning-based algorithms are currently in use at any institution in their clinical workflow.

Federated Machine Learning allows us to overcome the obstacles encountered by traditional machine learning models such as:

• Traditional machine learning occurs by moving all data sources to a centralized server to training and model building, but this may violate the rules of military organizations, especially when a third party is used to create, train and maintain the model.

• To train the model, a third party must prepare, clean, and restructure the data to be suitable for model training, however, this may violate the confidentiality of the data as the data data is processed to create the model.

• Traditional machine learning models also take a long time to build models with acceptable accuracy, leading to lag for organizations, especially startups.

• Traditional machine learning also requires a large amount of historical data to exist to train the model to provide acceptable accuracy

• A secure distributed machine learning method is needed to train the data data customers on their servers without a data security breach, saving compute power and fixing cold start, allowing customers to get immediate results (Abdul Salam, Taha & Ramadan, 2021).

Federated learning has the potential to solve these problems, as it allows dirty data servers to locally train their models and share their model gradients without violating privacy. of the patient. Thus, successful implementation of FL could have significant potential for large-scale precision drug activation, leading to unbiased decision-making models that optimally reflect an individual's physiology and sensitive to rare diseases while respecting governance and security issues. However, FL still requires rigorous technical review to ensure that the algorithm runs optimally without compromising patient safety or privacy. However, it has the potential to overcome the limitations of methods that require a single centralized data set. experimental studies performed only show the performance of linked learning models and currently, according to my literature search, no survey studies have been performed to understand the reliability and how strongly the user perceives affects the user's intention to accept.

The AI online marketplace provides a platform for buying and selling AI models between AI developers and customers (Kumar et al., 2020). The use of federated learning for machine learning to protect privacy among market players will be key to changing the business model in the AI market.

To assess the business value of investing in and implementing federated machine learning algorithms, it lacks the consideration of user-perceived reliability and robustness in federated machine learning, and Their intentions for user acceptance are based on my review of the material and opinions. In addition, given the novelty of federated learning in AI in medical imaging, it is, in my opinion, important to understand customer perceptions of the robustness and reliability of these association learning algorithms. FL includes a paradigm shift from centralized data lakes and it is important to understand its impact on various stakeholders in the FL ecosystem.

FL's promise is simple: address privacy and data governance challenges by enabling ML from non-localized data. In the FL environment, each data controller not only establishes their own governance processes and associated privacy policies, but also controls access to data and can revoke it. there. This includes both the training and validation phase. In this way, FL can create new opportunities, such as by enabling large-scale institutional validation or enabling new research in rare diseases, where incident rates are low and data sets are low. data of each organization is too small. Migrating from model to data rather than the other way around has another big advantage: large, archival medical data doesn't need to be copied from local institutions into a centralized group and copied to new data by each user using this data for the local model. train. Once the model is delivered to local organizations, it can scale naturally with a potentially growing global data set without disproportionately increasing data storage requirements (Rieke et al., 2020).

CHAPTER III:

METHODOLOGY

3.1 Overview of the Research Problem

This research project will evaluate the impact assessment of federated learning setting on user's perception of trust and robustness and likelihood of user adoption intent of AI-based tools in radiology medical imaging and test hypothesis whether user perception of trustworthiness and robustness of a machine learning algorithm in medical imaging has positive co-relation with federated machine learning approach. The outcome of this hypothesis testing could validate the theoretical understanding and brings value proposition in research and development in federated learning algorithms which is critical for Vendors and conceptualize the future business models of AI marketplace.

Similar research was conducted by McNair (2021), on Influencer marketing by conducting a quantitative study that researches the impact of two influencer attributes, trust, and content quality analyzing the co-relation with travel intentions of a follower. Due to the similarity of the research question, the methodology and data analysis procedures has been adapted.

3.2 Operationalization of Theoretical Constructs

Based on the hypothesis, there are three theoretical constructs - Perceived Trustworthiness, Perceived Robustness, and Intention to use AI-based tools using federated learning approach was described. Each of the measurement constructs was divided into 5 items. Perceived Trustworthiness as PT1, PT2, PT3, PT4 and PT5. Perceived Robustness as PR1, PR2, PR3, PR4 and PR5 and Intention to use AI-based tools using federated learning approach as INT1, INT2, INT3, INT4 and INT5. The measurement constructs for "Intention to use AI-based tools using federated learning approach" was directly adopted from survey study (Esmaeilzadeh, 2020).

Construct	Item	Wording

Perceived Trustworthiness	PT1	It can preserve privacy
	PT2	It can produce Results that are Unbiased
	PT3	It can produce Explainable machine learning models
	PT4	It can adapt to specific needs and stakeholder requirements
	PT5	It inspires trust in machine learning models/ AI-based tools in radiology
Perceived Robustness	PR1	It can produce Accurate results
	PR2	The algorithm performance is better over traditional machine learning
	PR3	It produces results that are Generalizable for the intended population
	PR4	Algorithm will always stay most updated
	PR5	It is reliable over traditional machine learning models
Intention to use AI-based tools using federated learning approach	INT1	I agree to use AI-based tools for clinical purposes
	INT2	Using AI-based tools for healthcare purposes is something I would consider
	INT3	I would like to use AI-based devices to manage my healthcare
	INT4	In the future, I am willing to use AI-based services for diagnostics and treatments
	INT5	I am very likely to use recommendations provided by AI-based tools for care planning

Table 1 Measurement Constructs (Esmaeilzadeh, 2020)

3.3 Research Design

The study analyzes the correlation between a participant’s perception of federated learning’s trustworthiness and robustness over traditional machine learning approach and ascertains whether this impacts the participant’s adoption intentions of AI-based tools in radiology medical imaging. Understanding these two aspects of the federated learning and user adoption relationship will allow businesses to invest in R&D and product development using federated learning approach for their projects and allow developers to facilitate the growth of suitable attributes for marketing success.

The quantitative study uses perceived trustworthiness and robustness as the two independent variables and user adoption intent of AI-based tool as the dependent variable.

To collect the data, online survey was planned as data collection instrument.
The online survey consists of research questionnaire on Likert scale. The main survey link was created and then shared with the participants. After the data collection via the survey, IBM SPSS statistical software was used for analysis.

To analyze the Likert scale data, parametric tests (Pearson's correlation and linear regression) was used to accept or reject the hypotheses. A multiple linear regression analysis can be conducted since there are two independent variables (trustworthiness of federated learning and robustness of federated learning) and one dependent variable (user adoption intent).

3.4 Expert Review of Survey Questionnaire

The review of survey questionnaire by experts was considered necessary because it may help to improve the quality and efficiency of the study. The review survey questionnaire was conducted by personally selected PhD holders (2) and subsection of the population (4 consumers). The potential limitation was the personal selection, but the author selected few people who can invest their time and provide quality feedback, giving the survey the best chance of providing significant insights.

For review, the Survey questionnaire was shared with participants. Based on the verbal feedback, definitions of terms were included for better clarity and additional question to distinguish the participant as users in radiology and non-radiology. This was considered since study cohort may contain both users in radiology and non-radiology. One of the reviewers suggested to assess the participant knowledge in Federated Learning on Likert scale instead of nominal "Yes/ No". Instead of changing the original question to likert scale, an additional question was included to first understand whether the participant is aware of Federated learning using nominal "Yes/ No" and in case answer is "Yes", then participants were asked to self-assess their level of knowledge on the 5-point scale, with 1 being "Very low" and 5 being "Very high". In addition to feedback on questionnaire, it was also suggested to conduct a sub-group of analysis of association of type of users with their perception of adoption intent as it can provide valuable information from business standpoint.

3.5 Population and Sample

The study population consists of community formed by enthusiasts of AI in healthcare Whatsapp messenger group. There were 250 members in the group consisting of Radiologists, Clinician, Academicians, Developers, Manufacturers, Enthusiasts etc. It was a convenience sample. Therefore, sample size of 250 was considered for this survey. Evans and Mathur (2018) found that response rate for most email surveys was closer to 11%. An expected response rate of around 15% was targeted. So, the expected response was considered as 40.

3.6 Participant Selection

Participants will be users (both prospective and existing) of AI-based tools in healthcare focusing on radiology. In general radiologists, Clinicians, Developers, Vendors, Technical Experts, Academicians and AI in Healthcare enthusiasts from Medical Imaging communities can be considered as study cohort.

3.7 Instrumentation

Online survey was chosen as data collection instrument. Online surveys are useful in conducting the research. It is found convenient way for questionnaire preparation, data collection, storing of data, visualization of data and for collaboration of work. Online surveys can be conducted at low cost, provide real-time access from multiple devices and in a short period of time. Some of the challenges related to online surveys were sampling, response rate, non-respondent characteristics, limited sampling and respondent availability, maintenance of confidentiality (Howard, 2019).

Online survey research is used more frequently and better accepted by researchers. Online surveys have become a preferred method for many researchers due to the rise of access to broadband internet, mobile phones, and social networks over the last 15 years. An online survey was used, since this is widely seen as the best way to survey a global population in a cost-effective way and most preferred by researchers in recent decade (Evans and Mathur, 2018).

3.8 Data Collection Procedures

To collect the data, online survey was used as data collection instrument.
After receiving the approval of Research proposal including survey questionnaire by the Internal Review Board at SSBM, the survey link was sent to the participants.

To collect the data for this study, an online survey will be created through Google Forms© and sent to participants. Google Forms© maintains security, privacy and regulatory compliance as stated on the website (Google, 2019).

Participants were given 3 weeks to answer the survey questions before the survey was closed. A follow-up reminder will be sent out 2 days before due date. The follow-up reminders are proven to increase the survey response rate (Evans and Mathur, 2018).

An online survey using Google Forms© was designed by formulating 12 questions. The link to the survey was distributed among radiologists, Clinicians and AI enthusiasts. The survey remained open for 3 weeks between 26 January and 22 February 2022 (Malwade, 2022).

3.9 Data Analysis

After the data collection via the survey, responses were stored in a spreadsheet (Google Forms©) that was later transformed into Microsoft Excel(.xlsx). IBM SPSS statistical software (Version 28) was used for analysis. To analyze the Likert scale data, parametric tests (Pearson's correlation and linear regression) was used to accept or reject the hypotheses. Sullivan and Artino (2013) stated that parametric tests are sufficiently robust to yield largely unbiased answers that are acceptably close to the truth when analyzing Likert scale responses.

For the analysis, Pearson's correlation will be used to test hypotheses 1 and 2. The correlation studied both trustworthiness and user adoption intent for hypothesis 1 and robustness and user adoption intent for hypothesis 2. Pearson's correlation was selected for this analysis because it analyzed two continuous variables and measured the strength of the association.

A multiple linear regression analysis was conducted since there are two independent variables (trustworthiness of federated learning and robustness of federated learning) and one dependent variable (user adoption intent). Since there are two independent variables (trustworthiness and robustness) and one dependent variable (user adoption intent), a multiple linear regression analysis was conducted. Hyman and Sierra (2016) states that multiple linear regression should be used when "identifying marketing strategies that influence purchases by a target customer." In this study, the "marketing strategies" are developing trustworthy and robust AI-based tools, and the "purchase" is user adoption intent. This test was used to answer hypothesis 3.

All inferential statistical research used 95% confidence interval. As a result, statistical significance was established at a level of 5% (= 0.05). Prior to doing the regression analysis, descriptive data for the various variables was provided, including mean, median, standard deviation, and kurtosis. These demonstrate the normality or distribution of the variables.

After that, a correlation analysis for the data was constructed and shown. The first stage in regression analysis was to establish correlation between the independent and dependent variables; only then can the precise influence of the independent factors on the dependent variables be determined. Additionally, the correlation matrix will demonstrate the link between the dependent variables. Correlation between independent variables increases the probability of multicollinearity, which may affect regression results.

According to Andersson et al. (2014), regression analysis may reveal just relationships between variables, not the underlying causal mechanism. Nonetheless, regression is an effective technique for predicting and estimate in the future. All variables were quantified using descriptive analysis, correlation and regression analysis were used to evaluate quantitative data, whereas content analysis was used to evaluate qualitative data.

After that, a correlation analysis for the data was constructed and shown. The first stage in regression analysis was to establish correlation between the independent and dependent variables; only then can the precise influence of the independent factors on the

dependent variables be determined. Additionally, the correlation matrix will demonstrate the link between the dependent variables. Correlation between independent variables increases the probability of multicollinearity, which may affect regression results.

A multiple linear regression analysis was conducted since there are two independent variables (trustworthiness of federated learning and robustness of federated learning) and one dependent variable (user adoption intent). Since there are two independent variables (trustworthiness and robustness) and one dependent variable (user adoption intent), a multiple linear regression analysis was conducted. Hyman and Sierra (2016) states that multiple linear regression should be used when "identifying marketing strategies that influence purchases by a target customer." In this study, the "marketing strategies" are developing trustworthy and robust AI-based tools, and the "purchase" is user adoption intent. This test was used to answer hypothesis 3.

3.10 Research Design Limitations

For this study, the sample consisted of community formed by enthusiasts of AI in healthcare Whatsapp messenger group. The community included Radiologists, clinician, Academicians, developers, manufacturers, enthusiasts etc. It was a convenience sample, having excluded a large population that lacked the opportunity to participate. According to participant exclusion criteria, participants are included if they are consumers, developers, AI enthusiasts etc. and these inclusions do not consider age, sex, socioeconomic status, or race. The primary disqualifying exclusion is that if a user did not read or skipped the Survey link or missed responses they were not considered for the study. This can impact the replication of this study since no one except the author has access to this particular sample, and future studies would be unable to test the same population.

The Likert scale does not allow the population to input their own answers and may be seen as “too closed”. This survey method could reduce the probability of participants completing the survey, especially if they had something they wanted to convey that was not covered by the available answers.

A quantitative study with a correlation-based analysis was chosen. There are inherent limitations with correlation studies, since a positive or negative correlation does not mean that the variables have a direct cause-and-effect relationship. The positive or negative correlation simply means that there is an association present when studying the two variables together.

The survey also deals with the idea of perception, since it asks participants to judge federated learning capabilities on deep topics such as trustworthiness and robustness. One participant's perception is potentially different from another's, but that does not mean that the federated learning approach is trustworthy, just that the individual perceives them as trustworthy.

The survey (Appendix 1) was used to capture the data for the study. The tool had multiple sections and the scale for each section will have 5-point Likert scale. Another potential limitation for the survey could be that it dealt with “perceived trustworthiness” and “perceived robustness,” since it was simply asking participants how they perceived these attributes in the federated learning approach. The research question was based on how the respondents perceived the federated learning's trustworthiness and robustness. This perception is not necessarily an accurate analysis of what something like "high robustness" means in the academic sense, but simply what it is perceived as by the respondent.

3.11 Conclusion

The study analyzes the correlation between a participant's perception of federated learning's trustworthiness and robustness over traditional machine learning approach and ascertains whether this impacts the participant's adoption intentions of AI-based tools in radiology medical imaging.

The quantitative study uses perceived trustworthiness and robustness as the two independent variables and user adoption intent of AI-based tool as the dependent variable. A Pearson correlation was calculated on each independent variable versus the dependent variable and a linear regression to test the correlation between both independent variables and the dependent variable.

The review of survey questionnaire by experts was considered necessary because it may help to improve the quality and efficiency of the study.

The survey questionnaire was prepared, and measurement contstucts were prepared. Few were adopted from research study (Esmaeilzadeh, 2020). The questionnaire was prepared using Likert scale to evaluate the user perception trustworthiness, robustness, and user adoption intent. The questionnaire was designed to collect respondents' attributes like experience, roles, knowledge of federated learning and whether they are already a customer of AI-based tools or intent/participating in research. These details could provide key insights into current trends in use of AI-based tools and established co-relations with adoption intent.

Few of limitations of this study design were choice of samples, use of Likert scale which was considered as closed.

CHAPTER IV:

RESULTS

4.1 Introduction

This chapter summarizes the findings of research. To begin, it gives information on the sample size, response rate, dependability, and validity of the study methods. Later, it discusses the respondents' backgrounds and conducts a descriptive analysis of the research variables. Finally, it explains the results of statistical analysis used to test the hypotheses and discusses the findings and conclusions drawn from them.

4.2 Sample

The survey was sent to 256 recipients. 53 responded, giving the survey an 20.71% response rate, with 51 fully completing the surveys. One of the respondents did not mention the country and another did not mention the years of experience. To maintain the integrity of the data, the 2 responses that were not fully completed were removed. These responses were a small percentage of the collected data (less than 10%), and Pigott (2001) stated that complete-case analysis can still represent the sample reliably when only a small percentage of surveys have not been completed.

4.3 Type of Respondents

Most of the respondents were Radiologists (45.10%), followed by Clinician (23.53%), AI enthusiasts (11.76%), Clinician/Researcher and Data Scientists (Figure 7).

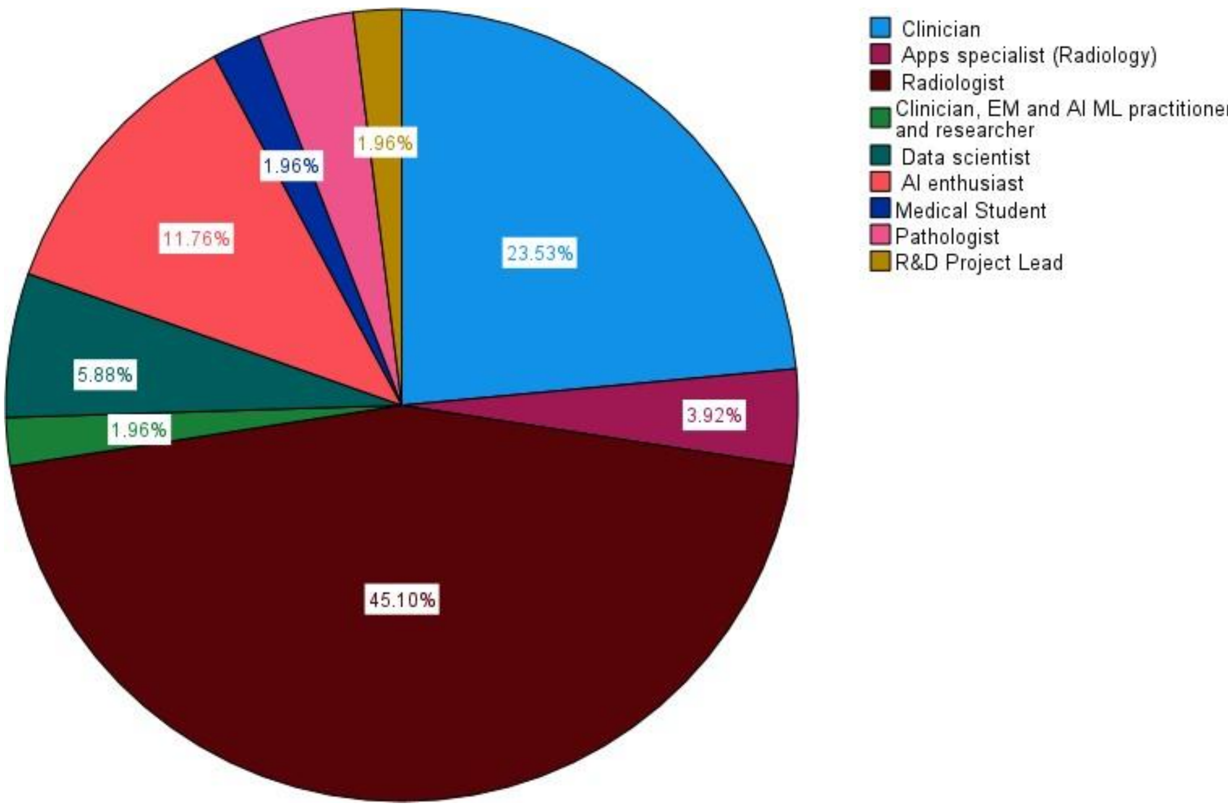

Figure 7 Distribution of Respondents

4.4 Experience of Respondents

The participants in this survey were predominantly in the range of 10-15 years of experience (39.22%). The years of experience had majority in the more than 10-15 years range (40.4%), followed by more than 15 years (21.2%) and 5-10 years (21.2%) and 0-5 years (15.4%) (Figure 8).

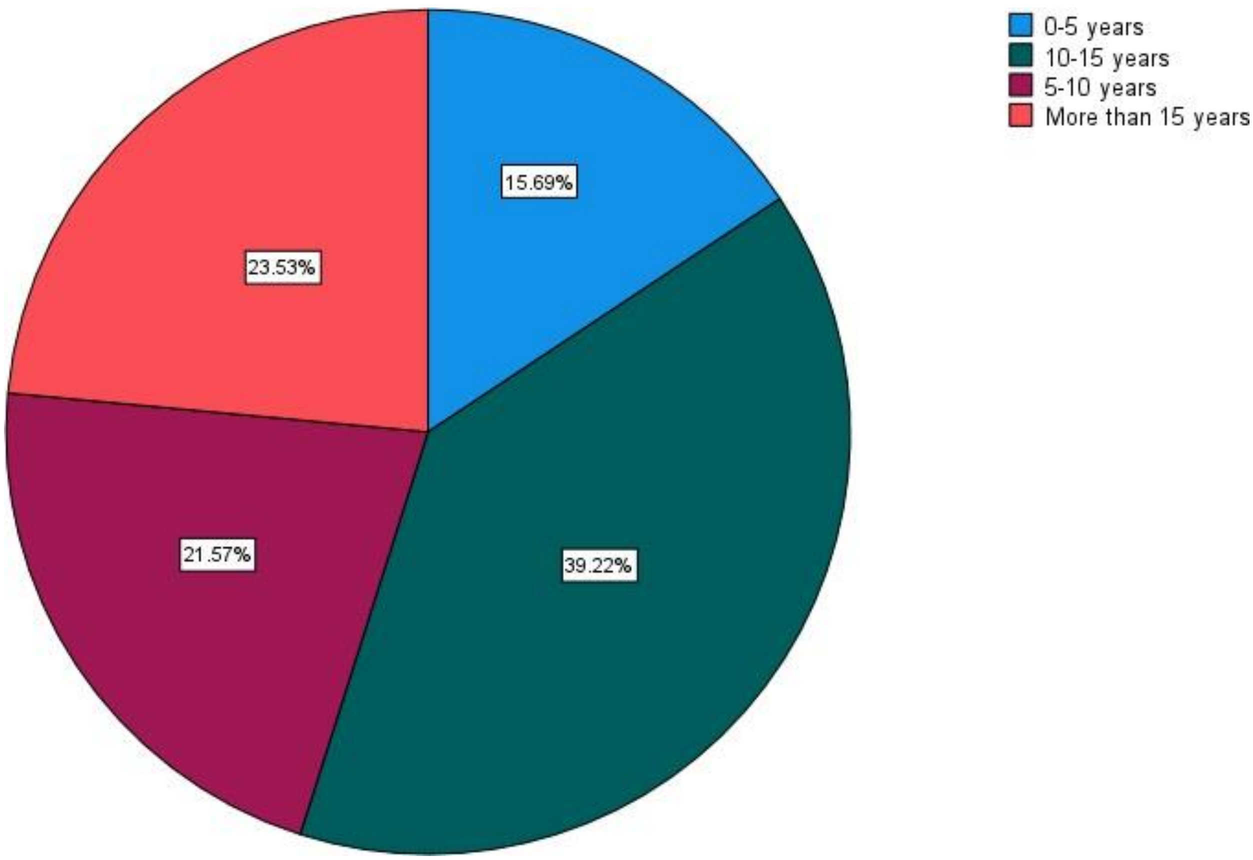

Figure 8 Experience of Respondents in relevant Field, Role, or Area

4.5 Respondents Location data

Most of the sample (35.29%) lived in India, followed by USA, (25.49%) and Germany (11.76%) the only other countries with a higher than 1% representation in the sample (Figure 9).

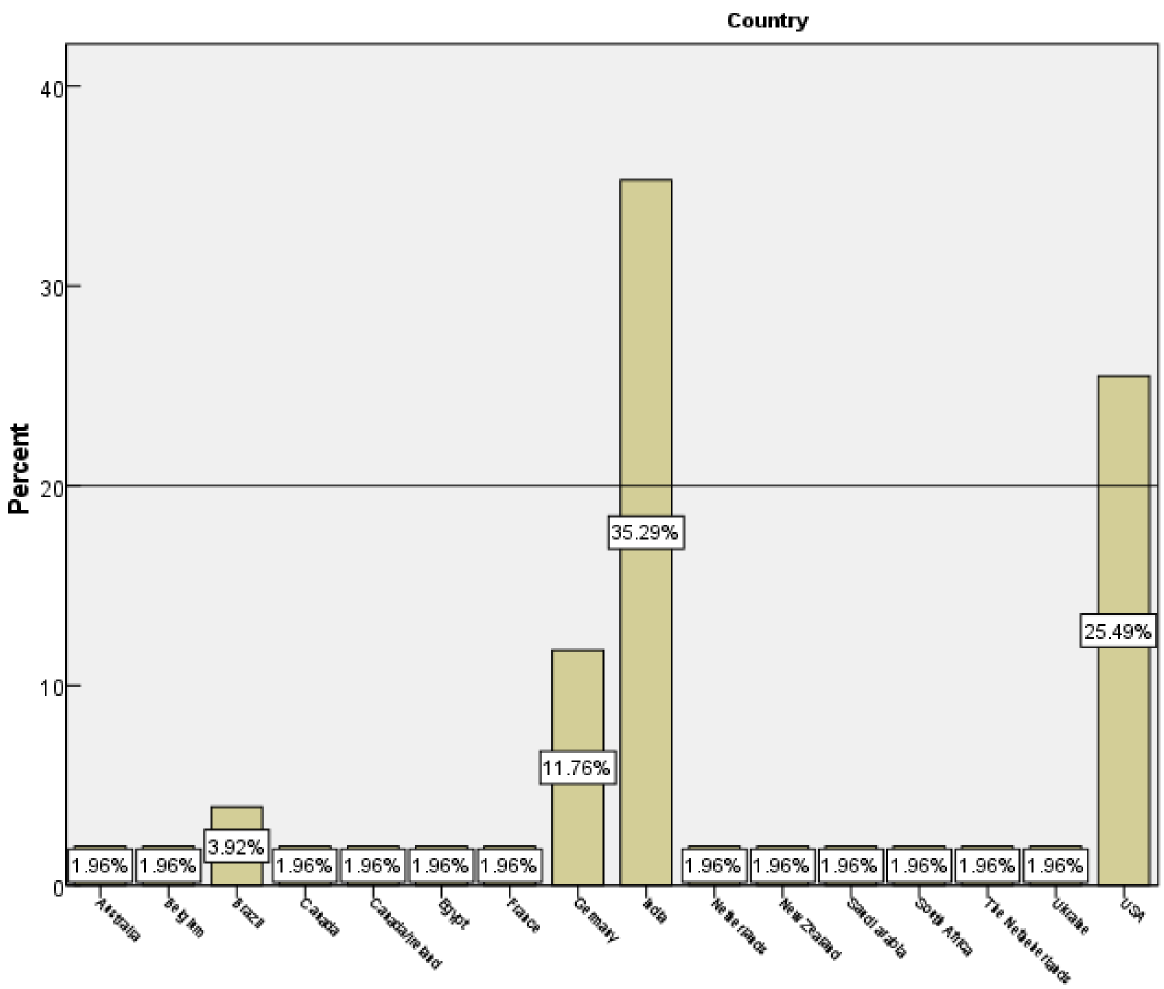

Figure 9 Respondents Location Data

Since knowledge of federated learning and experience in AI-based tools will impact the respondents' thoughts a question was asked on federated Learning whether they have knowledge of federated Learning.

4.6 Respondents Knowledge of Federated Learning

Of 51 respondents, 44(86.3%) had knowledge of federated learning and 7(13.7%) had no knowledge of federated Learning (Figure 5).

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	7	13.7	13.7	13.7
	Yes	44	86.3	86.3	100.0
	Total	51	100.0	100.0	

Table 2 Respondents Knowledge of Federated Learning

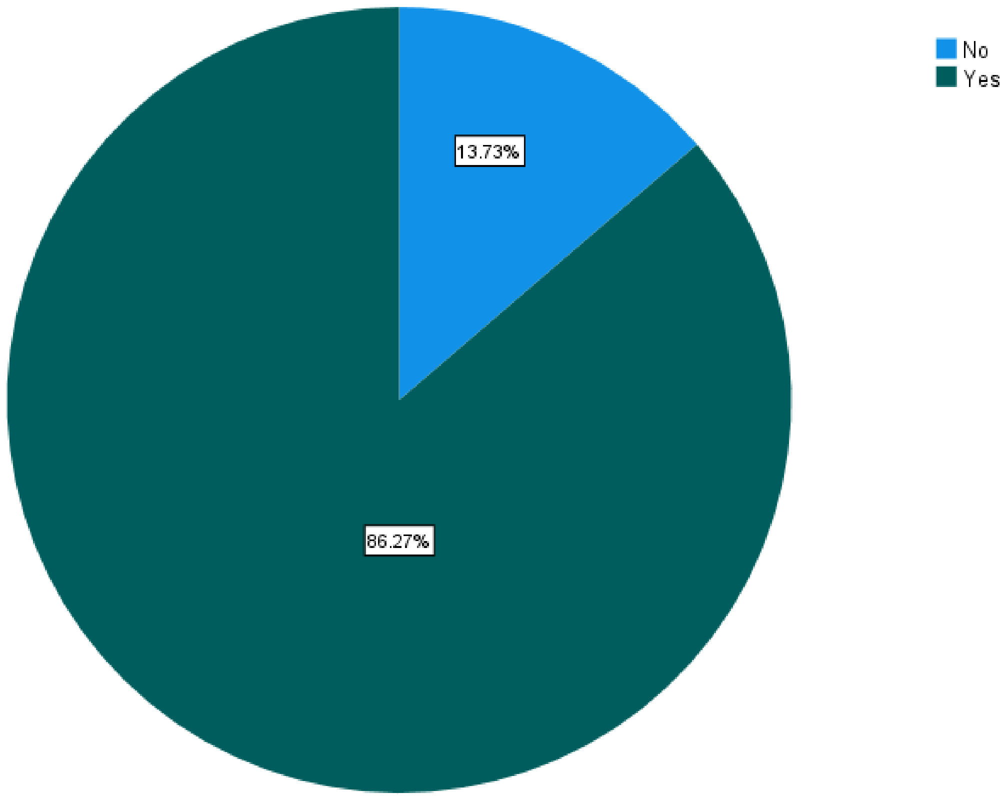

Figure 10 Respondents Knowledge of Federated Learning

4.7 Respondents Level of Knowledge of Federated Learning

Out of the 44 respondents who had knowledge of federated learning, 23 (52.27%) had moderate level of knowledge, 8 respondents had high level of knowledge, 1 respondent had very high and 5 (9.8%) had very low (Figure 6).

In case respondent chose Yes then level of perceived knowledge in Likert scale was

(1) Very low

(2) low

(3) moderate

(4) High

(5) Very High

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	5	9.8	11.4	11.4
	2	7	13.7	15.9	27.3
	3	23	45.1	52.3	79.5
	4	8	15.7	18.2	97.7
	5	1	2.0	2.3	100.0
	Total	44	86.3	100.0	
Missing	System	7	13.7		
Total		51	100.0		

Table 3 Respondents Level of Knowledge of Federated Learning

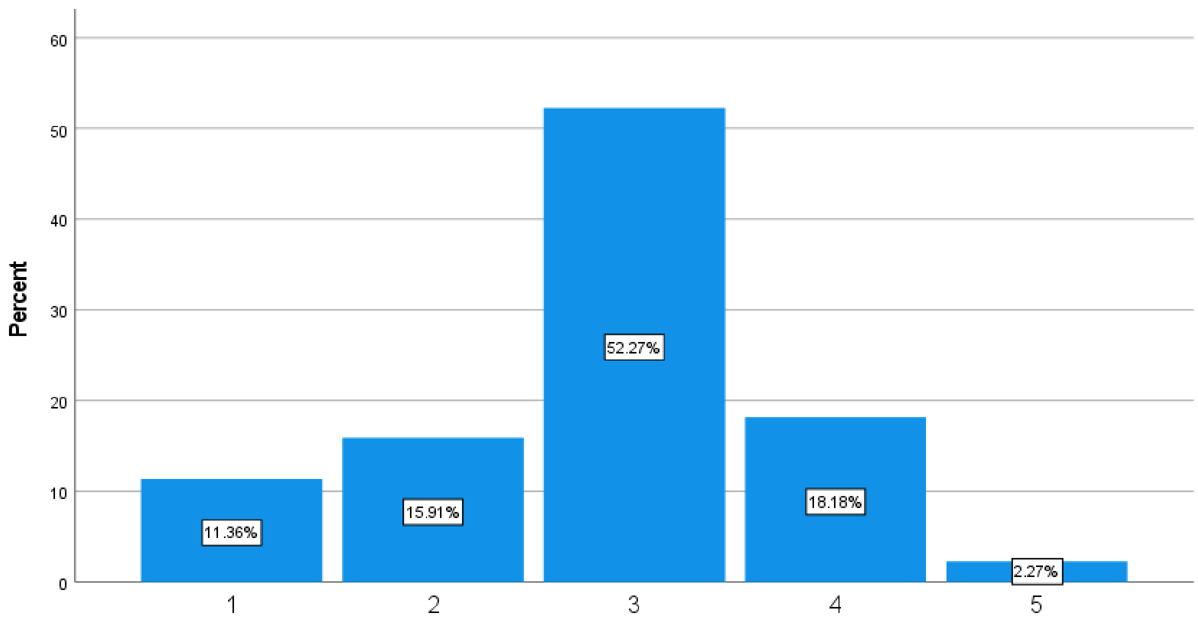

Figure 11 Respondents Level of Knowledge of Federated Learning

4.8 Respondents Currently Users of AI-Based Tool for Radiology Clinical Workflows

Out of 51 respondents 21 (41.2%) were currently a user of AI-based tools for radiology workflows (Figure 12).

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	30	58.8	58.8	58.8
	Yes	21	41.2	41.2	100.0
	Total	51	100.0	100.0	

Table 4 Respondents Currently Users of AI-based tool for Radiology Clinical Workflows

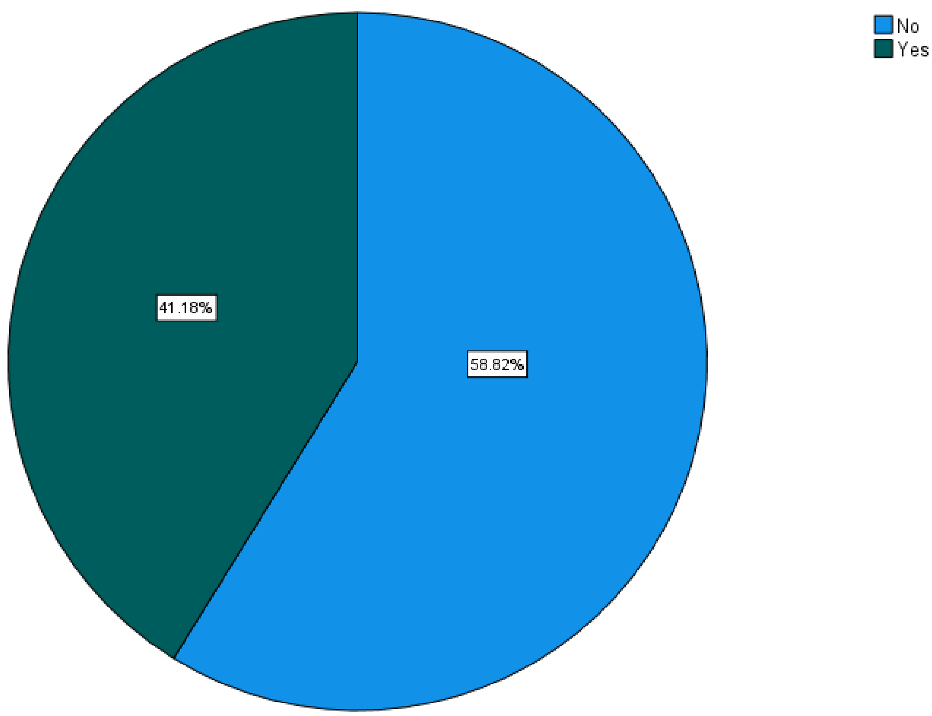

Figure 12 Respondents Currently Users of AI-based tool for Radiology Clinical Workflows

4.9 Respondents Currently Users of AI-Based Tool for other than Radiology Clinical Workflows (For example: Ophthalmology)

Out of 51 respondents 15 (29.4%) were currently a user of AI-based tools for other than radiology workflows (
Figure 13).

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	36	70.6	70.6	70.6
	Yes	15	29.4	29.4	100.0
	Total	51	100.0	100.0	

Table 5 Respondents Currently Users of AI-based tool for other than Radiology Clinical Workflows

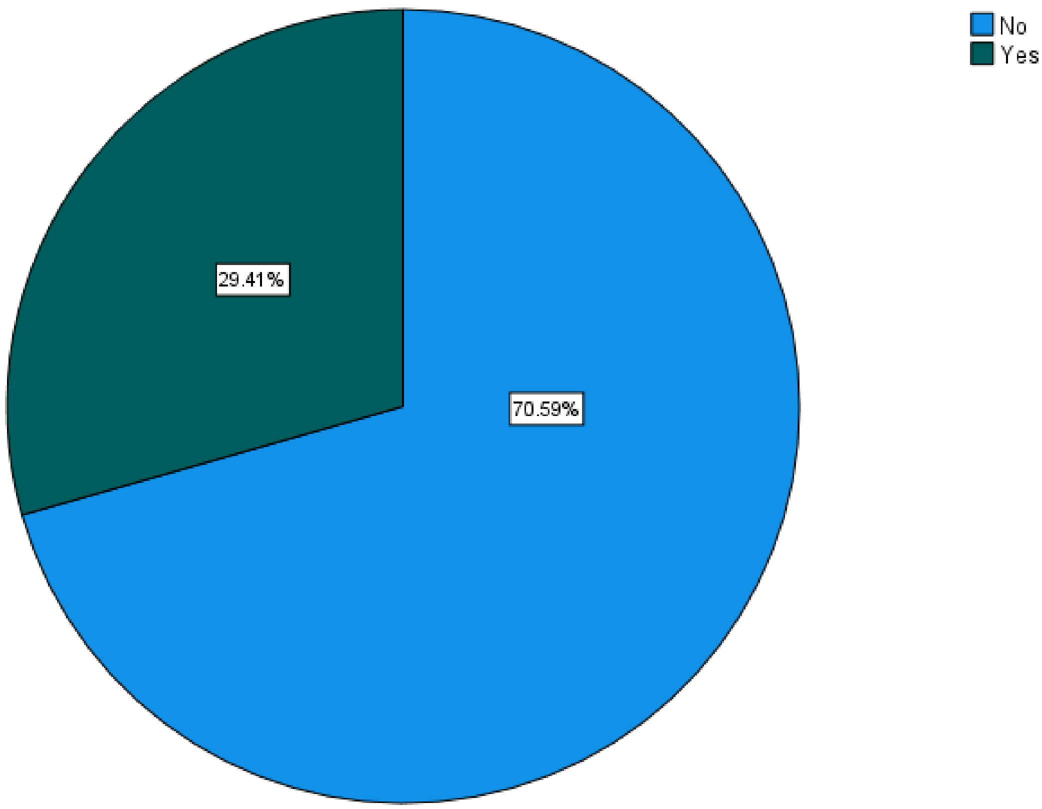

Figure 13 Respondents Currently Users of AI-based tool for other than Radiology Clinical Workflows

4.10 Respondents who Participated/Contributed to Research or Experiments related to Federated Machine Learning or Currently Intent to do.

Out of 51 respondents 22(43.1%) either participated or contributed to research or experiments related to federated learning or they intent to do as (Figure 9).

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	29	56.9	56.9	56.9
	Yes	22	43.1	43.1	100.0
	Total	51	100.0	100.0	

Table 6 Respondents who Participated/Contributed to Research or Experiments related to Federated Machine Learning or currently intent to do

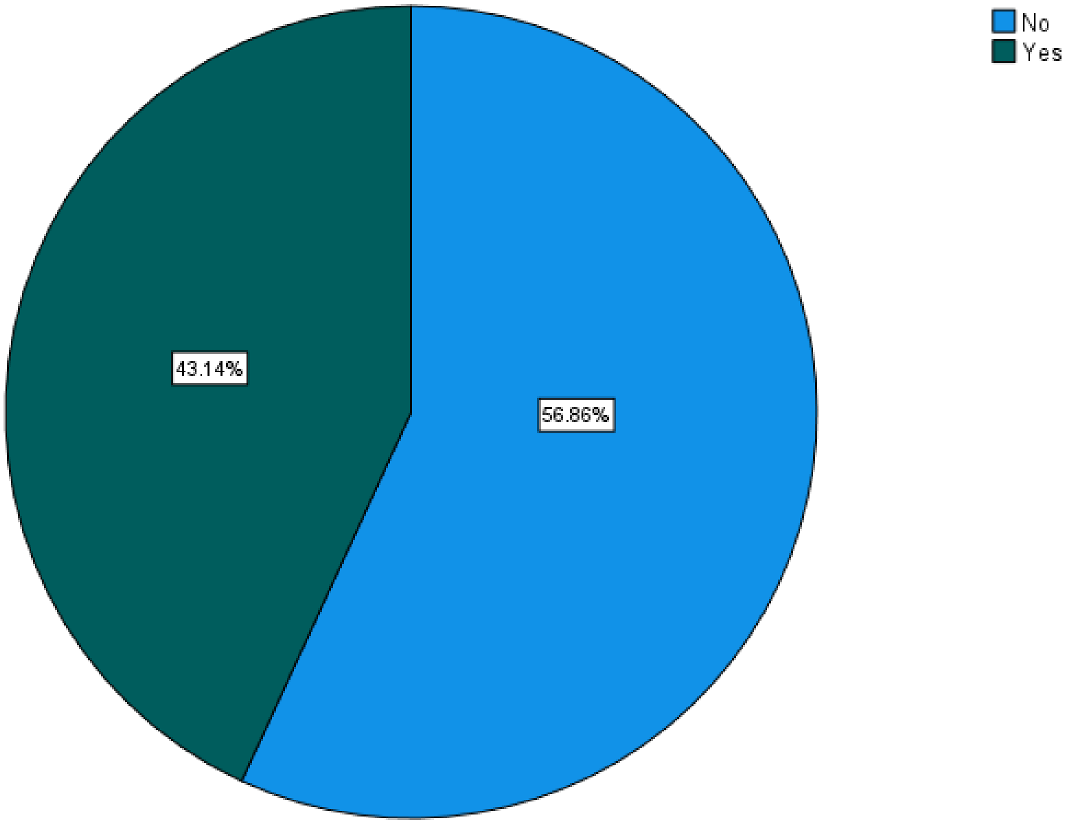

Figure 14 Respondents who Participated/Contributed to Research or Experiments related to Federated Machine Learning or Currently Intent to do

4.11 Respondents Perceived Trustworthiness vs Adoption intent

H1: User perception of federated learning's trustworthiness has a positive correlation with user adoption intent of AI-based tools in medical imaging in radiology.

H1 studied the correlation between how trustworthy an algorithm is and whether the perceived trust impacts user adoption intentions. For the analysis, a Pearson's correlation was run on the average of the responses from the five questions in the trustworthiness and adoption intent section, with 1 being "strongly disagree" and 5 being "strongly agree. The mean from the responses in the trustworthy section was 3.77 with a standard deviation of .67. The mean from the responses in the adoption intentions section was 4.01 with a standard deviation of .72.

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
PT	51	1.00	5.00	3.7725	.67174
INT	51	1.00	5.00	4.0157	.72756
Valid N (listwise)	51				

Table 7 Respondents Perceived Trustworthiness vs Adoption intent - Descriptive Statistics

To examine H1 ("User perception of federated learning's trustworthiness has a positive correlation with user adoption intent "), a Pearson's correlation was run on these two variables. The analysis shows a statistically significant ($r=.549$, $p<.001$) positive correlation between the two variables as per Table 8. The R-value is .549, so the effect size is medium since it is greater than 0.5 based on common effect size indices (Sullivan and Feinn, 2012).

Therefore, we can accept the hypothesis that states there is a positive relationship between perceived trustworthiness and users' adoption intent. Pearson product correlation of Perceived Trustworthiness vs Adoption intent was found to be moderately positive and statistically significant ($r=.549$, $p<0.01$) (Table 8). Hence H1 is supported, This shows that an increase in perceived trustworthiness would lead to higher adoption intent of users.

Correlations

		PT	INT
PT	Pearson Correlation	1	.549**
	Sig. (2-tailed)		<.001
	N	51	51
INT	Pearson Correlation	.549**	1
	Sig. (2-tailed)	<.001	
	N	51	51

**. Correlation is significant at the 0.01 level (2-tailed).

Table 8 Respondents Perceived Trustworthiness vs Adoption intent – Correlations

4.12 Respondents Perceived Robustness vs Adoption intent

H2: User perception of federated learning's robustness has a positive correlation with user adoption intent of AI-based tools in medical imaging in radiology

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
PR	51	2.60	5.00	3.7412	.52352
INT	51	1.00	5.00	4.0157	.72756
Valid N (listwise)	51				

Table 9 Respondents Perceived Robustness vs Adoption intent - Descriptive Statistics

H2 studied the connection between the user perceived robustness of federated learning algorithm and how it correlated with the user adoption intentions. To analyze the relationship, the average response from the five questions in the Perceived Robustness section of the survey was correlated against the average response from the five questions in the adoption intent section. The average mean from the survey responses in the perceived trustworthiness section was 3.74, with a standard deviation of .523. The average mean from the adoption intention section was the same as in the previous analysis at 4.01 with a standard deviation of .727 (Table 9).

To examine H2 (“User perception of federated learning’s robustness has a positive correlation with user adoption intent of AI-based tools in medical imaging in radiology”), Pearson product correlation of Perceived Robustness vs Adoption intent was found to positive with small effect size and statistically significant (r=0.303, p=0.05) (Table 10). Hence H2 is supported, This shows that an increase in perceived robustness would lead to higher adoption intent of users.

Correlations

		PR	INT
PR	Pearson Correlation	1	.303*
	Sig. (2-tailed)		.031
	N	51	51
INT	Pearson Correlation	.303*	1
	Sig. (2-tailed)	.031	
	N	51	51

*. Correlation is significant at the 0.05 level (2-tailed).

Table 10 Respondents Perceived Trustworthiness and Robustness vs Adoption intent

4.12 Respondents Perceived Trustworthiness and Robustness vs Adoption intent

H3: User perception of federated learning’s trustworthiness and robustness have a positive correlation with user adoption intent of AI-based tools in medical imaging in radiology.

H3 studied the impact of both perceived trustworthiness and robustness on user adoption intent. For this analysis, multiple linear regression was used. Trustworthiness and Robustness were the independent variables, and adoption intent was the dependent variable.

Model Summary[b]									
					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.549[a]	.302	.273	.62046	.302	10.376	2	48	<.001

a. Predictors: (Constant), PR, PT
b. Dependent Variable: INT

Table 11 Respondents Perceived Trustworthiness and Robustness vs Adoption intent - Model Summary

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	7.989	2	3.994	10.376	<.001[b]
	Residual	18.479	48	.385		
	Total	26.467	50			

a. Dependent Variable: INT
b. Predictors: (Constant), PR, PT

Table 12 Respondents Perceived Trustworthiness and Robustness vs Adoption intent – ANOVA

Coefficients[a]					
Model	Unstandardized Coefficients	Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B

		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	1.727	.660		2.615	.012	.399	3.055
	PT	.586	.154	.541	3.801	<.001	.276	.897
	PR	.020	.198	.015	.103	.918	-.378	.418

a. Dependent Variable: INT

Table 13 Respondents Perceived Robustness and Robustness vs Adoption intent – Coefficients

In the analysis, both perceived trustworthiness and robustness had a statistically significant impact on the user adoption intent. The results from indicate, that the average of the trustworthiness variable had a positive unstandardized beta of .586 and the average of Robustness had a positive unstandardized beta of 0.20. The model also has an adjusted r-squared of .273, which indicates a moderately positive relationship (Ratner, 2009).

The results from linear regression could find statistically significant influence perceived trustworthiness and robustness had positive increase in the adoption intentions of users. We can accept the hypothesis that there is a positive relationship between both trustworthiness and robustness and users' adoption intent.

4.13 Other Findings

After comparing the means of Perceived Trustworthiness, Perceived Robustness and Adoption intent, three additional factors were discovered, and these are presented below.

4.13.1 Respondents Difference between Experience Groups

The years of experience had majority in the more than 10-15 years range (40.4%), followed by more than 15 years (21.2%) and 5-10 years (21.2%) and 0-5 years (15.4%).

An analysis of the experience ranges showed a potential difference in perceived adoption intentions in respondents in range of 10-15 years. The response rate was significant in the range of 10-15 years (40.4%). The experience ranges from 0-5 years showed highest adoption intent but the respondents were lowest (15.4%) (Table 14). Future research could examine how the 0-5 years and 10-15 years' experience group's perception on adoption intent differ from other experience groups.

Respondents Difference between experience groups			
Mean			
Your experience in relevant field, role, or area.	PT	PR	INT
a. 0-5 years	3.3250	3.5500	4.2500
b. 5-10 years	3.8000	3.8000	4.0727
c. 10-15 years	3.8800	3.9200	3.8500
d. More than 15 years	3.8667	3.5167	4.0833
Total	3.7725	3.7412	4.0157

Table 14 Respondents Difference between experience groups – Mean

4.13.2 Respondents Level of Knowledge in Federated Learning

Of 51 respondents, 44 (86.3%) had knowledge of federated learning and 7 (13.7%) had no knowledge of federated learning.

Out of the 44 respondents who had knowledge of federated learning, 23 (52.27%) had moderate level of knowledge, 8 respondents had high level of knowledge, one respondent had very high and 5 (9.8%) had very low.

In case respondent chose Yes then level of perceived knowledge in Likert scale was

(1) Very low

(2) low

(3) moderate

(4) High

(5) Very High

Respondents Level of knowledge in Federated Learning

Mean

If answer is "Yes" for question 4, according to you what is your level of knowledge of Federated Learning? In case answer is "No" to question 4, please skip this question.	PT	PR	INT
1	3.4800	3.4400	3.6400
2	3.2857	3.5714	3.1429
3	3.9667	3.8083	4.2083
4	3.7500	3.8250	4.2250
5	3.2000	3.0000	3.8000
Total	3.7511	3.7156	3.9733

Table 15 Respondents Level of knowledge in Federated Learning – Mean

Another insightful observation was the correlation of level of knowledge in federated learning with the perceived adoption intent. The respondents with moderate and high level of knowledge had same mean for adoption intent (4.2). While respondents with very high knowledge also had significant mean for adoption intent (3.8) (Table 15). These could be stronger connection with the level of knowledge of federated learning. Further studies can be conducted in this area to understand whether increase in knowledge had positive or negative influence in perception of adoption intent.

4.13.3 Type of Role among Respondents

Most of the respondents were Radiologists (45.10%), followed by Clinician (23.53%), AI enthusiasts (11.76%), Clinician/Researcher and Data Scientists.

Another noteworthy finding that could encourage further research is related to the type of respondents and how that impacts the answers to question on trustworthiness, robustness, and adoption intent. There appears to be potential difference in type of role. The respondents who were radiologists have relatively lower level of perception when compared with other respondents. The other respondent like project leads and researcher had very high perception. This may also be due to the bias. The clinicians and AI enthusiasts had almost similar levels of perception.

	Type of role
Mean	

Which one among below roles describes you the best?	PT	PR	INT
Clinician	3.7500	3.8000	4.0667
Apps specialist (Radiology)	4.3000	4.5000	4.5000
Radiologist	3.6609	3.6783	3.8522
Clinician, EM and Ai ML practitioner and researcher	5.0000	3.0000	5.0000
Data scientist	3.8667	3.8000	4.3333
AI enthusiast	3.8333	3.8000	3.9667
Medical Student	3.4000	3.8000	3.4000
Pathologist	4.0000	3.5000	4.1000
R&D Project Lead	3.6000	3.6000	5.0000
Total	3.7725	3.7412	4.0157

Table 16 Type of role – Mean

4.14 Summary of Findings

The descriptive statistics, mean and standard deviation was calculated for two dependent variables – Trustworthiness and Robustness and independent variable – Adoption intent. Pearson correlation analysis was performed, and r values showed statistically significant with moderately positive correlation. Hence Hypothesis 1 and 2 was positively supported. A linear regression analysis was performed to understand the correlation of combination of Trustworthiness and Robustness with Adoption intent. Unstandardized Coefficients B value was positive for Perceived trustworthiness and perceived robustness. Therefore, it was concluded that Hypothesis 3 was also positively supported.

Also, it can be concluded that difference in experience, levels of knowledge in federated learning and type of role of respondents had difference in answering the questionnaire and their perception.

4.15 Conclusion

The results conclude that:

Hypotheses:

H1: The independent factor User perception of federated learning's trustworthiness are statistically significant and influence positive correlation with dependent factor user adoption intent of AI-based tools in medical imaging in radiology.

H2: The independent factor User perception of federated learning's robustness are statistically significant and influence positive correlation with dependent factor user adoption intent of AI-based tools in medical imaging in radiology.

H3: The independent factors User perception of federated learning's trustworthiness and robustness have a positive correlation with user adoption intent of AI-based tools in medical imaging in radiology.

CHAPTER V:

DISCUSSION

5.1 Discussion of Results

Federated learning is a technique for training machine learning models with the with remotely hosted datasets without the need to accumulate data and, therefore, compromise it. Data is collected, processed into a dataset, and taken to the central server to train the dataset into any model, and we achieve a predictive output. It helps us to take the algorithm to the data instead of doing this federated learning, and then carry the result to the central server. This implies that the user would not be asked to upload their individual information. Predictive maintenance is given by federated learning. According to the outcomes in the central server, predictive maintenance allows a forecast of when the system will need maintenance.

In the healthcare domain, federated learning use cases for devices would allow the user to learn a model of machine learning that will help patients improve certain aspects of their health without having to upload their data to a central cloud. Federated learning entails using a wide corpus of high-quality decentralized data distributed through several client devices for instruction. Since the model is trained on client computers, no data from the user is expected to be submitted. Keeping the client's personal data on their computer gives them clear and physical control of their information.

The goal of this research was to analyze the impact of trustworthiness and robustness of federated machine learning algorithm approach in AI-based tools in radiology medical imaging and ascertain how these affect the intentions of adoption of AI-based tools in radiology by users.

The study was limited in scope, but it provides a baseline and a preliminary look at the effectiveness of general factors that impact developers to pursue their research and build business models using federated learning approach.

The findings reveal a significant correlation between perceived robustness and trustworthiness and adoption intentions, while also postulating logical next steps for future research and highlighting the implications for the business and academic worlds.

This study sought to answer the question " What are the associations of federated machine learning approach over traditional machine learning towards perceived trustworthiness and robustness and its prediction with user's adoption intent of AI-based tools in radiology medical imaging?"

The question led to two hypotheses that observed each of the independent variables (trustworthiness and robustness) and their correlation with adoption intentions. The third hypothesis studied the relationship between trustworthiness/robustness and adoption intentions. The findings showed that the variables have a positive correlation with adoption intentions and that, in this study, user perceived trustworthiness and robustness positively sway user adoption intentions.

5.2 Discussion of Research Hypothesis 1 and 2

H1 & H2 examined specific user attributes (trustworthiness and robustness) and how they correlated with adoption intentions. Prior to the study, I presumed that these would have a positive impact but was unsure how impactful they would be.

Due to effect size (r=.549 for H1 and r=.303 for H2), it can also be stated that there is a moderately positive effect (medium) for trustworthiness and positively small effect for robustness individually in correlation with adoption intentions. Therefore, each of these variables will independently impact adoption intentions.

5.3 Discussion of Research Hypothesis 3

The results from linear regression could not find statistically Significant influence perceived robustness had positive increase in the adoption intentions of users. The average of the trustworthiness variable had a positive unstandardized beta of .586 and the average of Robustness had a positive unstandardized beta of .020. We can accept the hypothesis that there is a positive relationship between both trustworthiness and robustness with users' adoption intent.

CHAPTER VI:

SUMMARY, IMPLICATIONS, AND RECOMMENDATIONS

6.1 Summary

This study can also serve as a starting point for further empirical studies in the context of individual adoption of federated learning in AI-based medical devices.

To advance further with use of AI-based tools in medical imaging, machine learning vendors and users must adapt federated learning approach. To assess the business value of investing and deploying federated learning approach, it is critical to understand the influence of perceived trustworthiness and robustness of federated learning algorithm with user adoption intent.

The results of the findings indicate that there is positive correlation of trustworthiness and robustness of federated machine learning models which influences the user adoption intent of AI-based tools in radiology medical imaging.

6.2 Implications

6.2.1 Business Implication

A thorough understanding of user attitudes and perceptions is required for successful implementation of an AI-based system (Romero-Brufau et al., 2020). Due to concerns about the reliability and robustness of AI-based tools, healthcare professionals still express fundamental concerns about implementation leading to adoption challenges (Kaissis et al. events, 2020).

In this machine learning process, when the model shifts to using federated machine learning versus traditional machine learning, based on the results of document search and in my opinion, there is a gap in understanding of the accuracy of the data. perceived reliability and robustness of federated and correlated machine learning algorithms. for the purpose of user acceptance. Masud et al., (2019) conclude that in general, radiologists' perceptions have not been taken into account and the details of implementation methods for applying machine learning tools are unclear and unresolved. report.

Most respondents were radiologists (45.10%) and different countries, including the United States, Germany and India, were well represented in the study. Of the participants, 41% used AI-based tools in radiology. Perception of federated learning as perceived is average. Approximately 43.1% have participated or contributed to research or testing on federated learning.

The results provide valuable information on awareness, knowledge, and involvement in federated learning. Research indicates that there is a positive correlation between users' perception of trustworthiness and strong and their intention to accept. The results give confidence in the development and use of federated learning, as it represents a step change in the way AI-based services are delivered. User acceptance is essential as it will involve investment and accountability in using AI-based tools due to specific development management, expertise and IT infrastructure requirements as well as share machine learning models from a specific institution or hospital.

To dispel those doubts about the timing, whether related to the installation procedure or the time the radiologist will need to spend getting the proper training and familiarity with the software.

On the hospital/clinic side, there are methods, including Turnaround Time (TAT), that can be used to objectively measure productivity in radiology and thus understand time savings. time thanks to AI (Griffith, Kadom and Straus, 2019).

With regard to monetary matters, the hospital/clinic can always perform various analyzes to determine the return on financial investment of the RN. These analyzes are typically done in the form of a health technology assessment (HTA) that will provide the most transparent way to drive value for money in healthcare. In general ETS terminology, budgetary impact analysis (BIA) assesses the short- and medium-term financial consequences of introducing new technologies (Mauskopf et al., 2007). BIAs are often presented in conjunction with other economic assessments, such as cost-effectiveness analyses, assessing both costs and, importantly, the impact of alternative health interventions (such as survival rate or quality-adjusted life years (QALY).

6.2.2 Academic Implication

Federated learning offers a secure collaborative machine learning framework for different devices without sharing their private data. This attracted a lot of researchers and there is extensive research happening in this domain. There is a growing interest in research about AI-centric technologies, yet individuals have not integrated AI devices into many aspects of their lives

6.3 Recommendations for Future Research

One promising research direction would be to examine public opinion in other healthcare settings, such as when federated learning-based AI tools are deployed and used in hospitals and hospitals. Healthcare professionals recommend patients to use AI devices.

Further research is needed to examine user perceptions of value in situations where the use of AI devices under federated learning can be a required part of diagnostic performance and complete the treatment for the patient. There is little research on AI perception, and clinical researchers can provide insights into public perception.

In addition, "robust" as a predictor needs to be further evaluated after implementation in a real clinical setting, as federated learning is still in the conceptualization phase of R&D and Actual commercialization will differ in the performance of the algorithm compared to the traditional machine learning algorithm due to several factors and disadvantages discussed.

Finally, public perception of AI in healthcare using federated learning may be limited by their knowledge of AI definitions and capabilities, as highlighted by our finding that there is a need to improve public knowledge of AI. Therefore, it is not possible to assess the priority or importance of each perception and need.

Developing an AI technology

1. Identify regulatory requirements.

The regulatory framework for AI is evolving. While most regulatory frameworks deal with data protection, data security, and privacy, the new governance guidelines cover equal access and human autonomy. Compliance measures must be included in the development and updating of technology.

Specific Considerations:

– Comply with country or region-specific export guidelines and rules, such as EU GDPR, Singapore Data Protection Act or Data Portability United States and liability for health insurance.

- Define open concepts and open standards that must be specified for compliance, for example, in Article 22 GDPR, "far-reaching" in "People who may not only be subject to automated decision-making with far-reaching effects".

– Identify relevant open standards and concepts that can be demonstrated to affected parties and professionals with relevant knowledge of the application.

2. Establish data management plans.

Clear protection guidelines and management plans should be established for data collection, storage, organization, and access to ensure data security and maintain privacy and security.

Specific Considerations:

– Understand the requirements and regulations for data collection and sharing in potential user countries, sectors, and organizations, including legal requirements for managing privacy consent to the use of training data.

– Defines the type of data to be collected and where and how it will be stored.

– Evaluates the physical infrastructure and operational processes that can be used to ensure data integrity and security.

– Understand and define how security and privacy will be protected in different contexts.

– Establishes guidelines and protocols for the collection, storage, organization, access and use of personal data, proprietary data, and public data in various contexts.

– Define how long data is stored, when data can be shared, and other temporal considerations.

– Prioritize the use of anonymous data whenever possible.

– Identify who is responsible for data governance and ensure proper follow-up.

– Clearly define all teams that will have access to data throughout the product lifecycle.

– Identifies any type of secondary data usage that may be allowed.

3. Adopt standards and best practices.

Ensure compliance and/or interoperability of AI technology with other technologies to be incorporated into healthcare systems. One or more established international, regional, or national performance standards and/or benchmarks for AI technology must be applied in accordance with regulations, guidelines and application requirements, design plans and develop.

Specific Considerations (Sample Standard):

– ISO Standard (Security and Privacy)

– US National Institute of Standards and Technology (Security and Privacy)

– IEEE Series 7000 (Privacy and Equity)

– Health Level 7 (Transfer of administrative rights and clinical health data).

Clinical Deployment

Successful implementations of AI in clinical workflows must be free of inequities and errors. What is needed is an appropriate and ethical AI that translates into more equitable care. The more data it captures, the more accurate and general it becomes.

The recipe for successful deployment is to understand the data in which the models are built and the environment in which they are deployed. The four key considerations during development and implementation are: data assessment, model boundary planning, community involvement, and trust building.

Identifying inequities in the data and taking them into account will lead towards more equitable healthcare as data quality largely determines model quality. Procuring the right datasets is the key and it may depend on data collection techniques. There are various methods to remove bias in data. An expert discussion and labeling adjudication can address Individual-level bias.

A population-level bias can be addressed via missing data supplements and distributional shifts. To determine the generalizability of models across different populations, medical devices, resource settings, and practice models, international multi-institutional assessment is an effective method. In addition, using multitasking learning to train models to perform a variety of tasks makes them more useful and often more powerful, rather than a narrowly defined task, such as development. showing many cancers from histopathological images.

A well-established transparent reporting process can reveal potential weaknesses and help address model limitations. Sufficient controls should be in place to protect against worst-case scenarios such as minorities, layoffs, or automation bias. An understanding of the specific cases in which the model fails is required. Models should be assessed for demographic performance to check for potential biases.

Inadaequate distribution of data during model training could lead to limitations in performance. It is critical to develop detection of out-of-distribution data and help detect anomalies. Additionally, methods are being developed to understand the uncertainty around model performance. This is especially important when making patient-specific predictions that affect safety. Involvement of all stakeholders such as patients, physicians, IT professionals, and other stakeholders is key to a successful implementation. This identified causes of racial bias, specifically by uncovering bias in the dataset and identifying demographics where the models failed.

Usability test results are valuable to ensure they are suitable for real-world use. This is considered the best way to evaluate model results to support clinical decision making and deployment in resource-constrained environments, such as areas with intermittent connectivity. For example, when rolling out ML-powered diabetic retinopathy models in Thailand and India, the researchers found that the model's performance was influenced by socioeconomic factors and determined that the place where the model is most useful cannot be where the model is generated. It has also been found that ophthalmic models may need to be implemented in endocrinology care, as opposed to ophthalmology centres, due to accessibility issues in specific local settings. AI models will require rigorous evaluation through clinical trials to assess safety and effectiveness. Another effective way to give doctors confidence in AI results is to parallelize ML models with existing workflows (e.g. manual scoring). AI can also help support clinical trials through a number of applications, including patient screening, tumor monitoring, side effect detection, and more, creating an ecosystem in which AI can help design secure AI.

Trust in AI is the foundation for both clinicians and patients to adopt. The foundation of clinical confidence will largely come from rigorous prospective trials that validate AI algorithms in real clinical settings. These real-world usage environments incorporate responses where real-world conditions can be difficult to predict and that AI technologies must account for. The randomness and human factors of the clinical setting cannot be captured in retrospective studies. Prospective trials are best considered to reflect clinical practice with measurable benefit in real-world implementation. AI must be explainable. Predictive models should be able to describe why specific environmental or patient factors led them to their predictions. In addition to clinical trust, patient trust, privacy issues need to be addressed. A significant need is for next-generation regulations that take into account advances in privacy-preserving techniques. ML often does not require traditional identifiers to produce useful results, but there are important signals in the data that can be considered sensitive. To unlock insights from these types of sensitive data, the development of privacy protection techniques must continue and new advances must be made in areas such as federated learning and analysis.

Improving AI technology after deployment

Multiple stakeholders need to be involved and trained for implementation and maintenance. Ensuring a better understanding of needs and building tailored solutions for multi-stakeholder inclusion should be a cross-cutting priority.

Specific considerations:

- Clearly define responsibilities for what, when, and how.
- Design, discuss and validate the proposed approach with various stakeholders in all targeted areas, including policymakers, project owners and managers projects, project managers, engineers and solution developers, potential users, domain experts, and information privacy and ethics experts.

- Educate stakeholders on why, how, and when to use the tool, including its main purposes, functions and features, and differences between use cases, if applicable.

– Continually engage with stakeholders and support users.

2. Evaluate and improve performance.

The health care outcomes and impacts of AI technology must be formally assessed, and the design and development of the technology must be continuously improved according to the ethical principles that guide its development. as well as new guidelines, governance and all applicable legal obligations and regulations. The risks of the technology and its intended use in different healthcare environments must be assessed regularly to manage deployment, ongoing development, and maintenance.

The accuracy and error risk of AI technology needs to be assessed to assess the impact on:

– Integrate, verify, and validate changes to a tool or system.

– monitor and ensure the effectiveness and usefulness of a tool or system over time; – how long the results or technology can be used.

– tool or system update frequency; and

- the person responsible for updating

Transparency and explainability of AI-based devices

Machine learning black boxes create challenges for regulators, who may not fully appreciate new AI technologies because of the standard metrics used to assess their safety and effectiveness. medical knowledge as well as scientific understanding and clinical trials do not fit into the medical black box. Complex algorithms are difficult for regulators (partly due to their lack of expertise) and difficult to explain to developers.

Improving the scientific understanding (explainability) of algorithms is considered essential to ensure that regulators (as well as clinicians and patients) understand how a system makes decisions. Explainability is also a requirement of the EU GDPR and is being incorporated into law in other countries facing the proliferation of AI in healthcare and other areas. It is argued that, if there is a trade-off between transparency and accuracy, transparency prevails. However, this requirement may not be feasible or even undesirable in a medical context. While it is often possible to explain why a particular treatment is the best choice for a particular condition, it is not always possible to explain how that treatment works or how it works. inhibit its action, as medical interventions are sometimes used before their mode of action is understood.

Confidence in expert decisions and recommendations depends on the expert's ability to explain why a given system is the best choice for achieving a clinical goal. These explanations must be based on reliable evidence of the AI system's superior accuracy and accuracy over alternatives. Proof should be generated by testing the system in the future in randomized trials, not by its performance against existing data sets in the laboratory. There is strong, prospective evidence for the system's performance in future clinical trials. An explanation of how a system arrives at a particular decision can encourage the use of machine learning systems for purposes for which they are not suitable, since the models created by those systems are based on association. association between a bunch of variables, it doesn't have to be. reason. If the associations are causal, practitioners can rely on them to make decisions without the system being tested or validated. Requiring every clinical decision about AI to be "explainable" could also limit AI developers' ability to use AI technologies that outperform legacy systems but cannot solve them. like it.

Clinical trials ensure that the hazards and unintended consequences of AI-enabled applications can be completely identified, addressed, and avoided, and further tested and monitored by AI devices approved can demonstrate the performance of such equipment and any changes that may occur after approval. Clinical trials, especially those conducted with different populations, can also reveal whether AI technology is biased towards certain subgroups, races, or ethnicities. (see below). If an algorithm is expected to change over time with new data, the validity of the results may be questioned. However, clinical trials may not be suitable because it takes a long time to conduct the test properly and incurs great costs.

As AI-based technologies and products become increasingly personalized for smaller populations, it will be more difficult to test with enough people. Statistical analysis strategies and clinical trial design need to be reevaluated and innovation encouraged in these areas of AI validation. While AI needs to be validated in clinical trials or other means of adoption, AI itself has the potential to enable more precise testing of device or drug efficacy with large numbers of patients. than. approach. This may become relevant during the COVID-19 pandemic, as recruitment and access to medical facilities are difficult.

Regulators could introduce “softer pre-market assessments” in lieu of clinical trials for AI technologies for health, assessing the protections that regulators offer development, quality of data used, development techniques, validation process, and "robust post-market monitoring". However, this can be difficult to do in practice, especially with post-marketing surveillance of new algorithms, and it may be too late to avoid harming particularly vulnerable people, such as For example, people who do not have access to a health care provider can protect them from being misdiagnosed. or wrong advice. The transparency of the original data set can be improved, including where the data comes from and how it is processed, as well as the transparency of the system architecture. Such transparency will allow others to independently validate AI technology and increase user trust.

While greater transparency of components of an AI system, including source code, input data, and analytical methods, can facilitate regulatory oversight, some transparency can can distract attention. Reviewing lines of code is time consuming and may not provide information on system performance, functionality, and accuracy before and after integration into a healthcare system. Management should provide incentives to encourage developers to identify and avoid biases. One example is adding metrics to a pre-certification program organized by the US Food and Drug Administration, the Digital Health Innovation Action Plan. The program evaluated medical software on the basis of excellence criteria, including quality. Quality and other criteria established by regulatory bodies may include the risk of bias in training data. Robust post-market monitoring to identify biases in machine learning algorithms, including working with vendors and communities that may be impacted by biased algorithms, can improve regulatory supervision.

The proper management of the private sector must overcome a number of obstacles. One is the strength of the many companies involved in providing AI for healthcare. Many of them employ former government and regulatory officials who are required to lobby and influence policymakers and regulators responsible for overseeing the work. using AI for healthcare. This could affect governments' ability to act independently of companies. A second challenge is that many technologies developed by companies are increasingly difficult to evaluate and monitor, in part due to their increasing complexity, including the use of black box algorithms and deep machine learning methods. Increasing complexity has encouraged both governments and businesses to consider "co-management" models, in which each relies on the other to evaluate and regulate a technology. While such surveillance models can help governments understand technology, they can limit the government's ability to exercise independent judgment and encourage them to believe that companies are willing to strictly self-regulate their activities. Improving private sector governance in other ways will require expertise and more independent inside information so that governments can effectively assess and regulate corporate activities. Therefore, improving the capacity of public regulators and transparency will play an important role in improving government oversight of the private sector. These measures could include greater transparency of data collected and used by private companies, how ethical and legal principles are integrated into company operations, and how products and services in practice, including how algorithms change over time.

Governments are increasingly required to disclose the use of algorithms in services and operations to promote accountability for AI use, and many data protection laws require decisions not to be made. show. prevented in certain contexts. In France, the government is required to provide a general explanation of the operation of any algorithm used by the government, a personalized explanation of the decisions made by the algorithm, an explanation of the decisions and publish the source code as well as other material on the algorithm.

In general, governments are increasingly expected to be transparent about their use of AI, including whether they are investing in AI, entering into enterprise partnerships, or developing the technology in a transparent manner. independent in public companies or government agencies. Governments are also expected to be transparent about any harm caused by the use of AI and the steps taken to remedy any such harm. A study by the UK Commission on Standards in Public Life found that the UK government (during the period under review) failed to adhere to established principles of openness and noted that " According to the principle of openness, the current lack of information about government uses of AI risks undermining transparency."

However, transparency may not be enough to ensure that government use of algorithms will not cause undue harm, especially to marginalized communities and populations. Greater public participation from a wide range of stakeholders is needed to ensure that decisions about the introduction of AI systems in healthcare and elsewhere are not made solely by public and business officials. out, but based on the public participation of more stakeholders. include representatives of public interest groups and leaders of vulnerable groups who are not normally involved in making such decisions. Their opinions must be gathered before and not only after an adverse impact has been identified, i.e. it is too late.

It is often argued that strict regulations limit innovation and deprive health care systems, providers, and patients of beneficial innovation. A balance must be struck between protecting the public and promoting growth and innovation. The medical use of AI is still new and often untested, and policymakers and regulators need to consider a variety of ethical, legal, and human rights issues. For example, regulators must identify AI-based applications and devices that could be described as "snake oil," a euphemism for misleading, fraudulent marketing in the industry. healthcare or fraud. follow health advice that may be contrary to their health.

Applications that do not provide a therapeutic or health benefit may only be introduced to collect health and biological data for use in commercial marketing or to incentivize patients to pay for measures. unrelated or unproven health interventions. For example, academic data obtained from 300,000 Facebook users was told the data was for "psychological testing". Their data and the data of approximately 50 million other users associated with them ("Facebook friends") were then sold to Cambridge Analytica, which used this data to create predictive and influencing software. affect the selection at the ballot box. Malicious use of data collected on behalf of academic or healthcare purposes can expose healthcare systems, healthcare providers and service providers to Healthcare-related AI services are at significant risk. Regulation may vary based on risk, so that particularly vulnerable people, including people with mental illness, children and the elderly, are protected from misinformation and bad advice from health apps instead of tapping to help these people. People living in resource-poor environments, in countries that do not have the resources to regulate and monitor the adverse consequences of AI applications, and suffer from diseases that lead to marginalization and discrimination. Treatments, such as HIV/AIDS or tuberculosis, also need more protection and oversight from regulators than users of lifestyle or healthcare apps. AI uses complex computer algorithms to simulate human perception, but is scalable to analyze large data sets. The field of AI is growing rapidly and has significantly infiltrated almost every aspect of human life, including healthcare. Incorporating AI-based tools and techniques is expected to improve healthcare delivery by making healthcare accessible and affordable, and at the same time improve the quality of care provided. For example, AI as well as radiologists can read CT scans automatically. Tuberculosis screening can be performed by AI using chest X-rays with comparable performance to molecular tests, and mammograms can be used to predict the onset of breast cancer before when the visual cues appear. Therefore, AI for health has been recognized by researchers as well as the government as one of the important areas.

An ethically sound policy framework is essential to guide the development of AI technologies and its application in healthcare. In addition, as AI technologies evolve and are applied in clinical decision-making, it is important to have procedures in place to discuss liability in the event of errors to protect and protect guard. Like any other diagnostic tool, AI-based solutions themselves cannot be held accountable for their own decisions and judgments. Therefore, it is important to hold responsibility at all stages of AI development and implementation for health.

Despite all the potential benefits, the application of AI to healthcare highlights a number of ethical, legal, and social concerns, especially regarding its development and implementation. The field can be largely guided by well-established health research principles, but the development and implementation of AI-based solutions in healthcare face a number of challenges. challenges, including those related to data security, data sharing, data privacy, and so on. For example, AI-based solutions can empower the masses by enabling early and easy diagnosis and access to healthcare facilities, but the use of these tools and techniques without unsupervised risks. Therefore, a legal and ethical framework is required before health AI becomes part of health research and healthcare delivery. While general principles related to biomedical research and healthcare delivery are applicable to health AI, the field also has unique ethical considerations.

When developing AI technology for health care applications, similar ethical principles can be followed. However, as AI technology presents some unique methodological and interpretive challenges and in the context of rapidly changing healthcare scenarios, the guidelines have been developed in consultation with experts. from these two fields. The purpose of these guidelines is not to constrain innovation or recommend specific disease-specific diagnostics or treatments, but rather to guide the effective development, implementation, and adoption of AI-based technologies. but safe in biomedical research and healthcare delivery. These guidelines will be used by experts and ethics committees reviewing research proposals related to the use of AI-based tools and technologies. There is no standard and universally accepted term and thus the term AI technology has been used to refer to AI technologies, AI applications, AI models, AI products, AI-based solutions or AI-based solutions throughout the document.

Trustworthiness

Trustworthiness is the most desirable quality of AI-based tool in healthcare. Clinicians need to build trust in the tools that they use and the same goes for AI. In order to effectively use AI, clinicians and healthcare providers need to have a simple, systematic and trustworthy way to test the validity and reliability of AI technologies. In addition to providing accurate analysis of health data, a trustworthy AI-based solution should also be:

i. Lawful, i.e., it must adhere to all applicable laws and regulations.

ii. Ethical, to ensure adherence to ethical principles and values cherished by the community. Agencies involved in developing and deployment of AI should cultivate trust in the general public by adopting ethical principles at all stages of development

iii. Reliable and valid, both from technical and social perspectives, to ensure predictability in the results and outcomes of AI-based solutions when applied in variety of clinical settings. The results thus obtained also should be in sync with standard assessment tools.

iv. Explainable, i.e., the results and interpretations provided by AI-based algorithms should be explainable based on scientific plausibility. It should be possible to understand the logic behind the results obtained so that AI technology is valid, reliable and responsible. The lack of information about the decision-making by AI algorithms has prompted some to label it as a "black box" which can prove a deterrent to its wider adoption. A well explainable AI-based solution is expected to improve the confidence of both the patient and the health professionals.

v. A diagnostic AI technology may produce results that are not in line with the physicians' views/ decision on disease. Such situations may question the credibility of the system as well as the doctor. In such cases, the physician may seek the help of their colleagues or may consult with AI developers. The patient should be informed about the recommendation from both the doctor(s) and the AI technology. The patient must have the ultimate autonomy to choose over whether to accept or reject the AI technology-generated decision.

vi. Transparent, i.e., Details about the development and deployment must be easily available to all the stakeholders to enable them to make an informed decision. AI developers should ensure transparency in every step so that consumers can make informed choices about sharing their data and using AI. The end user must be provided with adequate information in a language they can understand to ensure that they are not being manipulated by the AI technologies. The end-user must be informed about the intention, outcome and limitation of using AI technologies. In absence of transparent information about the processes involved it is difficult to expect large scale adoption of AI for Health. This is also important for legal and regulatory purposes in cases where undesirable clinical outcomes may arise out of the inaccurate interpretation and or recommendation by AI-technologies. Therefore, for the regulation, acceptance, and deployment of AI technologies transparency, explainability and functional understanding is necessary. Limitation in transparency of the system impairs validation, clinical recommendations, and make it difficult to identify errors and biases.

vii. Sufficient information must be published widely before deploying AI technologies in the health care sector. An adequate platform must be there to ensure the input of public consultation and debate regarding design, usage, safety security, etc. such information must be published regularly and must be documented.

viii. All AI technologies must comply with legal norms. Developers must be able to demonstrate and interpret how the AI technology complies with data and privacy laws. All software/ privacy policy updates in an already established AI technology must comply with legal norms.

ix. AI technologies must have an ethical responsibility to be as transparent as domestically developed AI technology and comply with the law. The assessment will cover all stages as with any domestically developed AI technology.

x. Conflict of interest arising at any stage of development must be disclosed and made public on public platforms.

Data Privacy

AI-based technology must ensure privacy and protect personal data at all stages of development and implementation. Maintaining the trust of all stakeholders, including healthcare recipients, in the safe and secure use of their data is paramount to successful AI implementations and widely. Data security should aim to prevent unauthorized access, modification and/or loss of personal data. The application of artificial intelligence to personal data must not unreasonably restrict the actual or perceived freedom of individuals. These practices are important in the healthcare industry, where medical information is sensitive data that, if misused, could harm patients or expose them to discrimination, even if it is not. love. Personal patient data should preferably be anonymized unless keeping it in an identifiable format is essential for research or clinical purposes. All algorithms that process patient data must ensure proper anonymity before any form of data sharing. It is important to know that patient identifiers can be presented as "metadata" and "image" data, and both must be effectively anonymized. Data ownership issues are complex and vary depending on national or regional laws and regulations. It also depends on the level of anonymity of the data. Because data to build AI applications is often gathered from a variety of sources (e.g., medical and insurance records, pharmaceutical data, genetic data, social media, data GPS, etc.), tracing this data back to the patient could potentially become easier and (intentionally or not) defeat privacy objectives. The existing data protection law in India is the Computer Act 2000. Subject to Section 43A, legal entities possess, process or otherwise process sensitive personal data or information in computing resources in their possession, control or operator will be liable for damages such as compensation to data subjects if they are negligent in implementing and maintaining reasonable security processes and procedures to protect data. or sensitive personal information. To ensure the privacy and security of healthcare data, the government of India is enacting a new healthcare data protection law – the Healthcare Digital Information Privacy Bill (DISHA) and personal data protection (PDP); these will tie into the ethical principles of AI technology.

d. Users should have control over the data that has been collected from them for the purpose of developing and designing AI technologies for healthcare. Users must be able to access, modify or delete this AI technology data at any time.

ii. End users should be clearly informed about safeguards designed to protect privacy. They must have knowledge of the type of data collected and how it is used, to develop AI algorithms or to interpret or store it.

iii. AI technology's prediction algorithms can produce inconsistent results that could compromise patient privacy. Consent must be obtained before running the prediction algorithm on participants/patients.

iv. AI technologies that require human biometric data should have additional security measures in place to protect the data. EC/regulatory authority approval is required to use this data. Such accidental data leakage could have unprecedented consequences

v. An impact assessment should be carried out by the relevant authorities before implementing AI for widespread use. It should focus on key areas such as human rights, privacy, and ethical principles

vi. The manufacturer is responsible for preventing re-identification from the dataset and preventing leakage of identifying information.

vii. Sharing data may expose patients/participants to privacy threats. Additional patient consent is required to share data if not done previously. Consent must include the nature of the data, to what extent it is shared, and the damages that may result from sharing the data.

vii. Redundant data is collected contributing to redundant data. Reusing redundant data without patient/participant consent is unethical. Storing redundant data for future use may require additional consent, if not done earlier.

Accountability and Liability

Accountability is described as the obligation of an individual or organization to be accountable for its activities, to accept responsibility for its actions, and to disclose results transparently. AI technologies intended for deployment in the healthcare sector must be available for review by relevant authorities at any time. AI technologies undergo regular internal and external tests to ensure they function optimally. These audit reports must be made public. AI developers must allow independent analysis and testing of their

i system. AI innovators may not be familiar with medical ethics, research regulations, and regulatory guidelines that apply to the field. Therefore, it is important to have healthcare industry representation at all stages of the development and implementation of AI-based tools and technologies.

ii. Automation is the most direct benefit of AI-based solutions. Machine-assisted decision algorithms are increasingly being used in clinical medicine. However, the inherent risks of misinterpretation in clinical settings with full automation require caution. Therefore, unlike other areas of AI technology that often deploy unattended, AI for healthcare must always be appropriately supervised. Open source software is also subject to ethical best practices to ensure that ethical considerations are not compromised at any time in the name of innovation.

iii. The "Human In The Loop" (HITL) concept puts humans in a supervisory role and is more suitable for healthcare purposes. This will ensure healthcare professionals make personalized decisions that are in the best interests of the patient at the center. Applying HITL principles throughout the development and implementation of healthcare AI also contributes to optimal sharing of responsibility by the team involved in the development and implementation of intelligence-based algorithms. artificial AI.

iv. It is essential to ensure that the entity or entities seeking liability has the appropriate legal and technical certifications in the field of health AI technology.

v. AI-based solutions can malfunction, perform poorly, or make wrong decisions that have the potential to harm recipients, especially if left unattended. The healthcare professional who will use the technology will assign responsibilities. Like other diagnostic and decision-making tools used in clinical practice, the responsibility for optimal use of technology rests with the healthcare professional using AI-based solutions to deliver services. health care.

vi. When implementing tools based on AI technology, liability for its use must be determined prior to clinical application or public use.

vii. Liability for damage caused by the failure of AI technology depends on the nature of the cause of the damage. If the problem is primarily due to a functional defect, the designer, developer, or manufacturer may be held liable. If damage occurs as a result of faulty technology implementations, the end user or organization may be held liable. A clear understanding and allocation of responsibilities is required before implementing AI technology.

vii. If AI technology has caused damage, then an appropriate mechanism is needed to determine the relative roles of the parties involved in the damage, from the manufacturer to the user, and their liability. All stakeholders involved in conceptualizing the distribution chain must cooperate and work together to minimize harm.

Optimization of Data Quality

AI is a data-driven technology, its results mainly depend on the data used to train and test the AI. This is especially important in the field of health AI, as a skewed and insufficiently large dataset can lead to problems related to data bias, errors, discrimination, and more. Data trends are seen as the biggest threat to data-driven technologies like AI for healthcare due diligence is needed to ensure "training data" is free of biases known and representative of the majority of the target population.

One of the main concerns he raised is that pre-existing bias arises in AI models when making decisions against a particular group of people, mainly due to human participation in the formation. into that data, obscuring the AI's judgment. The challenges inherent in machine learning, the logistical difficulties of implementation, and consideration of barriers to adoption as well as the necessary itineraries or sociocultural changes.

Before implementing AI technologies, opportunities for bias should be carefully considered, identified, and considered. The training data should have no sampling bias.

Such sampling bias can affect the quality and accuracy of the data. Researchers must ensure the quality of the data.

ii. The datasets used in AI technologies must adequately represent the population for which the technologies are intended to be used. Data on ethnic minorities, marginalized populations and remote areas must be fully representative, otherwise oversampling may be necessary to obtain the same quality of results observed with population groups are better represented.

iii. The existence of bias in the data set has the potential to affect the operation of AI technology. If there are allegations of discrimination or signs of bias in AI technology, the operation of such a system will be temporarily halted. The manufacturer is responsible for eliminating bias. Proving unbiased AI technology with optimal functionality to the competent authority is required to continue operating.

iv. Data collection and the development of AI algorithms present various challenges and trade-offs, and developers and researchers need to ensure that the best data can be used for the field. their specific use case.

v. These inherent data problems can be mitigated through rigorous clinical validation prior to the use of any AI-based technology in healthcare.

vi. All emerging technologies, including AI, must go through a well-established review process that applies to all areas of biomedical research and clinical care. In fact, it is prudent to implement a "test-before-deploy" process at each new location where AI is deployed so that AI performance can be guaranteed locally. A strong mechanism is needed to monitor data collection methods, which can check the fairness and completeness of data collection and can point out inadequacies and misrepresentations

vii . Poor data quality, inappropriate and incomplete data representation can lead to bias, discrimination, errors, and suboptimal performance of AI technology.

6.4 Conclusion

Digital health and AI are becoming increasingly important in medical imaging as the industry adopts new technologies and treatments. The next wave of digital disruption will be the use of artificial intelligence, and businesses should prepare for that now. Given the hype and gloom, AI has yet to be deployed on a large scale commercially. The adoption of AI-based tools in medical imaging has become a key driver of the technology. Machine learning (ML) is a way to enable AI's "learning" capabilities. It involves the use of a set of learning algorithms driven by mathematical techniques that allow machines to learn from data, rather than being explicitly programmed to perform certain tasks. The training process uses a learning algorithm to derive relationships between data points from training data, usually a subset of historical data. The results of the training are trained machine learning models, which can make predictions or make decisions based on observed data patterns from user-supplied input data or queries.

This growth has sparked new discussions about industry regulations, data collection, and most importantly, the enormous potential for treatment and patient outcomes. There have been significant advances in the use of AI algorithms in the field of medical image analysis. Special deep learning has made continuous breakthroughs in radiology practice. The main consumer proposition for AI companies in the radiology market stems from the exceptional ability of AI tools to recognize complex patterns in imaging data.

Medical imaging is said to have seen some of the most important developments in AI technology due to parallel improvements in machine vision. However, security and privacy concerns are not limited to medical imaging.

As the adoption of AI applications in healthcare accelerates, the urgent need for proper governance to address ethical, regulatory, and trust concerns is urgent.

The application of AI-based tools in medical imaging radiology is not yet recognized as a reliable tool to assist users. Current challenges related to data silos, generalization, and privacy issues need to be addressed in accordance with regulatory requirements.

Federated Machine Learning has emerged as a promising method for building precise and robust algorithms that serve as tools to help users free up their workloads. In order to understand user acceptability, understanding the reliability and robustness of federated learning is essential to both the R&D investment of the developers of such tools.

This study provides a basis for understanding users' perceptions of the reliability and strength of federated learning algorithms as well as their current perceptions of intention to adopt AI-based tools in radiology medical imaging.

The results are encouraging as they indicate a positive correlation between user perceptions of reliability and durability as well as intention to use. Future research can be conducted to gain a deeper understanding of users, especially radiologists, on how implementing federated learning will change the current standard of practice. Alternatively, patient perception of federated learning can be investigated after implementing federated learning algorithms in real-world clinical practice.

Implementations of AI applications are considered special clinical projects similar to other clinical initiatives in hospitals. One of the key stakeholders and drivers for successful adoption of federated learning will be IT professionals in hospitals and operations. They will play a leading role in the structure. Further research involving these key stakeholders can be conducted to develop the governance framework. Medical imaging is a vast and complex field that encompasses many imaging modalities, disease states, and diagnostic protocols. Machine learning, on the other hand, is an active field of research with thousands of new techniques being published every year. The combined diversity of the two fields as well as inconsistent hospital practice, limited data sharing regulations, and lack of standard outcome reporting make it difficult to clearly assess its role and potential machine learning applications in medical imaging. Machine learning has great potential to improve diagnostic accuracy, reduce reporting times, reduce radiologists' workloads, and ultimately improve healthcare delivery. However, to realize this potential, a concerted effort is required from physicians, radiologists, patients, hospital managers, data scientists, software developers and stakeholders is different jurisdiction.

Hospital procedures and practices vary widely between hospitals, even within the same geographic region. This increases the difficulty of seamlessly integrating predictive models into hospital workflows. Workflow heterogeneity also raises the question of whether the reported performance of one model is reproducible in another clinical setting. This is ongoing research and satisfactory solutions have not been found.

- The implementation of the FL pipeline for medical imaging could alleviate privacy concerns to a large extent. However, the peculiarities of healthcare organizations and medical imaging can lead to specific barriers that are significantly different from those encountered with other types of data.

- Healthcare facilities often lack the on-premises or cloud-based computing facilities needed to establish interconnected networks. They may also need to prepare data management and standardization pipelines and have robust network connections.

- Key functional challenges include intra-hospital bias, data heterogeneity, local model performance, and safety issues.

- Several FL algorithms have been designed to solve these problems. Some promising results improve security, communication costs, data heterogeneity, and model performance. Research is underway and a universal solution has yet to be found.

CHAPTER VII:

APPENDIX

APPENDIX A

SURVEY QUESTIONAIRE

<Message to participant>

Dear Participant,

My name is Pavan Kumar Malwade and am pursuing Doctor of Business Administration (DBA) from Swiss School of Business and Management (SSBM) Geneva.

I would like to invite you to participate in my dissertation thesis survey. My research project will test the hypotheses whether there is positive correlation of trustworthiness and robustness of Federated machine learning models which influences the user adoption intent of AI-based tools in radiology medical imaging

The survey has 12 questions.

The survey will be conducted using Google Forms which complies with security, privacy, and regulatory requirements. The data will be securely stored.

As token of appreciation of your investment of time, I would be happy to share the survey results within 1-2 months of completion. Please provide your email address to do so.

Here is the link for the survey <Insert Survey URL>

The Survey closes on <insert date>. Kindly complete at your earliest convenience.

I sincerely appreciate your valuable time and grateful for fully completing the survey.

Thank you,
Pavan Kumar Malwade

A. Background:

The adoption of AI-based tools is currently challenging as healthcare professionals still express fundamental concerns about the implementation due to concerns with trustworthiness and robustness (Kaissis et al., 2020). The major barriers in adaption the machine learning tools by customers were generalizability, lack of trust due to bias and safety mechanisms and complying with regulations through transparent and explainable algorithm while preserving data privacy and security.

Federated Learning is expected to address the problem of bias, generalizability, and privacy concerns of Traditional machine learning. Federated learning (FL) is a learning paradigm seeking to address the problem of data governance and privacy by training algorithms collaboratively without exchanging the data itself. (Reike et. al., 2020)

The experimental studies conducted in various modalities of radiology medical imaging has positive results showing federated learning performs better than traditional machine learning and theoretically improve the robustness of algorithm while preserving the patient data. Federated learning offers easy scalability, flexible training scheduling, and large training datasets through multi-site collaborations, all essential conditions to the successful deployment of an AI solution (Stripelis et al., 2021, Yang et al., 2021, Linardos et al. 2021, Sarma et al., 2021, Dayan et al., 2021 and Sheller et al., 2020).

Although there are several advantages of federated learning, important challenges like security concerns, regulatory compliance, model performance monitoring etc. must be addressed before federated learning is optimally able to build acceptable AI models (Xu et al., 2021 and Rieke et al., 2020).

B. Definitions of Terms:

Bias: systematic difference in treatment of certain objects, people, or groups in comparison to others (DRAFT ISO/IEC DIS 22989, 3.4.4. p 10)

Robustness: ability of a system to maintain its level of performance under any circumstances (DRAFT ISO/IEC DIS 22989, 3.4.11. p 11)

Trustworthiness: ability to meet stakeholders' expectations in a verifiable way (DRAFT ISO/IEC DIS 22989, 3.4.11. p 11)

Federated learning (FL): A learning paradigm seeking to address the problem of data governance and privacy by training algorithms collaboratively without exchanging the data itself. (Reike *et. al.*, 2020).

C. Questionnaire:

Trustworthiness

Please evaluate how much you agree with the statements below using the 5-point scale, with 1 being Strongly disagree and 5 being Strongly agree.

1. My perception of trustworthiness of federated learning over traditional machine learning is that

	1	**2**	**3**	**4**	**5**
a. It can preserve privacy	○	○	○	○	○
b. It can produce Results that are Unbiased	○	○	○	○	○
c. It can produce Explainable	○	○	○	○	○

machine learning models					
d. It can adapt to specific needs and stakeholder requirements	○	○	○	○	○
e. It inspires trust in machine learning models used in AI-based tools in radiology	○	○	○	○	○

Robustness:
Please evaluate how much you agree with the statements below using the 5-point scale, with 1 being Strongly disagree and 5 being Strongly agree.

2. My perception of robustness of Federated learning over traditional machine learning is that

	1	**2**	**3**	**4**	**5**
a. It can produce Accurate results	○	○	○	○	○
b. The algorithm performance is better over traditional machine learning	○	○	○	○	○
c. It produces results that are Generalizable for the intended population	○	○	○	○	○
d. Algorithm will always stay most updated	○	○	○	○	○

e. It is reliable over traditional machine learning models	○	○	○	○	○

Intention to use AI based tools

3. Does use of Federated Learning approach influence your adoption intent of AI based tools?

	1	**2**	**3**	**4**	**5**
a. I agree to use AI-based tools for clinical purposes	○	○	○	○	○
b. Using AI-based tools for healthcare purposes is something I would consider	○	○	○	○	○
c. I would like to use AI-based devices to manage my healthcare	○	○	○	○	○
d. In the future, I am willing to use AI-based services for diagnostics and treatments	○	○	○	○	○
e. I am very likely to use recommendations provided by AI-based tools for care planning	○	○	○	○	○

Participant Experience and Knowledge

4. Do you have knowledge of Federated Learning?
   a. Yes
   b. No

5. If answer is “Yes” for question 4, according to you what is your level of knowledge of Federated Learning?
   Please evaluate using the 5-point scale, with 1 being “Very low” and 5 being “Very high”.

(1) Very low	(2) low	(3) moderate	(4) High	(5) Very High
○	○	○	○	○

6. Are you currently a user of AI-based tool for radiology clinical workflows?
   a. Yes
   b. No

7. Are you currently a user of AI-based tool for other than radiology clinical workflows? (For example: Ophthalmology)
   a. Yes
   b. No

8. Have you participated/contributed for experiments related to Federated Machine Learning experiments or currently intent to do?
   a. Yes
   b. No

9. Which one among below roles describes you the best?

   a. Clinician
   b. Radiologist
   c. Algorithm Developer
   d. Data scientist
   e. Algorithm Vendor
   f. AI enthusiast
   g. others.

10. Your experience in relevant field or area. Eg: Healthcare, Data Science, Artificial Intelligence etc.
    a. 0-5
    b. 5-10
    c. 10-15
    d. More than 15

11. which country do you belong?
    < To provide option to select from Drop down list of countries or manually type name of country>

Optional

12. In case you wish to receive a copy of survey results, please provide your email address. The survey results would be shared within 1-2 months of completion.
    <Space for email address>

Actual Questionnaire (Google Form) sent to Respondents.

Doctoral Research Survey: Federated Machine Learning Algorithm in AI-based tools in Radiology Medical Imaging - Association of user perceived Trustworthiness and Robustness with their adoption intent.

Dear All,

My name is Pavan Kumar Malwade. I am a Regulatory Affairs Professional with over 13 years of experience in the Healthcare industry. I am pursuing Doctoral program (DBA) from Swiss School of Business and Management (SSBM) Geneva, Switzerland.

My research topic is "THE ASSOCIATIONS OF FEDERATED MACHINE LEARNING ALGORITHM'S PERCEIVED TRUSTWORTHINESS AND ROBUSTNESS ON USER'S ADOPTION INTENT OF AI-BASED TOOLS IN RADIOLOGY MEDICAL IMAGING". The outcome of the research is expected to provide valuable insights about correlation of user perception with adoption intent of Federated Learning technology.

I would request 8 -10 mins of your time to participate in my dissertation thesis survey consisting of TWELVE questions.

Please note that the survey is being conducted using Google Forms which complies with security, privacy, and regulatory requirements. The data will be securely stored.

As token of appreciation of your time, I would be happy to share the survey results within 1-2 months of completion. So, please provide your email address at the end of the survey. This is optional :).

The Survey closes on 20 February 2022. Kindly complete at your earliest convenience.

I sincerely appreciate your valuable time and am grateful for fully completing the survey.

Thank you in advance for your support.

Pavan Kumar Malwade

(LinkedIn: https://www.linkedin.com/in/pavanmalwade/)

Background:

The adoption of AI-based tools is currently challenging as healthcare professionals still express fundamental concerns about the implementation due to concerns with trustworthiness and robustness (Kaissis et al., 2020). The major barriers in adaption the machine learning tools by customers were generalizability, lack of trust due to bias and safety mechanisms and complying with regulations through transparent and explainable algorithm while preserving data privacy and security.

Federated learning (FL) is a learning paradigm seeking to address the problem of data governance and privacy by training algorithms collaboratively without exchanging the data itself. (Reike et. al., 2020)

The experimental studies conducted in various modalities of radiology medical imaging has positive results showing federated learning performs better than traditional machine learning and theoretically improve the robustness of algorithm while preserving the patient data. Federated learning offers easy scalability, flexible training scheduling, and large training datasets through multi-site collaborations, all essential conditions to the successful deployment of an AI solution (Stripelis et al., 2021, Yang et al., 2021, Linardos et al. 2021, Sarma et al., 2021, Dayan et al., 2021 and Sheller et al., 2020).

Although there are several advantages of federated learning, important challenges like security concerns, regulatory compliance, model performance monitoring etc. must be addressed before federated learning is optimally able to build acceptable AI models (Xu et al., 2021 and Rieke et al., 2020).

Definitions of Terms:

Federated learning (FL): A learning paradigm seeking to address the problem of data governance and privacy by training algorithms collaboratively without exchanging the data itself. (Reike et. al., 2020).

Trustworthiness: ability to meet stakeholders' expectations in a verifiable way (DRAFT ISO/IEC DIS 22989, 3.4.11. p 11)

Robustness: ability of a system to maintain its level of performance under any circumstances (DRAFT ISO/IEC DIS 22989, 3.4.11. p 11)

Bias: systematic difference in treatment of certain objects, people, or groups in comparison to others (DRAFT ISO/IEC DIS 22989, 3.4.4. p 10)

* Required

Skip to question 1Skip to question 1

Questionnaire

Please evaluate how much you agree with the statements below using the 5-point scale, with 1 being Strongly disagree and 5 being Strongly agree.

Trustworthiness

1. 1. My perception of trustworthiness of federated learning over traditional machine learning is that *

Mark only one oval per row.

	(1) Strongly Disagree	(2) Disagree	(3) Neither agree nor disagree	(4) Agree	(5) Strongly Agree
a. It can preserve privacy	⬭	⬭	⬭	⬭	⬭
b. It can produce results that are unbiased	⬭	⬭	⬭	⬭	⬭
c.It can produce explainable machine learning models	⬭	⬭	⬭	⬭	⬭
d.It can adapt to specific needs and stakeholder requirements	⬭	⬭	⬭	⬭	⬭
e.It inspires trust in machine learning models used in AI-based tools in radiology	⬭	⬭	⬭	⬭	⬭

Robustness

2. 2. My perception of robustness of Federated learning over traditional machine learning is that *

Mark only one oval per row.

	(1) Strongly Disagree	(2) Disagree	(3) Neither agree nor disagree	(4) Agree	(5) Strongly Agree
a. It can produce accurate results	⬭	⬭	⬭	⬭	⬭
b. The algorithm performance is better over traditional machine learning	⬭	⬭	⬭	⬭	⬭
c. It produces results that can be generalized for the intended population	⬭	⬭	⬭	⬭	⬭
d. Algorithm will always stay most updated	⬭	⬭	⬭	⬭	⬭
e. It is reliable over traditional machine learning models	⬭	⬭	⬭	⬭	⬭

User adoption Intent of AI-based tools

3. 3. Does use of Federated Learning approach influence your adoption intent of AI based tools? *

Mark only one oval per row.

	(1) Strongly Disagree	(2) Disagree	(3) Neither agree nor disagree	(4) Agree	(5) Strongly Agree
a. I agree to use AI-based tools for clinical purposes	⬭	⬭	⬭	⬭	⬭
b. Using AI-based tools for healthcare purposes is something I would consider	⬭	⬭	⬭	⬭	⬭
c. I would like to use AI-based devices to manage my healthcare	⬭	⬭	⬭	⬭	⬭
d. In the future, I am willing to use AI-based services for diagnostics and treatments	⬭	⬭	⬭	⬭	⬭
e. I am very likely to use recommendations provided by AI-based tools for care planning	⬭	⬭	⬭	⬭	⬭

Skip to question 4

Participant Experience and Knowledge

4. 4. Do you have knowledge of Federated learning? *

Mark only one oval.

- Yes
- No

5. 5. If answer is "Yes" for question 4, according to you what is your level of knowledge of Federated Learning? In case answer is "No" to question 4, please skip this question.

Please evaluate using the 5-point scale, wherein 1 refer to "Very low", 2 refers to "Low", 3 refers to "Fair", 4 refers to "High" and 5 refers to "Very high".

Mark only one oval.

	1	2	3	4	5	
Very Low						Very High

6. 6. Are you currently a user of AI-based tool for radiology clinical workflows? *

Mark only one oval.

- Yes
- No

7. 7. Are you currently a user of AI-based tool for other than radiology clinical workflows? (For example: Ophthalmology) *

Mark only one oval.

- Yes
- No

8. 8. Have you participated/contributed in research or experiments related to Federated Machine Learning or currently intent to do? *

Mark only one oval.

- Yes
- No

9. 9. Which one among below roles describes you the best? *

Mark only one oval.

- a. Clinician
- b. Radiologist
- c. Algorithm Developer
- d. Data scientist
- e. Algorithm Vendor
- f. AI enthusiast
- Apps specialist (Radiology)
- Pathologist
- R&D Project Lead
- Clinician, EM and Ai ML practitioner and researcher
- Medical Student
- Other: ______________

10. 10. Your experience in relevant field, role or area.

Mark only one oval.

- a. 0-5 years
- b. 5-10 years
- c. 10-15 years
- d. More than 15 years

11. 11. Which country do you belong?

Optional

12. 12. In case you wish to receive a copy of survey results, please provide your email address. The survey results would be shared within 1-2 months of completion.

CHAPTER VIII:

REFERENCES

Abdul Salam, M., Taha, S. and Ramadan, M. (2021). COVID-19 detection using federated machine learning. *PLOS ONE,* 16(6), p.e0252573. [online] Available at:10.1371/journal.pone.0252573(Accessed: 27 December 2022)

Allen, B., Dreyer, K., Stibolt, R., Agarwal, S., Coombs, L., Treml, C., Elkholy, M., Brink, L. and Wald, C. (2021). Evaluation and Real-World Performance Monitoring of Artificial Intelligence Models in Clinical Practice: Try It, Buy It, Check It. Journal of the American College of Radiology, [online] 18(11), pp.1489–1496. Available at: 10.1016/j.jacr.2021.08.022 (Accessed: 27 December 2022)

Attia, Z.I., Noseworthy, P.A., Lopez-Jimenez, F., Asirvatham, S.J., Deshmukh, A.J., Gersh, B.J., Carter, R.E., Yao, X., Rabinstein, A.A., Erickson, B.J., Kapa, S. and Friedman, P.A. (2019). An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. The Lancet, [online] 394(10201), pp.861–867. Available at: 10.1016/S0140-6736(19)31721-0 (Accessed: 27 December 2022)

Arora, A. (2020). Conceptualising Artificial Intelligence as a Digital Healthcare Innovation: An Introductory Review. *Medical Devices: Evidence and Research*, [online] Volume 13, pp.223–230. doi:10.2147/mder.s262590 (Accessed: 27 December 2022)

Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices. (2021). *FDA.* [online] Available at: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices (Accessed: 27 December 2022)

Barzescu, C. (2020). *How do users' perceptions of AI influence adoption into value-*

based radiology departments? [online] Aidence. Available at: https://www.aidence.com/articles/perceptions-ai-adoption/ (Accessed 5 September 2022)

Beck, T. (2022). *Clinical implementation of AI in radiology: how to face the challenges*. [online] www.quantib.com. Available at: https://www.quantib.com/blog/clinical-implementation-of-ai-in-radiology/how-to-face-the-challenges?utm_medium=email&_hsmi=224379252&_hsenc=p2ANqtz--U603-tgaipe-MFL8sQT5HgEW-Lvaoln6_okDWjf9b6MhYYrCA9PxmkmnWx9aPxXxRW8M37-ihLJE6UmVBY_PAPyLxxw&utm_content=224379252&utm_source=hs_email (Accessed 7 Sep. 2022)

Belenguer, L. (2022). AI bias: exploring discriminatory algorithmic decision-making models and the application of possible machine-centric solutions adapted from the pharmaceutical industry. [online] *AI and Ethics*. Available at:10.1007/s43681-022-00138-8 (Accessed: 27 December 2022)

Bell, D., Gachuhi, N. and Assefi, N. (2017). Dynamic Clinical Algorithms: Digital Technology Can Transform Health Care Decision-Making. *The American Journal of Tropical Medicine and Hygiene*, [online] 98(1), pp.9–14. Available at:10.4269/ajtmh.17-0477(Accessed: 27 December 2022)

Benjamens, S., Dhunnoo, P. and Meskó, B. (2020). The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *npj Digital Medicine*, [online] 3(1), pp.1–8. Available at:10.1038/s41746-020-00324-0 (Accessed: 27 December 2022)

Bjurlin, M.A., Carroll, P.R., Eggener, S., Fulgham, P.F., Margolis, D.J., Pinto, P.A., Rosenkrantz, A.B., Rubenstein, J.N., Rukstalis, D.B., Taneja, S.S. and Turkbey, B. (2020). Update of the Standard Operating Procedure on the Use of Multiparametric Magnetic Resonance Imaging for the Diagnosis, Staging and Management of Prostate Cancer. *Journal of Urology*, [online] 203(4), pp.706–712. Available at 10.1097/ju.0000000000000617(Accessed: 27 December 2022)

Brendan, M.H., Moore, E., Ramage, D., Hampson, S. and Arcas, Blaise Agüera y (2016). *Communication-Efficient Learning of Deep Networks from Decentralized Data*. [online] arXiv.org. Available at: https://arxiv.org/abs/1602.05629 (Accessed: 27 December 2022)

Caliskan, A., Bryson, J.J. and Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, [online] 356(6334), pp.183–186. Available at:10.1126/science.aal4230 (Accessed: 27 December 2022)

Cath, C. (2018). Governing artificial intelligence: ethical, legal and technical opportunities and challenges. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, [online] 376(2133), p.20180080. Available at:10.1098/rsta.2018.0080 (Accessed: 27 December 2022)

Char, D.S., Shah, N.H. and Magnus, D. (2018). Implementing Machine Learning in Health Care — Addressing Ethical Challenges. *New England Journal of Medicine*, [online] 378(11), pp.981–983. Available at:10.1056/nejmp1714229(Accessed: 27 December 2022)

Ching, T., Himmelstein, D.S., Beaulieu-Jones, B.K., Kalinin, A.A., Do, B.T., Way, G.P., Ferrero, E., Agapow, P.-M., Zietz, M., Hoffman, M.M., Xie, W., Rosen, G.L., Lengerich, B.J., Israeli, J., Lanchantin, J., Woloszynek, S., Carpenter, A.E., Shrikumar, A., Xu, J. and Cofer, E.M. (2018). Opportunities and obstacles for deep learning in biology and medicine. *Journal of The Royal Society Interface*, [online] 15(141), p.20170387. Available at:10.1098/rsif.2017.0387(Accessed: 27 December 2022)

Chen, Irene, Fredrik D. Johansson, and David Sontag. "Why Is My Classifier Discriminatory?" ArXiv:1805.12002 [Cs, Stat], [online]. Available at: http://arxiv.org/abs/1805.12002 (Accessed: 27 December 2022)

Chew, H.S.J. and Achananuparp, P. (2022). Perceptions and Needs of Artificial

Intelligence in Health Care to Increase Adoption: Scoping Review. *Journal of Medical Internet Research*, [online] 24(1), p.e32939. Available at:10.2196/32939. (Accessed: 27 December 2022)

Chowdhury, A., Kassem, H., Padoy, N., Umeton, R. and Karargyris, A. (2022). A Review of Medical Federated Learning: Applications in Oncology and Cancer Research. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pp.3–24. Available at:10.1007/978-3-031-08999-2_1(Accessed: 27 December 2022)

Choy, G., Khalilzadeh, O., Michalski, M., Do, S., Samir, A.E., Pianykh, O.S., Geis, J.R., Pandharipande, P.V., Brink, J.A. and Dreyer, K.J. (2018). Current Applications and Future Impact of Machine Learning in Radiology. *Radiology*, [online] 288(2), pp.318–328. Available at:10.1148/radiol.2018171820 (Accessed: 27 December 2022)

Coombs, C., Hislop, D., Taneva, S.K. and Barnard, S. (2020). The strategic impacts of Intelligent Automation for knowledge and service work: An interdisciplinary review. *The Journal of Strategic Information Systems*, [online] 29(4), p.101600. Available at:10.1016/j.jsis.2020.101600 (Accessed: 27 December 2022)

Dawson, D., Schleiger, E., Horton, J., McLaughlin, J., Robinson, C., Quezada, G., Scowcroft, J. and Hajkowicz, S. (2019). Artificial Intelligence: Australia's ethics framework. *apo.org.au*. [online] Available at: https://apo.org.au/node/229596 (Accessed 6 October 2021)

Davis SE, Greevy RA, Fonnesbeck C, Lasko TA, Walsh CG, Matheny ME. A nonparametric updating method to correct clinical prediction model drift. J Am Med Inform Assoc. 2019 [online] Available at: https://doi.org/10.1093/jamia/ocz127 (Accessed 6 October 2021)

Dayan, I., Roth, H.R., Zhong, A., Harouni, A., Gentili, A., Abidin, A.Z., Liu, A., Costa, A.B., Wood, B.J., Tsai, C.-S., Wang, C.-H., Hsu, C.-N., Lee, C.K., Ruan, P., Xu, D., Wu, D., Huang, E., Kitamura, F.C., Lacey, G. and de Antônio Corradi, G.C. (2021).

Federated learning for predicting clinical outcomes in patients with COVID-19. *Nature Medicine*, [online] pp.1–9. Available at:10.1038/s41591-021-01506-3 (Accessed 6 October 2021)

Daye, D., Wiggins, W.F., Lungren, M.P., Alkasab, T., Kottler, N., Allen, B., Roth, C.J., Bizzo, B.C., Durniak, K., Brink, J.A., Larson, D.B., Dreyer, K.J. and Langlotz, C.P. (2022). Implementation of Clinical Artificial Intelligence in Radiology: Who Decides and How? *Radiology*, [online] p.212151. Available at:10.1148/radiol.212151(Accessed 6 October 2021)

DREYER, K.J. and COOMBS, L.P. (2020). *The AI Lifecycle Evaluating And Monitoring Ai Algorithms*. [online] Available at: https://www.fda.gov/media/135715/download (Accessed 13 December 2022)

Dwivedi, Y.K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P.V., Janssen, M., Jones, P., Kar, A.K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B. and Medaglia, R. (2019). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, [online] 57. Available at:10.1016/j.ijinfomgt.2019.08.002 (Accessed 6 October 2021)

Esmaeilzadeh, P. (2018). The Effects of Public Concern for Information Privacy on the Adoption of Health Information Exchanges (HIEs) by Healthcare Entities. *Health Communication*, [online] 34(10), pp.1202–1211. Available at:10.1080/10410236.2018.1471336 (Accessed 6 October 2021)

Ebrahimian, S., Kalra, M.K., Agarwal, S., Bizzo, B.C., Elkholy, M., Wald, C., Allen, B. and Dreyer, K.J. (2022). FDA-regulated AI Algorithms: Trends, Strengths, and Gaps of Validation Studies. *Academic Radiology*, 29(4), pp.559–566. Available at:10.1016/j.acra.2021.09.002 (Accessed 6 October 2021)

Esmaeilzadeh, P. (2020). Use of AI-based tools for healthcare purposes: a survey study from consumers' perspectives. *BMC Medical Informatics and Decision Making*, [online] 20(1). Available at:10.1186/s12911-020-01191-1(Accessed 6 October 2021)

Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., Liu, Y., Topol, E., Dean, J. and Socher, R. (2021). Deep learning-enabled medical computer vision. *npj Digital Medicine*, [online] 4(1). Available at:10.1038/s41746-020-00376-2 (Accessed 6 October 2021)

Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M. and Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), pp.115–118. Available at:10.1038/nature21056 (Accessed 6 October 2021)

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S. and Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*, 25(1), pp.24–29. Available at:10.1038/s41591-018-0316-z (Accessed 6 October 2021)

European Commission . (2021). *Ethics guidelines for trustworthy AI | Shaping Europe's digital future*. [online] Available at: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (Accessed 6 October 2021)

Evans, J. and Mathur, A. (2018). The Value of Online Surveys Retailing and Retail Management View project Value Chain and Relationships View project. *Emerald Insight*, [online] 28(4). Available at: 10.1108/10662240510590360 (Accessed 6 October 2021)

Giansanti, D. (2022). The Regulation of Artificial Intelligence in Digital Radiology in the Scientific Literature: A Narrative Review of Reviews. *Healthcare*, [online] 10(10), p.1824. Available at:10.3390/healthcare10101824 (Accessed 6 October 2021)

Google (2019). *Google Forms: Free Online Surveys for Personal Use*. [online]

Google.com. Available at: https://www.google.com/forms/about/ (Accessed 6 October. 2021)

Griffith, B., Kadom, N. and Straus, C.M. (2019). Radiology Education in the 21st Century: Threats and Opportunities. *Journal of the American College of Radiology*, [online], 16(10), pp.1482–1487. Available at:10.1016/j.jacr.2019.04.003 (Accessed 6 October 2021)

Gunčar, G., Kukar, M., Notar, M., Brvar, M., Černelč, P., Notar, M. and Notar, M. (2018). An application of machine learning to haematological diagnosis. *Scientific Reports*, [online], 8(1). Available at:10.1038/s41598-017-18564-8(Accessed 6 October 2021)

Hägele, M., Seegerer, P., Lapuschkin, S., Bockmayr, M., Samek, W., Klauschen, F., Müller, K.-R. and Binder, A. (2020). Resolving challenges in deep learning-based analyses of histopathological images using explanation methods. *Scientific Reports*, 10(1). [online], Available at:10.1038/s41598-020-62724-2 (Accessed 6 October 2021)

Health Sciences Authority (2019). *Regulatory Guidelines for Software Medical Devices - A Lifecycle Approach*. [online], Available at: https://www.hsa.gov.sg/docs/default-source/announcements/regulatory-updates/regulatory-guidelines-for-software-medical-devices--a-lifecycle-approach.pdf (Accessed 13 December 2022)

Hengstler, M., Enkel, E. and Duelli, S. (2016). Applied artificial intelligence and trust—The case of autonomous vehicles and medical assistance devices. *Technological Forecasting and Social Change*, [online] 105(C), pp.105–120. Available at: https://econpapers.repec.org/article/eeetefoso/v_3a105_3ay_3a2016_3ai_3ac_3ap_3a105-120.htm (Accessed 6 October 2021)

Holstein, K., Wortman Vaughan, J., Daumé III, H., Dudik, M., & Wallach, H. (2019, April). Improving fairness in machine learning systems: What do industry practitioners need?. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (p. 600). [online], ACM. Available at: https://arxiv.org/abs/1812.05239 (Accessed 6 October 2021)

Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L.H. and Aerts, H.J.W.L. (2018). Artificial intelligence in radiology. *Nature Reviews Cancer*, [online] 18(8), pp.500–510. Available at: 10.1038/s41568-018-0016-5 (Accessed 6 October 2021)

Howard, C. (2019). *Advantages and Disadvantages of Online Surveys | Cvent Blog*. [online] Cvent.com. Available at: https://www.cvent.com/en/blog/events/advantages-disadvantages-online-surveys (Accessed 1 August 2022)

Hu, K., Li, Y., Xia, M., Wu, J., Lu, M., Zhang, S. and Weng, L. (2021). Federated Learning: A Distributed Shared Machine Learning Method. *Complexity*, 2021, pp.1–20. Available at: 10.1155/2021/8261663(Accessed 1 August 2022)

Hyman, M. and Sierra, J. (2016). Statistical methods marketing research consultants might use (Part 1). *NM State University Business Outlook*, [online] 14(11). Available at: https://www.researchgate.net/publication/283350875_Statistical_methods_marketing_research_consultants_might_use_Part_1 (Accessed 1 December 2021)

ITU (2019). *New ITU data reveal growing Internet uptake but a widening digital gender divide*. [online] Available at: https://www.itu.int/en/mediacentre/Pages/2019-PR19.aspx (Accessed 5 September 2022)

Kaissis, G.A., Makowski, M.R., Rückert, D. and Braren, R.F. (2020). Secure, privacy-preserving and federated machine learning in medical imaging. *Nature Machine Intelligence*, [online] 2(6), pp.305–311. Available at: 10.1038/s42256-020-0186-1(Accessed 6 October 2021)

Kashani, K., Barracca, A., Contini, M., Ledda, S., Mancosu, G., Pintore, G. and Ronco, C. (2020). Evolution of Clinical Medicine: From Expert Opinion to Artificial Intelligence. *Journal of Translational Critical Care Medicine*, [online] 2(4), p.78. Available at::10.4103/jtccm.jtccm_6_21(Accessed 6 October 2021)

Kelly, C.J., Karthikesalingam, A., Suleyman, M., Corrado, G. and King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*, [online] 17(1). doi:10.1186/s12916-019-1426-2(Accessed 6 October 2021)

Kumar, A., Finley, B., Braud, B., Tarkoma, S. and Hui, P. (n.d.). *Marketplace for AI Models*. [online] Available at: https://arxiv.org/pdf/2003.01593.pdf (Accessed 6 October 2021)

Laï, M.-C. ., Brian, M. and Mamzer, M.-F. . (2020). Perceptions of artificial intelligence in healthcare: findings from a qualitative survey study among actors in France. *Journal of Translational Medicine*, [online] 18(1). Available at:10.1186/s12967-019-02204-y (Accessed 6 October 2021)

Lee, J.-G., Kim, K.J., Lee, S. and Shin, D.-H. (2015). Can Autonomous Vehicles Be Safe and Trustworthy? Effects of Appearance and Autonomy of Unmanned Driving Systems. *International Journal of Human-Computer Interaction*, [online] 31(10), pp.682–691. Available at:10.1080/10447318.2015.1070547(Accessed 6 October 2021)

Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B. and Sánchez, C.I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, [online] 42, pp.60–88. Available at:10.1016/j.media.2017.07.005 (Accessed 6 October 2021)

Linardos, A., Kushibar, K., Walsh, S., Gkontra, P. and Lekadir, K. (2021). Federated Learning for Multi-Center Imaging Diagnostics: A Study in Cardiovascular Disease. *arXiv:2107.03901 [cs, eess]*. [online] Available at: https://arxiv.org/abs/2107.03901 (Accessed 6 October 2021)

Liu, Y., Zhang, L., Ge, N. and Li, G. (2020). *A Systematic Literature Review on Federated Learning: From A Model Quality Perspective*. [online] Available at: https://arxiv.org/pdf/2012.01973.pdf (Accessed 19 January 2021)

Liu, Q., Chen, C., Qin, J., Dou, Q. and Heng, P.-A. (2021). FedDG: Federated Domain

Generalization on Medical Image Segmentation via Episodic Learning in Continuous Frequency Space. *arXiv:2103.06030 [cs]*. [online] Available at: https://arxiv.org/abs/2103.06030 (Accessed 6 October 2021)

Lockey, S., Gillespie, N., Holm, D. and Someh, I.A. (2021). A Review of Trust in Artificial Intelligence: Challenges, Vulnerabilities and Future Directions. *Proceedings of the 54th Hawaii International Conference on System Sciences*. [online] Available at:10.24251/hicss.2021.664 (Accessed 5 May 2022)

Malwade, P.K. (2022). Doctoral Research Survey: Federated Machine Learning Algorithm in AI-based tools in Radiology Medical Imaging - Association of user perceived Trustworthiness and Robustness with their adoption intent. [online] Google Docs. Available at: https://forms.gle/4du7oGqF1GRCHLZh8 (Accessed 5 May 2022)

Markets, R. and (2022). *AI In Medical Imaging Market Global Report 2022: Growing Inclination of Radiologists Toward AI-Based Medical Imaging Presents Opportunities*. [online] GlobeNewswire News Room. Available at: https://www.globenewswire.com/en/news-release/2022/07/06/2474675/28124/en/AI-In-Medical-Imaging-Market-Global-Report-2022-Growing-Inclination-of-Radiologists-Toward-AI-Based-Medical-Imaging-Presents-Opportunities.html (Accessed 5 September 2022)

Masud, R., Al-Rei, M. and Lokker, C. (2019). Computer-Aided Detection for Breast Cancer Screening in Clinical Settings: Scoping Review. *JMIR Medical Informatics*, [online] 7(3), p.e12660. Available at:10.2196/12660 (Accessed 5 September 2022)

Mauskopf, J.A., Sullivan, S.D., Annemans, L., Caro, J., Mullins, C.D., Nuijten, M., Orlewska, E., Watkins, J. and Trueman, P. (2007). Principles of Good Practice for Budget Impact Analysis: Report of the ISPOR Task Force on Good Research Practices—Budget Impact Analysis. *Value in Health*, [online] 10(5), pp.336–347. Available at:10.1111/j.1524-4733.2007.00187.x (Accessed 5 September 2022)

Mazurowski, M.A., Clark, K., Czarnek, N.M., Shamsesfandabadi, P., Peters, K.B. and Saha, A. (2017). Radiogenomics of lower-grade glioma: algorithmically-assessed tumor shape is associated with tumor genomic subtypes and patient outcomes in a multi-institutional study with The Cancer Genome Atlas data. *Journal of Neuro-Oncology*, [online] 133(1), pp.27–35. Available at:10.1007/s11060-017-2420-1(Accessed 5 September 2022)

Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.-A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N. and Corso, J.J. (2015). The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging*, [online] 34(10), pp.1993–2024. Available at:10.1109/tmi.2014.2377694 (Accessed 5 September 2022)

McKinney, S.M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., Back, T., Chesus, M., Corrado, G.C., Darzi, A., Etemadi, M., Garcia-Vicente, F., Gilbert, F.J., Halling-Brown, M., Hassabis, D., Jansen, S., Karthikesalingam, A., Kelly, C.J., King, D. and Ledsam, J.R. (2020). International evaluation of an AI system for breast cancer screening. *Nature*, [online] 577(7788), pp.89–94. Available at:10.1038/s41586-019-1799-6 (Accessed 5 September 2022)

McNair, J. (2021). An Analysis of the Impact of an Influencer's Perceived Trustworthiness and Content Quality on Their Follower's Travel Intent. *Doctor of Business Administration (DBA)*. [online] Available at: https://digitalcommons.georgefox.edu/dbadmin/41/ (Accessed 5 January 2022)

Meskó, B. and Görög, M. (2020). A short guide for medical professionals in the era of artificial intelligence. *npj Digital Medicine*, [online] 3(1). Available at:10.1038/s41746-020-00333-z (Accessed 5 January 2022)

Muehlematter, U.J., Daniore, P. and Vokinger, K.N. (2021). Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–

20): a comparative analysis. *The Lancet Digital Health*, [online] 3(3), pp.e195–e203. Available at:10.1016/S2589-7500(20)30292-2(Accessed 5 January 2022)

Naseri, M., Hayes, J. and De Cristofaro, E. (2021). Toward Robustness and Privacy in Federated Learning: Experimenting with Local and Central Differential Privacy. *arXiv:2009.03561 [cs]*. [online] Available at: https://arxiv.org/abs/2009.03561 (Accessed 6 October 2021)

Nestor B, McDermott MBA, Chauhan G, Naumann T, Hughes MC, Goldenberg A, et al. Rethinking clinical prediction: why machine learning must consider year of care and feature aggregation. In: Machine Learning for Health (ML4H): NeurIPS; 2018. [online] Available at: https://arxiv.org/abs/1811.12583 (Accessed 1 May 2022)

Ng, D., Lan, X., Yao, M.M.-S., Chan, W.P. and Feng, M. (2021). Federated learning: a collaborative effort to achieve better medical imaging models for individual sites that have small labelled datasets. *Quantitative Imaging in Medicine and Surgery*, [online] 11(2), pp.852–857 Available at:10.21037/qims-20-595(Accessed 1 May 2022)

Obermeyer, Z., & Lee, T. H. (2017). Lost in thought: The limits of the human mind and the future of medicine. The New England journal of medicine, 377(13), 1209. [online] Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5754014 (Accessed 1 May 2022)

Pianykh, O.S., Langs, G., Dewey, M., Enzmann, D.R., Herold, C.J., Schoenberg, S.O. and Brink, J.A. (2020). Continuous Learning AI in Radiology: Implementation Principles and Early Applications. *Radiology*, 297(1), [online] pp.6–14. Available at: 10.1148/radiol.2020200038 (Accessed 1 May 2022)

Prasad, Dr.K. and Vaidya, Dr.R.W. (2020). Perceived Psychological Wellbeing of an

Academician in Higher Education is a function of Organizational Support, Supervisor Support, Family Support and Social Support during Covid-19 Pandemic. *International Journal of Management and Business*, [online] 11(7), pp.20–34. Available at:10.34218/IJM.11.7.2020.003 (Accessed 1 May 2022)

Pupic, N., Ghaffarizadeh, S.A., Forster, B.B. and Darras, K.E. (2022). Artificial Intelligence: The Medium and The Message. *Canadian Association of Radiologists Journal*, [online] p.084653712211145. Available at:10.1177/08465371221114590 (Accessed 1 May 2022)

Qayyum, A., Ahmad, K., Ahsan, M.A., Al-Fuqaha, A. and Qadir, J. (2021). Collaborative Federated Learning For Healthcare: Multi-Modal COVID-19 Diagnosis at the Edge. *arXiv:2101.07511 [cs]*. [online] Available at: https://arxiv.org/abs/2101.07511 (Accessed 6 October 2021)

Qian, Y., Hu, L., Chen, J., Guan, X., Hassan, M.M. and Alelaiwi, A. (2019). Privacy-aware service placement for mobile edge computing via federated learning. *Information Sciences*, 505, pp.562–570. [online] Available at: doi:10.1016/j.ins.2019.07.069. (Accessed 27 August 2022)

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347-1358. [online] Available at: https://pubmed.ncbi.nlm.nih.gov/30943338/ (Accessed 27 December 2022)

Ribeiro M, Singh S, Guestrin C. “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations; 2016. [online] Available at: https://doi.org/10.18653/v1/n16-3020 (Accessed 27 December 2022)

Rimmer, A. (2017). Radiologist shortage leaves patient care at risk, warns royal college. *BMJ*, [online] p.j4683. Available at:10.1136/bmj.j4683(Accessed 27 December 2022)

Reddy, S., Allan, S., Coghlan, S. and Cooper, P. (2019). A governance model for the application of AI in health care. *Journal of the American Medical Informatics Association*, [online] 27(3). Available at:10.1093/jamia/ocz192(Accessed 27 December 2022)

Reddy, S., Fox, J. and Purohit, M.P. (2021). Artificial intelligence-enabled healthcare delivery. *Journal of the Royal Society of Medicine*, [online] 112(1), pp.22–28. Available at:10.1177/0141076818815510 (Accessed 27 December 2022)

Rieke, N., Hancox, J., Li, W., Milletarì, F., Roth, H.R., Albarqouni, S., Bakas, S., Galtier, M.N., Landman, B.A., Maier-Hein, K., Ourselin, S., Sheller, M., Summers, R.M., Trask, A., Xu, D., Baust, M. and Cardoso, M.J. (2020). The future of digital health with federated learning. *npj Digital Medicine*, [online] 3(1), pp.1–7. Available at:10.1038/s41746-020-00323-1 (Accessed 27 December 2022)

Roth, H.R., Chang, K., Singh, P., Neumark, N., Li, W., Gupta, V., Gupta, S., Qu, L., Ihsani, A., Bizzo, B.C., Wen, Y., Buch, V., Shah, M., Kitamura, F., Mendonça, M., Lavor, V., Harouni, A., Compas, C., Tetreault, J. and Dogra, P. (2020). Federated Learning for Breast Density Classification: A Real-World Implementation. *arXiv:2009.01871 [cs, eess]*, [online] 12444, pp.181–191. Available at:10.1007/978-3-030-60548-3_18 (Accessed 27 December 2022)

Romero-Brufau, S., Wyatt, K.D., Boyum, P., Mickelson, M., Moore, M. and Cognetta-Rieke, C. (2020). A lesson in implementation: A pre-post study of providers' experience with artificial intelligence-based clinical decision support. *International Journal of Medical Informatics*, [online] 137, p.104072. Available at:10.1016/j.ijmedinf.2019.104072 (Accessed 27 December 2022)

Rowthorn, R.E. (1975). What Remains of Kaldor's Law? *The Economic Journal*, [online] 85(337), p.10. Available at:10.2307/2230525 (Accessed 27 December 2022)

Rylands-Monk, F. (2022). *Niessen puts emphasis on need for trust and collaboration in AI*. [online] AuntMinnieEurope.com. Available at: https://www.auntminnieeurope.com/index.aspx?sec=log&URL=https%3a%2f%2fwww.auntminnieeurope.com%2findex.aspx%3fsec%3dsup%26sub%3dpac%26pag%3ddis%26itemId%3d621060 (Accessed 5 September 2022)

Safi, S., Thiessen, T. and Schmailzl, K.J. (2018). Acceptance and Resistance of New Digital Technologies in Medicine: Qualitative Study. *JMIR Research Protocols*, [online] 7(12), p.e11072. Available at:10.2196/11072 (Accessed 5 September 2022)

Salam, M. A., Taha, S., & Ramadan, M. (2021). COVID-19 detection using federated machine learning. *PLoS ONE*, *16*(6 June), [online]. Available at https://doi.org/10.1371/journal.pone.0252573(Accessed 5 September 2022)

Samek, W., Wiegand, T. and Müller, K.-R. (2017). *Explainable artificial intelligence: understanding, visualizing and interpreting deep learning models*. [online] Available at: https://www.itu.int/dms_pub/itu-s/opb/journal/S-JOURNAL-ICTF.VOL1-2018-1-P05-PDF-E.pdf (Accessed 5 September 2022)

Santomartino, S.M. and Yi, P.H. (2022). Systematic Review of Radiologist and Medical Student Attitudes on the Role and Impact of AI in Radiology. *Academic Radiology*. [online] Available at: 10.1016/j.acra.2021.12.032 (Accessed 5 September 2022)

Sarma, K.V., Harmon, S., Sanford, T., Roth, H.R., Xu, Z., Tetreault, J., Xu, D., Flores, M.G., Raman, A.G., Kulkarni, R., Wood, B.J., Choyke, P.L., Priester, A.M., Marks, L.S., Raman, S.S., Enzmann, D., Turkbey, B., Speier, W. and Arnold, C.W. (2021). Federated learning improves site performance in multicenter deep learning without data sharing. *Journal of the American Medical Informatics Association: JAMIA*, [online] 28(6), pp.1259–1264. Available at:10.1093/jamia/ocaa341 (Accessed 5 September 2022)

Scherer, J., Nolden, M., Kleesiek, J., Metzger, J., Kades, K., Schneider, V., Bach, M., Sedlaczek, O., Bucher, A.M., Vogl, T.J., Grünwald, F., Kühn, J.-P., Hoffmann, R.-T., Kotzerke, J., Bethge, O., Schimmöller, L., Antoch, G., Müller, H.-W., Daul, A. and Nikolaou, K. (2020). Joint Imaging Platform for Federated Clinical Data Analytics. *JCO clinical cancer informatics*, [online] 4, pp.1027–1038. Available at:10.1200/CCI.20.00045(Accessed 5 September 2022)

Shah, U., Dave, I., Malde, J., Mehta, J. and Kodeboyina, S. (2021). *Maintaining Privacy in Medical Imaging with Federated Learning, Deep Learning, Differential Privacy, and Encrypted Computation*. [online] IEEE Xplore. Available at: https://ieeexplore.ieee.org/document/9417997/ (Accessed 27 December 2022)

Sheller, M.J., Edwards, B., Reina, G.A., Martin, J., Pati, S., Kotrotsou, A., Milchenko, M., Xu, W., Marcus, D., Colen, R.R. and Bakas, S. (2020). Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Scientific Reports*, [online] 10(1), p.12598. Available at:10.1038/s41598-020-69250-1(Accessed 27 December 2022)

Sheller, M.J., Reina, G.A., Edwards, B., Martin, J. and Bakas, S. (2019b). Multi-institutional Deep Learning Modeling Without Sharing Patient Data: A Feasibility Study on Brain Tumor Segmentation. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, [online] pp.92–104. Available at:10.1007/978-3-030-11723-8_9(Accessed 27 December 2022)

Sohn, K. and Kwon, O. (2020). Technology acceptance theories and factors influencing artificial Intelligence-based intelligent products. *Telematics and Informatics*, [online] 47, p.101324. Available at:10.1016/j.tele.2019.101324 (Accessed 27 December 2022)

Statista Research Department (2022). *MRI scan examinations in the Netherlands 2008-2019*. [online] Statista. Available at: https://www.statista.com/statistics/609151/magnetic-resonance-imaging-scan-examinations-in-the-netherlands [Accessed 5 Sep. 2022].

Steyerberg, E. W. (2019). Clinical prediction models. Springer International Publishing [online] Available at: https://link.springer.com/chapter/10.1007/978-3-030-16399-0_17(Accessed 27 December 2022)

Stripelis, D., Ambite, J.L., Lam, P. and Thompson, P. (2021). Scaling Neuroscience Research using Federated Learning. *arXiv:2102.08440 [cs]*. [online] Available at: https://arxiv.org/abs/2102.08440 (Accessed 6 October 2021)

Sullivan, G.M. and Artino, A.R. (2013). Analyzing and Interpreting Data From Likert-Type Scales. *Journal of Graduate Medical Education*, [online] 5(4), pp.541–542. Available at: 10.4300/jgme-5-4-18(Accessed 6 October 2021)

Sullivan, G.M. and Feinn, R. (2012). Using effect size—or why the P value is not enough. *Journal of Graduate Medical Education*, [online] 4(3), pp.279–282. Available at: 10.4300/jgme-d-12-00156.1(Accessed 6 October 2021)

Sun, T.Q. and Medaglia, R. (2019). Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. *Government Information Quarterly*, [online] 36(2), pp.368–383. Available at: 10.1016/j.giq.2018.09.008(Accessed 6 October 2021)

Turja, T., Aaltonen, I., Taipale, S. and Oksanen, A. (2020). Robot acceptance model for care (RAM-care): A principled approach to the intention to use care robots. *Information & Management*, [online] 57(5), p.103220. Available at:10.1016/j.im.2019.103220 (Accessed 6 October 2021)

United Nations Development Programme (2015). *Sustainable Development Goals*. [online] Sustainable Development Goals. Available at: https://www.undp.org/sustainable-development-goals (Accessed 5 September 2022)

U.S. Food & Drug Administration (2021). Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices. *FDA.* [online] Available at:

https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical devices?utm_medium=email&utm_source=govdelivery (Accessed 11 October 2022)

Vayena, E., Blasimme, A. and Cohen, I.G. (2018). Machine learning in medicine: Addressing ethical challenges. *PLOS Medicine*, [online] 15(11), p.e1002689. Available at: 10.1371/journal.pmed.1002689 (Accessed 5 September 2022)

Venugopal, V.K., Takhar, R., Gupta, S. and Mahajan, V. (2022). Clinical Explainability Failure (CEF) & Explainability Failure Ratio (EFR) – Changing the Way We Validate Classification Algorithms. *Journal of Medical Systems*, [online] 46(4). Available at: 10.1007/s10916-022-01806-2(Accessed 5 September 2022)

Wernick, M., Yang, Y., Brankov, J., Yourganov, G. and Strother, S. (2010). Machine Learning in Medical Imaging. *IEEE Signal Processing Magazine*, [online] 27(4), pp.25–38. Available at:10.1109/msp.2010.936730 (Accessed 5 September 2022)

Whittlestone, J., Nyrup, R., Alexandrova, A., Dihal, K. and Cave, S. (n.d.). *Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research B.* [online] Available at: https://www.nuffieldfoundation.org/sites/default/files/files/Ethical-and-Societal-Implications-of-Data-and-AI-report-Nuffield-Foundat.pdf (Accessed 5 September 2022)

Wooldridge, J.M. (2000). *Introductory Econometrics. A modern approach*. 5th ed. [online] Cincinnati: South-Western Cengage Learning, p.300. Available at: https://economics.ut.ac.ir/documents/3030266/14100645/Jeffrey_M._Wooldridge_Introductory_Econometrics_A_Modern_Approach_2012.pdf (Accessed 22 April 2022)

www.oecd-ilibrary.org. (2021). *Home*. [online] Available at: https://www.oecd-ilibrary.org/sites/ed023875-en/index.html?itemId=/content/component/ed023875-en (Accessed 5 September 2022)

Xie, J., Jin, C., Liu, M., Sun, K., Jin, Z., Ding, Z. and Gong, X. (2022). MRI/Transrectal Ultrasound Fusion-Guided Targeted Biopsy and Transrectal Ultrasound-Guided Systematic Biopsy for Diagnosis of Prostate Cancer: A Systematic Review and Meta-analysis. *Frontiers in Oncology*, [online] 12. Available at:10.3389/fonc.2022.880336 (Accessed 5 September 2022)

Xu, J., Glicksberg, B.S., Su, C., Walker, P., Bian, J. and Wang, F. (2020). Federated Learning for Healthcare Informatics. *Journal of Healthcare Informatics Research*. [online] Available at:10.1007/s41666-020-00082-4 (Accessed 5 September 2022)

Yang, D., Xu, Z., Li, W., Myronenko, A., Roth, H.R., Harmon, S., Xu, S., Turkbey, B., Turkbey, E., Wang, X., Zhu, W., Carrafiello, G., Patella, F., Cariati, M., Obinata, H., Mori, H., Tamura, K., An, P., Wood, B.J. and Xu, D. (2020). Federated Semi-Supervised Learning for COVID Region Segmentation in Chest CT using Multi-National Data from China, Italy, Japan. *arXiv:2011.11750 [cs, eess]*. [online] Available at: https://arxiv.org/abs/2011.11750 (Accessed 6 October 2021)

Zech, J.R., Badgeley, M.A., Liu, M., Costa, A.B., Titano, J.J. and Oermann, E.K. (2018). Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLOS Medicine*, [online] 15(11), p.e1002683. doi:10.1371/journal.pmed.1002683 (Accessed 6 October 2021)