Miina Schildt

Ethics of Using AI in Healthcare Diagnostics

- a scoping review

Metropolia University of Applied Sciences

Master of Business Administration

Health Business Management

Thesis

13 April 2022

Abstract

Author(s) Title Number of Pages Date	Miina Schildt Ethics of Using AI in Healthcare Diagnostics - a scoping review 48 pages + 2 appendices 13 April 2022
Degree	Master of Business Administration
Degree Programme	Master's Degree Programme in Health Business Management
Specialisation option	Ethical AI
Instructor	Docent, PhD, Principal Lecturer Eija Metsälä

Introduction: Diagnostic AI systems are providing promising results with accurate output, and they offer a great possibility to ease the problems of insufficient human and financial resources. Diagnostic AI systems use sensitive health data, so they carry a high risk of violating fundamental human rights. For that, it is crucial to update the outdated ethical regulations and guidelines for the field of AI. This scoping review aims to define the main ethical aspects of using AI in healthcare diagnostics.

Methods: Literature searches were carried out in databases ProQuest Central, Science Direct and PubMed to find all relevant research published on ethics of diagnostic AI. JBI´s three-step search strategy recommended for scoping reviews was being followed. It included the pilot search, the actual search, and the analysis of the results first on title-level, next on abstract-level, then on full-text level. In addition, more sources were searched in the reference lists of the chosen articles.

Results: Following a systematic search process, 12 articles were included into this review. The inductive content analysis was used to analyse the articles, and the ethical issues recognised from the articles, were organised under the four ethical principles of trustworthy AI defined by European Commission (2019): respect for human autonomy, prevention of harm, fairness, and explicability.

Conclusion: Ethical diagnostic AI system consists of deep understanding and consideration of ethical issues around it from the point of view of all active stakeholders. In practice this requires ethical discussion to be a fixed and coordinated as a part of the development process of diagnostic AI system, involving system developers, healthcare professionals and experts on AI ethics.

Recommendation for future research: One major finding of this review was that empirical research regarding the topic is non-existent. The author recommends empirical research to be done on how governance of AI ethics is implemented in companies that develop diagnostic AI systems; and then, assessing the effectiveness of these measures from the end-user point of view.

Keywords	Ethics of diagnostic AI, ethics of medical AI

Contents

Appendices

1 Introduction

This thesis is a scoping literature review about the main ethical issues with reference to using artificial intelligence (AI) in healthcare diagnostics. The topic concerns several stakeholders, such as the developers of the technology, healthcare organizations who purchase the systems, healthcare professionals who work with AI-driven diagnostic systems, as well as patients who get treated with the assistance of the system, and even society and future generations in the bigger picture.

Service and user experience (UX) designs are fixed part of contemporary development process of new systems, products, and services. To create something that truly adds value to the end users, their opinions, needs and experiences should be carefully studied, considered, and included in the development process. (Stickdorn, Hormess, Lawrence, & Schneider 2018: 14.) This also applies to AI systems used for healthcare diagnostics (Bitkina, Kim & Park 2020). The developers have the power to build the system and everything it contains, but they need the expertise of healthcare professionals to know what end users actually require from the system. Consequently, it is important that both these parties are aware of the ethical issues concerning the topic. (Keskinbora 2019.) The aim of this thesis is to provide this information to both system developers and healthcare professionals as end users of the systems, but the target group is not specified in the research question.

AI-based solutions used in healthcare diagnostics are getting more common and new solutions are being developed every day, which makes this issue very topical. Using AI systems for healthcare diagnostics generates great new possibilities to treat people better and to answer to the growing demand of healthcare services both in quantity and quality. AI technologies enable people to manage their own health in great detail. Also, they facilitate the prevention of health issues and prediction of both worsening of health condition and emerging of new diseases. They also help us to improve diagnosing and treatment. (Anom 2020; EIT Health 2019.)

The field of healthcare is constantly suffering from lack of resources (Topol 2019). The world population is growing, and the share of ageing population is increasing significantly. We are now able to treat an enormous number of diseases and other medical conditions, so that people can live longer and still have a good quality of life.

However, all this requires resources such as financial support and manpower, as well as facilities, medicines, and other medical equipment. (Mintzberg 2017:17-18.) Processing health data to make decisions takes time and there are a lot of people waiting for their health data to be analysed. Human capabilities to analyse data and make diagnoses is limited. That is why the shortage in number of physicians to make diagnosis creates a big challenge. (De Fauw et al. 2018). AI offers a great number of possibilities to address and find solutions to this dilemma, for example by significantly augmenting the human capabilities in diagnosis (Arieno, Chan & Destounis 2019; De Fauw et al. 2018). It is possible that improving the diagnostic process through the application of AI is one of the most promising areas of health innovation, and it has great potential to change the society and peoples´ lives for the better (Bartoletti 2019).

When new possibilities arise, new challenges emerge at the same time. Ethics has always played an extremely important role in healthcare, and it requires continuous consideration and orientation to the issue in many aspects of everyday work when dealing directly with life and death (Véliz 2019). The fast development of AI technology has overtaken the ethical guidelines, laws and regulations leaving them outdated (European Commission 2018: 8; Rigby 2019: 121). That is why it is crucial to take a deeper look into the ethical issues arising when using new technologies, like in this case using AI in healthcare diagnostics, and update the ethical guidelines.

The purpose of this thesis is to summarise the main ethical aspects of using AI in healthcare diagnostics according to the latest research for the benefit of the companies that develop AI solutions for healthcare diagnostics; for the healthcare institutions who purchase these systems; and for the healthcare staff who use them. The aim is to help these stakeholders to understand and be aware of the ethical issues, so that they can consider them during the whole life cycle of the system from planning and development to using it in practice and following and evaluating the effects continuously.

2 Theoretical background

AI systems can be used either to support physicians in diagnostics (Lee at el. 2021) or as autonomous systems that make clinical diagnostic decisions without human oversight (Abràmoff, Tobey, & Char 2020). Using the AI system to support diagnostics process is much more common than autonomous AI systems, but there is for example an AI system that provides a direct diagnostic recommendation for the point-of-care diagnosis of diabetic retinopathy without physician´s interpretation (Abràmoff et al. 2020). Normally physicians have the medical liability, but according to Abràmoff et al. (2020), in case of an autonomous AI system, the medical liability is on the creator of that system. However, the 2021 Coordinated Plan on Artificial Intelligence (European Commission 2021) underlines that the final decision should always involve human oversight.

The benefits of using AI in healthcare diagnostics are numerous and current systems are giving very promising results. When using AI in healthcare diagnostics the results are based on the data, so they are always evidence-based (Arieno et al. 2019). It has turned out that diagnoses produced by AI are as objective and correct or even better than diagnoses made by human professionals (Arieno et al. 2019; Bohr & Memarzadeh 2020; Miller & Brown 2018; McDougall 2018). For instance, the level of accuracy in detecting breast and skin cancer and some cardiovascular diseases by using AI technologies is very impressive (EIT Health 2019). When testing the ability to classify skin cancers, the AI system achieved the same level of performance as 21 expert dermatologists (McDougall 2019). AI systems also offer an important solution for the shortage of professionals who make diagnoses. Compared to a human healthcare professional, the AI system can read an enormous number of screenings and process data of numerous patients at the same time. This is something a human cannot do, and it explains unquestionably the benefits in efficiency. This is relevant as misdiagnoses, or delayed diagnoses are the most common forms of preventable harm in healthcare. Diagnostic errors can have severe consequences and lead to death or serious disability. (Newman-Toker, Schaffer, Yu-Moe, Nassery & Saber 2019.) For example, Geras, Mann and Moy (2019) claim that many breast cancers that could be detected in earlier stage are being missed due to lack of professional readers. Despite the impressive results of using AI technologies in healthcare diagnostics, no technology can replace the professionalism gained from years of experience, intellectual curiosity, and dedication that human healthcare practitioners have (Arieno et al. 2019). AI is not an autonomous operator and cannot or should not act as one, but it is a tool for a human and a machine to work in collaboration (Rusanen & Lappi 2018).

Ethics consists of principles, values, and ideals concerning right and wrong, good and bad, and it describes how we should live and act with each other in the society and the environment. Ethics helps us to make choices in life and to evaluate our actions and the reasons behind them. (ETENE 2001.) The aim of healthcare is to prevent diseases, treat them, promote health, and relieve suffering. The decisions made in the healthcare field concerning health, sickness, and death have a significant impact on people´s lives, which lays a major responsibility and influence on that instance. This is the basis of healthcare ethics. (Leino-Kilpi & Välimäki 2014: 14.) AI solutions cannot be used for purposes of healthcare diagnostics unless the AI systems are trustworthy. Ethics can be used as the basis for securing the trustworthiness of AI (European Commission 2019: 6). This explains why the review question of this study,
“What are the main ethical issues using AI in healthcare diagnostics?” consists of two different ethical themes: healthcare ethics and ethics of AI.

2.1 Ethics of AI

Ethics is a study and a system of generally accepted beliefs on what is morally good or bad, right or wrong (Cambridge Dictionary). Ethical issues concerning development, deployment and use of AI are in the core of AI ethics. According to European Commission´s Ethics Guidelines for Trustworthy AI (2019), trustworthiness of AI means that it is lawful, ethical, and robust. Ethics of AI are based on respecting the fundamental rights within democracy and rule of law. These fundamental rights concern issues like dignity, freedom, equality, solidarity, citizens’ rights, and justice; and they can be found defined in International human rights law, EU Treaties, and the EU Charter. (European Commission 2019: 6, 11-12.)

Ethical reasoning that considers the context and its details cannot be substituted by general ethical guidelines. To develop trustworthy AI, ethical discussion should be a fixed part of the technological development process of the systems, education, and practical learning of software engineering and the public debate. (European Commission 2019:

11; Borenstein & Howard 2020.) In comparison, the healthcare field already has a well-established code of ethics since 1970´s, and healthcare facilities have ethics committees that educate healthcare professionals, support, and provide consultation when facing ethical issues, and monitor and take actions in ethical problems as well as form and review ethical policies in that particular community (Véliz 2019).

When processing people´s health data for the use of AI, there is a risk that the human dignity vanishes and people are being treated as objects to be sifted, sorted, scored, herded, conditioned, or manipulated for the use of technology. To maintain and respect human dignity, it is vitally important not to forget the role of a human as moral operator in this process. In the context of AI, equality means that the results the technology provides must not discriminate anyone. This requires a lot of consideration when feeding data to the algorithms. Data should be as comprehensive as possible and represent different population groups, paying special attention to potentially vulnerable groups such as children, women, ethnic minorities, disabled people etc. Ethics guidelines for trustworthy AI list four ethical principles that are: respect for human autonomy, prevention of harm, fairness, and explicability. (European Commission 2019: 11-14.)

2.1.1 The principle of respect for human autonomy

AI is to be designed to augment, complement, and empower human cognitive, social, and cultural skills. AI must not discriminate, manipulate, cause deception, herding or conditioning but to enable fundamental rights. Because of the wide capacity of AI, there is a risk that instead of enabling fundamental rights, the system might hinder them, and this risk needs to be evaluated. That is why the impact on fundamental rights should be assessed in the beginning of the development process of the system. (European Commission 2019: 14, 19.)

AI solutions should always ensure human oversight in the co-operation between human and technology. This means that users need to be provided with knowledge and tools that enable them to understand the system and how it works at sufficient level, so that they can make independent decisions and assess the output. The oversight can be implemented in different ways. Like for example, the user can affect the decision process, s/he can monitor the system´s operation and intervene in it if necessary, or s/he can monitor the overall activity of the system. (European Commission 2019: 14, 19.) Human autonomy also refers to patient´s right to self-determination. In context of diagnostic AI systems, it means that patients should have right to decide whether AI technology is being used in their diagnosis or not. To be able to decide this, patients need to be provided with sufficient level of information about the AI system and how it works. The information must be presented in a form that is understandable for patients, as non-experts of AI. (Bartneck, Lütge, Wagner & Welsh 2021: 30-31.)

2.1.2 The principle of prevention of harm

The principle of prevention of harm concerns technical robustness and safety of the system itself and its usage, privacy, and data protection. Harm can mean physical and mental harm, and harm caused by asymmetry in power structures or information. Like any other software, also AI systems and the data in it must be adequately protected against cyber-attacks and there must be a fallback plan in case problems occur. The system´s level of accuracy to produce correct judgement is also part of safety and prevention of harm, and so are its reliability and reproducibility. (European Commission 2019: 15, 20, 21.)

Privacy is a fundamental right, and it can be strongly affected by AI systems. Data governance is a fixed part of protecting privacy, and it consists of the quality and integrity of the data fed into the system; its relevance in that specific context; and the system´s capability to process it; as well as determining who has access to the data. AI systems are fed with large amounts of personal data, and during the process of interaction with the user even more data is produced in the form of output the system gives. Privacy and protection of this data must be always ensured so that it cannot be used unfairly or unlawfully against people whose data is being processed. Ensuring the quality of the data means that it must be carefully assessed before feeding it into the system, so that it doesn´t contain socially constructed biases, inaccuracies, errors, or mistakes. Integrity of the data is crucial because wrong kind of data affects the output and distorts it. To protect the integrity of the data, careful testing and documentation of the results must be conducted throughout the whole lifecycle of the system from planning, training, and testing to deployment. To protect privacy, access to the data must be thoroughly recorded into a protocol which defines who can have access to it, how and at which point of the process. (European Commission 2019: 21.)

There are several ongoing projects seeking to connect different types of health databases, such as databases of medical images and genomic repositories. While these projects aim at developing AI technologies for the use of medical diagnoses, it is important that people have control over their own health data. Even though that data would be used for the benefit of AI technologies, people need to be able to trust that their personal health data is protected and through that their privacy is secured. The General Data Protection Regulation (GDPR) has a significant role in data protection and securing people´s privacy as it sets strict rules on the use of health data. (EIT Health 2019.)

2.1.3 The principle of fairness

The principle of fairness refers to non-discriminative function and results of the system, to respecting diversity, and to equal opportunity to access the benefits of the system. “If unfair biases can be avoided, AI systems could even increase societal fairness.” (European Commission 2019: 15.)

Although AI technologies can process big amounts of data and learn from it, the data itself is given to the system by humans. If the given data is insufficient, incomplete, or biased, it can easily lead to conclusions that are unreliable, unsafe, or biased, which in turn can have serious consequences for patients (EIT Health 2019). Data sets fed into AI systems can easily contain unintended biases that can lead to discriminative output and aggravate prejudices and exclusion of some groups or people (Davenport & Kalakota 2019). Software companies who develop diagnostic AI systems can avoid unfair bias by systematically monitoring the processes to ensure transparency, the purpose of the system, its requirements, and decisions by spotting and deleting bias from the data, and by nurturing diversity among their own employees who develop the system. Also, it is recommendable to co-operate and involve other stakeholders, such as end users or even patients, in the system’s development throughout its lifecycle. (European Commission 2019: 22.)

In the best-case scenario AI systems can have positive impact even on future generations. For this to happen, also the environment and other living creatures are to be considered as stakeholders besides humans, and that is why sustainability and ecological responsibility are central objectives in the lifecycle of AI systems and the ethical discussion around it. One way to enhance sustainability is to pay attention to and measure the environmental footprint of the system’s supply chain. Environmental wellbeing is important but so is social wellbeing. In the long-run, AI systems will affect various areas of our lives. Consequently, they will have a social impact as well, which might influence people´s physical and mental wellbeing. This must be kept in mind and these effects should be attentively monitored. Social impact does not concern only individuals, and therefore it is important to observe the issue also from the viewpoint of institutions, democracy, and society. (European Commission 2019: 23.)

In the business-to-consumer context, accessibility and universal design are important components of ethical AI and the principle of fairness. (European Commission 2019: 22). This means that systems should be designed so that their usability does not discriminate anyone, for example disabled people. However, AI systems used for healthcare diagnostics work in business-to-business context where the end-user is a physician or other healthcare professional with required knowledge to use the system and the ability to understand the interaction with it.

2.1.4 The principle of explicability

The principle of explicability means that the processes must be transparent, and the purpose of the system is openly communicated. In addition, the output of the system should be explainable, but this cannot always be fully accomplished, as some decisions that come from complicated AI algorithms cannot be interpreted or explained by humans. These are called ”black box” algorithms (see 2.2 AI in healthcare diagnostics) which are often deep learning algorithms used for image analysis. They raise a difficult issue with transparency. (Davenport & Kalakota 2019.) In spite of this, also these outputs can be opened to some extent by using other explicability measures (European Commission 2019: 15).

Transparency in this context refers to the data, to the system, and to the business models that are being used with that particular system. All the data and its processes as well as the outputs must be carefully documented so that in case of a mistake it is possible to trace down where the mistake has happened and how it can be fixed. Secured traceability supports explicability. Explicability means that the system´s processes, decision-making, and the way in which the given output was reached can be understood and explained to any stakeholder in an understandable manner. (European Commission 2019: 21-22.) According to Bartneck et al. (2021: 36) in practice this means that for example an expert AI programme can understand the system and is able to explain it to users or judges if needed.

People have the right to know when they are interacting with AI and have the possibility to decline from this interaction and have human interaction instead to ensure their fundamental rights. Also, the capacity, level of accuracy, and possible limitations of the system must be communicated clearly to system users. (European Commission 2019: 22.)

2.2 AI in healthcare diagnostics

The amount of data that is and can be gathered from people´s medical records, is enormous. The benefits of properly analysing this data are significant for individuals´ health and wellbeing as well as for society, for instance in the form of savings in many areas. Different forms of artificial intelligence are being used to facilitate the analysis and processing of this big data.

Logic is a science that deals with valid reasoning, and reasoning is an important part of intellectual functions and trustworthiness of AI. With computers, algorithms are used to implement logic. Algorithms in turn mean detailed descriptions or instructions to be followed in problem-solving, for instance in calculations. Although logic is needed, in real life most problem-solving situations are ambiguous and therefore cannot be explained logically. That is why practical reasoning is more decisive than valid reasoning. (Suomen Koodikoulu 2019: 15.)

Artificial intelligence (AI) is not one technology, but a collection of different kinds of technologies (Davenport & Kalakota 2019). Machine learning, a subfield of AI, and deep learning that is respectively a subfield of machine learning, are the most common techniques for AI systems used for medical diagnosis (Gerke, Minssen & Cohen 2020: 296). In machine learning a computer is given a dataset and algorithms to work with. The system can then learn from that data and improve its performance without being explicitly programmed to do exactly that (Mehta & Devarakonda 2018).Machine learning technique imitates the function of human brain in form of artificial neural networks. Neural networks are formed of connected neurons that transmit signals between each other. Over time, the most used connections become stronger. This means that the machine learns to find patterns from the data processing it with prewritten algorithms. Deep learning refers to multi-layered artificial neural networks that identify patterns in massive amounts of data. (Suomen Koodikoulu 2019.) Where machine learning requires a human to identify certain features, deep learning deduces the features that predict the outcomes by itself (Mehta & Devarakonda 2018). AI can continuously train itself, and that is why it is probable that the already promising performance in healthcare diagnostics will improve in the future.

Machine learning can be supervised or unsupervised. In supervised machine learning algorithms are trained with datasets that already have inputs and outputs, so they learn which output comes from which kind of input. Unsupervised machine learning, in comparison, is based on processing the data and organizing it into clusters when patterns and similarities are found. With deep learning the process that leads to the output is often complex and opaque, and it is commonly referred to as “black box”. This means that the details of the process leading to the outcome are so complex that they are unavailable and impossible to track down even to the programmer of the system. (Mehta & Devarakonda 2018.)

3 Purpose and aims

The purpose of this review is to produce information about the main ethical aspects of using AI in healthcare diagnostics. The aim is to define the main ethical aspects of using AI in healthcare diagnostics.

The review question of this thesis is: What are the main ethical issues using AI in healthcare diagnostics?

4 Research methods and data collection

This thesis is a scoping literature review. A scoping review provides a broad overview of a certain topic (Peters et al. 2020). It is used to give an understanding of the quality, quantity and/or perspective of the research done on particular theme, and it summarises the evidence of existing research for further use of downstream user (Stolt, Axelin & Suhonen 2016: 10-11). This type of review is useful for examining emerging evidence when other more specific questions are still unclear. The scoping review works best in a situation where the interest of research is in the identification and mapping of certain characteristics and concepts in sources of evidence; and reporting on the findings and discussing them (Peters et al. 2020). Techniques of using AI in healthcare diagnostics are developing fast and ethical discussion is falling behind. To identify and map the concepts and characteristics of the ethical viewpoint in this context, a scoping review is an appropriate choice of literature review type.

A scoping review protocol provides a plan for the review, and it assures transparency of the study and the whole process. The protocol for this scoping review follows the JBI´s Scoping review framework (Peters et al. 2020). The framework was developed by Arksey and O´Malley in 2005, and JBI presents an option for enhancements to it, proposed later by other researchers (Peters et al. 2020). Figure 1. The Scoping Review Framework by Arksey and O´Malley describes the framework and its steps.

Stage 1 Identifying the research question (see Figure 1) is the starting point of any systematic review. Doing this thoughtfully is important as it later affects the planning of the search strategy. The form and level of detail in research question should be considered carefully so that all relevant studies will appear in searches, while the material still remains manageable. (Arksey & O´Malley 2005.) Levac, Colquhoun and O´Brien (2010) also suggest clarifying the purpose of the scoping review with the research question, which will ease decision-making later when figuring out how to execute the selection of source of evidence and data extraction. In this paper the research question is referred to as review question.

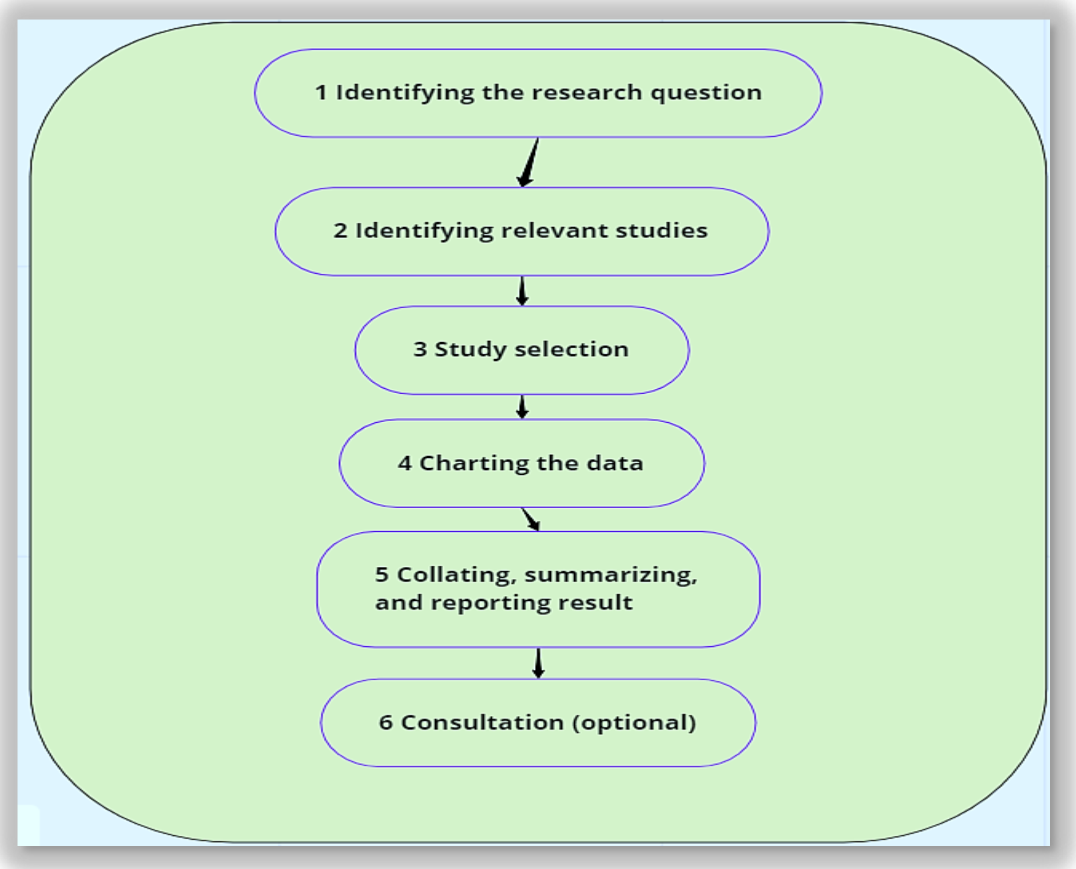

1 Identifying the research question
2 Identifying relevant studies
3 Study selection
4 Charting the data
5 Collating, summarizing, and reporting result
6 Consultation (optional)

Figure 1. Scoping Review Framework by Arksey and O´Malley (Peters et al. 2020)

Stage 2 Identifying relevant studies (see Figure 1) goes to the core purpose of the scoping review: to find and map comprehensively all the published and unpublished research articles and reviews that discuss the topic and answer the review question (Arksey & O´Malley 2005). Comprehensiveness, depth, and breadth can be seen as the main strengths of scoping reviews, but the researcher must also understand the limits of resources like time and money when planning the depth of the study (Levac, Colquhoun & O´Brien 2010). Source of evidence should be widely searched from different databases, reference lists from other selected studies, manually from professional journals and publications from relevant organizations, conferences, and networks (Arksey & O´Malley 2005).

After identifying all relevant studies, it is time to screen them and reject the ones that do not properly address the review question and select the studies to be included in the review (Arksey & O´Malley 2005). Inclusion/exclusion criteria must be defined, and they should give clear instructions on how to execute stage 3 study selection (see Figure 1). Defining inclusion/exclusion criteria is an iterative process. While reading the studies, one’s understanding of the topic grows and makes it easier to determine what kind of studies are relevant for the purpose. Levac, Colquhoun, and O´Brien (2010) strongly suggest involving a multidisciplinary team in the process of study selection, which will strengthen trustworthiness and transparency. However, with this thesis there is no possibility for using a team effort for the process.

4.1 Search strategy

Defining a search strategy for the literature review is relevant as errors that occur during the search process lead to skewed conclusions (Stolt, Axelin & Suhonen 2016: 25). The search strategy started with identifying review question and keywords that were used in search of evidence for this review. To identify keywords and to formulate the review question PICO tool is often used in case of quantitative research and PICo tool with qualitative research (Murdoch University). A combination of PICO and PICo models was used in this thesis. The concepts refer to P= population, I= Intervention, Co= Context and O= Outcome and by dividing the review question into categories, the review question can be adequately formed, and suitable search terms can be selected (Aromataris et al. 2020). The outcome of this research is targeted to organizations who design and develop AI systems for healthcare diagnostics, to organizations who purchase them, and to healthcare professionals as end-users of these systems. Nevertheless, the target groups are not mentioned in the review question as explained earlier (see chapter 1. Introduction). That is why the “population” concept doesn´t appear in PICO/PICo.

The review question is: What are the main ethical issues (O) of using AI (I) in healthcare diagnostics (Co)? The keywords were identified from PICO/PICo categories. These keywords and their synonyms and alternative spellings were tested during an initial search, and the final search terms were selected based on that, as shown in Table 1 PICO/PICo and selection of search terms. The U.S. Medical Subject Headings (MeSH) official terminology search engine is widely used globally in medical research and study (U.S. National Library of Medicine, 2019). MeSH terms were also searched at the initial phase to make sure the right terminology is in use when realizing the actual search. This revealed that the MeSH term ”deep learning” was only added to the terminology in 2018, which would indicate that by using the aforementioned term, the search would give recently published articles.

Table 1. PICO/PiCo and selection of search terms

PICO/PICo concepts	Search terms including alternative spellings, synonyms, and abbreviations:
I=Intervention	AI, artificial intelligence, machine learning, deep learning
Co=Context	Healthcare, health care, care, medical care, medical Diagnostics, diagnosis
O=Outcome	Ethics, ethical/moral issue/problem/matter/question

4.1.1 Inclusion/exclusion criteria

Inclusion/exclusion criteria are the characteristics that the article must have in order to be included in or excluded from the review. These criteria guide the selection of studies to be included in the review. (Stolt, Axelin & Suhonen 2016: 26.) The inclusion and exclusion criteria were defined in advance when planning the search strategy. During the iterative search and evaluation process the necessity of refining the criteria emerged a couple of times to ensure quality and consistency of the chosen texts. The criteria for article selection are listed in Table 2. Inclusion and Exclusion Criteria.

The inclusion criteria included articles written only in English or Finnish and published not earlier than January 2010. The language criteria were set based on author´s personal language skills, and the limit to the publishing date can be justified with the development of artificial intelligence techniques that have grown explosively just during the last decade. Articles needed to be empirical quantitative or qualitative, peer reviewed studies, systematic or integrated reviews, expert-driven guidelines, or professional organization/institution reports. Books, letters, conference proceedings, and theses were excluded from this review. To spot the most relevant evidence out of a big amount of technical and medical literature all three key concepts, AI, healthcare diagnostics, and ethics had to be found in the abstract for the article to be included in the review. When screening the search results on full text level, studies that treated one of the key concepts only superficially were excluded.

Table 2. Inclusion and Exclusion Criteria

Inclusion criteria	**Exclusion criteria**
Studies in English or Finnish	Any other languages
Empirical quantitative and qualitative peer reviewed studies, systematic, integrated and narrative reviews	Deals with AI in healthcare without ethical aspect

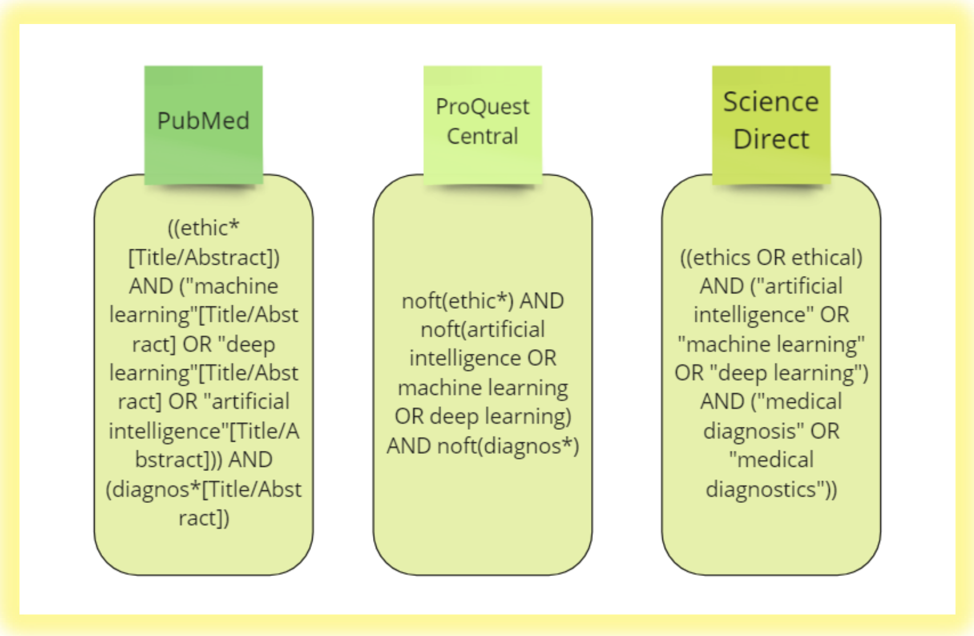
Expert-driven guidelines	Deals with digital health in general
Professional organization/institution reports	Deals with ethics in other than AI context
Full text articles that discuss all three key concepts more than superficially	Deals with AI in other setting than healthcare diagnostics
	Deals with healthcare diagnostics without AI context
	Published before 1.1.2010
	Full text articles where one or more of the three key concepts are only treated superficially

4.1.2 Literature search and selection of studies

The databases used for the literature search were: ProQuest Central, Science Direct and PubMed, and they were selected because they are relevant to the research question and have content from medical, ethical, and technical research. To identify right sources for this review, the recent research about ethics of using AI in healthcare diagnostics was searched and studied carefully by following JBI´s three-step search strategy recommended for scoping reviews. It includes the pilot search, the actual search, and the analysis of the results first on title-level, next on abstract-level, then on full-text level, and lastly, the search for more sources in the reference lists of chosen articles (Peters et al. 2020). All the steps of the search were documented carefully, step by step, and duplicates were removed on abstract level using RefWorks citation manager.

The pilot search was done in all the chosen databases in December 2020. The keywords and concepts of the titles and abstracts were analysed from the result of the first search. The initial search indicated that “AI” cannot be used as a search term because the databases did not recognise it well and the searches gave a huge number of results not concerning the topic of the review question. The results also confirm that AI subfields machine learning and deep learning are the most used techniques in systems for healthcare diagnostic purposes. That is why they were chosen as search terms for the actual search. Another important finding was that using the search term “medical diagnosis” instead of “healthcare diagnosis” gave significantly better search results in terms of quantity and relevance to the topic. Later, when iterating the search, the decision was made to only use the search term “diagnostics” or “diagnoses” without a reference to healthcare, as the search results still didn´t significantly include articles concerning other topics than healthcare. The initial search also showed that abstracts of the right kind of source of evidence, mentioned *diagnostics* in healthcare setting, *artificial intelligence*, and *ethics*. Based on this, these three were chosen as *key concepts* of this research.

The actual search was done during March 2021 in databases Science Direct, ProQuest Central and PubMed. The full search sentences in each database are shown in Figure 2. Search sentences in chosen databases.



PubMed

((ethic*[Title/Abstract]) AND ("machine learning"[Title/Abstract] OR "deep learning"[Title/Abstract] OR "artificial intelligence"[Title/Abstract])) AND (diagnos*[Title/Abstract])

ProQuest Central

noft(ethic*) AND noft(artificial intelligence OR machine learning OR deep learning) AND noft(diagnos*)

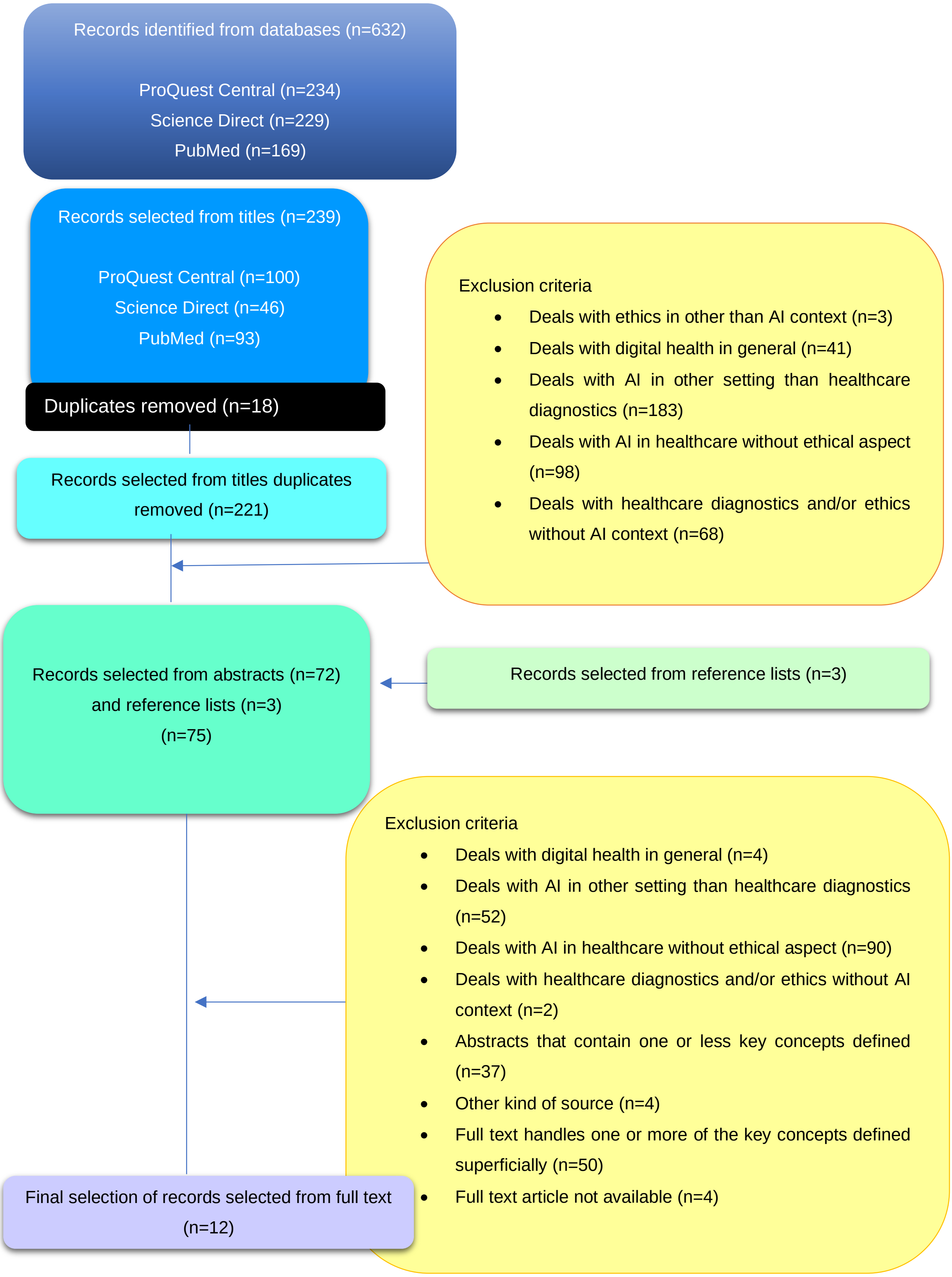
Science Direct

((ethics OR ethical) AND ("artificial intelligence" OR "machine learning" OR "deep learning") AND ("medical diagnosis" OR "medical diagnostics"))

Figure 2. Search sentences in chosen databases

The search terms were connected with Boolean operators “AND” and “OR” in all the databases. The same way of using search terms did not work in all the chosen databases, but the search had to be modified in order to get reasonable and good quality results. In PubMed the appearance of search terms in the title or abstract was added to the search criteria. In ProQuest Central the command for appearance of search terms was “anywhere except full text”, which in search sentence is shown as “noft” (see Figure 2). To get relevant results, more filters needed to be added concerning subject areas (medicine and dentistry, computer science, nursing, and health professions) and article types (review articles, research articles, practice guidelines) when running a search in Science Direct.

The results of search strategy are illustrated in detail in flow diagram in Figure 3. Search strategy. The actual search gave altogether 632 results divided between three chosen databases as follows: ProQuest Central n=234, Science Direct n=229, and PubMed n=169. All the titles were read and analysed based on inclusion/exclusion criteria, and 393 studies were excluded. Duplicates (n=18) were removed from remaining titles (n=239) leaving 221 titles to the next stage of data selection process. The abstracts (n=221) were carefully read and again compared with the inclusion/exclusion criteria and as a result, 149 studies were excluded at this phase. The remaining articles (n=72) were read carefully through, and the final selection of studies to be included in this review was made according to the inclusion/exclusion criteria. Many of these remaining 72 studies concerned AI in healthcare diagnostics and they did contain a short section about the ethical point of view. However, this part of the study often remained superficial, and rather than being an actual part of the research, the ethical discussion was limited to the authors´ reflections on the topic. That is why the decision was made to include only studies that focus on ethical issues of using AI in healthcare diagnostics. To recognise these studies an addition was made to the inclusion/exclusion criteria as shown in Table 2. Inclusion and Exclusion Criteria. After carefully reading the full text studies, 12 were included in the final selection of this review.



Records identified from databases (n=632)

ProQuest Central (n=234)
Science Direct (n=229)
PubMed (n=169)

Records selected from titles (n=239)

ProQuest Central (n=100)
Science Direct (n=46)
PubMed (n=93)

Duplicates removed (n=18)

Records selected from titles duplicates removed (n=221)

Exclusion criteria

- Deals with ethics in other than AI context (n=3)
- Deals with digital health in general (n=41)
- Deals with AI in other setting than healthcare diagnostics (n=183)
- Deals with AI in healthcare without ethical aspect (n=98)
- Deals with healthcare diagnostics and/or ethics without AI context (n=68)

Records selected from abstracts (n=72) and reference lists (n=3)
(n=75)

Records selected from reference lists (n=3)

Exclusion criteria

- Deals with digital health in general (n=4)
- Deals with AI in other setting than healthcare diagnostics (n=52)
- Deals with AI in healthcare without ethical aspect (n=90)
- Deals with healthcare diagnostics and/or ethics without AI context (n=2)
- Abstracts that contain one or less key concepts defined (n=37)
- Other kind of source (n=4)
- Full text handles one or more of the key concepts defined superficially (n=50)
- Full text article not available (n=4)

Final selection of records selected from full text (n=12)

Figure 3. Search strategy (modified from PRISMA Flow Diagram 2009 by the author)

5 Data charting

Data charting is a process that presents in a simple and logical manner why the source of evidence included in the review was selected and how it relates to the review question and the objectives of the review. It defines details and characteristics of the source of evidence key findings that are relevant to the review. (Peters et al. 2020.) The entire data charting of the selected studies is seen in Appendix 1. Data charting, and the key characteristics chosen to chart were author(s), year of publication, country of origin, aims and purpose of the study, study design, description of data and methods, and main results of the study.

Among the twelve articles selected there were five review articles, one editorial, one research article, a white paper, an extended essay, a comment paper, a bulletin of the WHO, and a condensed summary of an international multisociety statement. All the articles were published between 2018 and 2021, which makes the evidence up to date. The majority of the selected studies were from European origin (66,66%) and the rest were published in North America or Australia. The countries of origin of the studies were Switzerland (n=3), Germany (n=2), UK (n=2), Italy (n=1), Australia (n=1), Canada (n=1) and USA (n=1). The condensed summary of an international multisociety statement was from Europe, USA, and Canada. (See Appendix 1. Data charting.)

5.1 Quality appraisal and assessment of bias

Critical appraisal of quality and assessment of bias of the studies that are included in the review is as important part of the literature review as is analysing the results from them. Quality appraisal can be described as “a systematic examination of literature to evaluate its reliability, value, and relevance in a particular context”. (Toronto & Remington 2020: 45-46.) The purpose of quality appraisal is to enhance the quality and validity of the review and its outcomes, and to minimise the errors and bias that emerge from the possible uncertainties in the research process or synthesis of the original studies. The errors and bias that the studies in scope contain can seriously distort inferences from the analysis of the data, which directly affects the trustworthiness of the review. (Acosta, Garza, Shu & Goodson 2020.)

While systematic reviews report the development of accurate clinical guidelines and recommendations, scoping reviews aim to produce an overview of the existing evidence. That is why quality assessment of the literature is normally not conducted with scoping reviews. (Peters et al. 2020.) However, Metropolia University of Applied Sciences has included assessment of the quality of sources to its requirements for master´s thesis. Many kinds of quality assessment tools exist for evaluating the quality of studies. After testing several quality assessment tools, such as SANRA (Baethge, Goldbeck-Wood & Mertens 2019), which is a quality assessment tool of narrative review articles, and STROBE checklist for cohort, case-control, and cross-sectional studies (combined) (STROBE 2022), JBI critical appraisal checklist for text and opinion (JBI 2017) was found to be the most suitable assessment tool for these selected articles in question and was chosen as tool to carry out the quality assessment. JBI critical appraisal checklist for text and opinion includes six questions, and studies can score from 0 to 12 points (see Appendix 2. Quality Assessment). All the studies that scored 9 points or more in terms of quality were included in this review.

The entire quality assessment process was carefully charted and can be seen in Appendix 2. Quality Assessment. Six of the selected studies (50%) scored full 12 points, four of them scored 11 points and two scored 10 points, so they all passed the set limits for quality assessment and were included in the review.

5.2 Analysing the data

Unlike systemised reviews, scoping reviews do not aim to synthesise the results of the included data, but rather to map the findings (Peters et al. 2020). The data of this review was processed by using inductive content analysis and the analysing process followed the stages presented by Erlingsson and Brysiewicz (2017) and illustrated in Figure 4. Inductive Content Analysis and Example. Inductive content analysis is a method to analyse data by reducing the volume of the text, identifying and grouping categories from it, and drawing conclusions from that information (Bengtsson 2016).

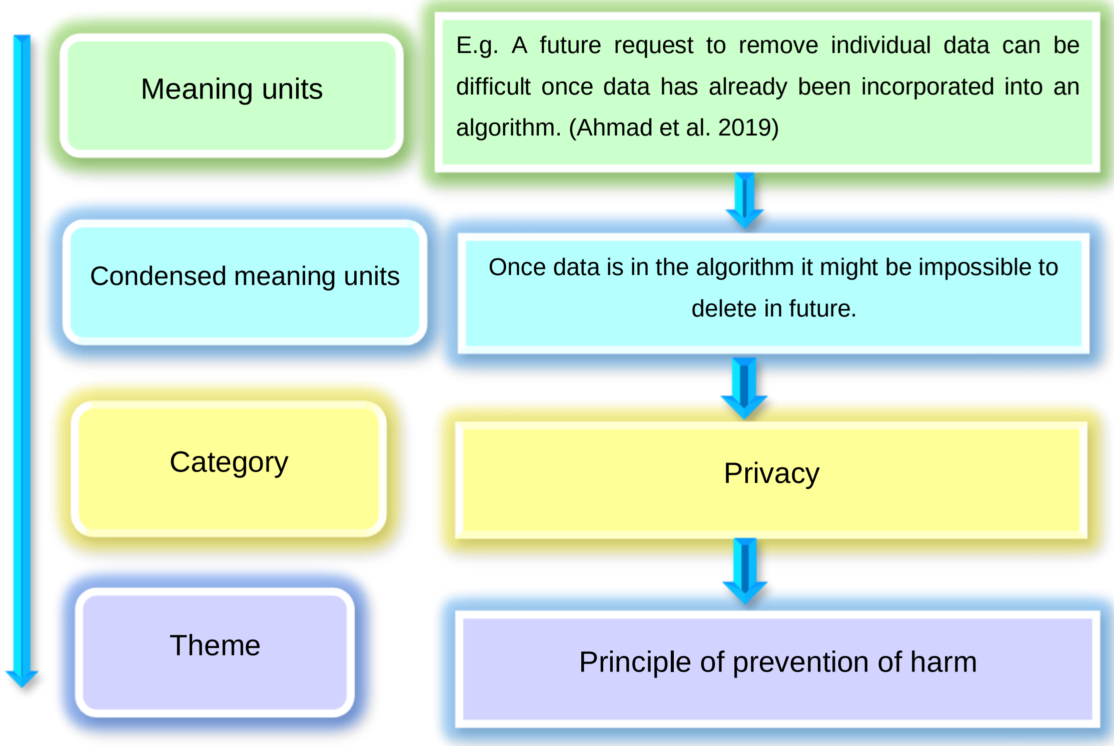

Meaning units
E.g. A future request to remove individual data can be difficult once data has already been incorporated into an algorithm. (Ahmad et al. 2019)
Condensed meaning units
Once data is in the algorithm it might be impossible to delete in future.
Category
Privacy
Theme
Principle of prevention of harm

Figure 4. Inductive Content Analysis and Example (Erlingsson & Brysiwicz 2017, modified by the author)

In the condensation stage, the studies included in the review were carefully read and re-read one by one. Next, the information that answers the review question was recognised and charted in meaning units which again were processed further into condensed meaning units. It was essential to make sure that the core information did not change in the condensation process. According to Erlingsson and Brysiewicz (2017), the subsequent step in the analysis process would have been coding these condensed meaning units with descriptive labels. Instead, during this analysis process, the condensed meaning units naturally formed categories and the coding was done only after that stage. The coding was done by using different colours. At the last stage, categories were grouped together under similar themes. An example of analysis process is shown in Figure 4. Inductive Content Analysis and Example.

The content analysis process produced four themes and twelve categories (see Figure 4. Inductive Content Analysis). While doing the content analysis and identifying the categories and themes, the results very naturally and automatically led to the direction of the same main subjects that were already introduced in the theory chapter of this thesis. So, the themes were named according to the four ethical principles defined by “Ethics guidelines for trustworthy AI” as: Principle of Respect for Human Autonomy, Principle of Prevention of Harm, Principle of Fairness, and Principle of Explicability (see Chapter 2.1 Ethics of AI). As the subjects in concern are closely linked with each other, they are often being discussed at same time and it is not always possible, necessary, or even appropriate to separate them from each other. For that reason, some of the condensed units discussed subjects so that they fit into two different categories and/or themes, therefore secondary categories and themes were created for them.

6 Findings

The findings from the content analysis were presented by following the themes and categories formed during the content analysis process. Altogether four themes and twelve categories under them were formed in content analysis process as illustrated in Figure 5. Themes and Categories.

6.1 Principle of prevention of harm

Under the theme Principle of prevention of harm, four categories were recognised. As illustrated in Figure 5. Themes and Categories, they were Intention, Safety, Privacy and Accuracy. The Intention Category discussed the ethics of the motives behind the development, purchase, and usage of diagnostic AI systems. The Safety and Privacy Categories were discussed together under one sub-chapter, as the matters under them were so closely connected with each other throughout the reviewed articles. Safety and Privacy form a particularly important part of the ethics of diagnostic AI, and they were extensively discussed from various angles. The Accuracy Category was its own sub-chapter reviewing different aspects of the topic.

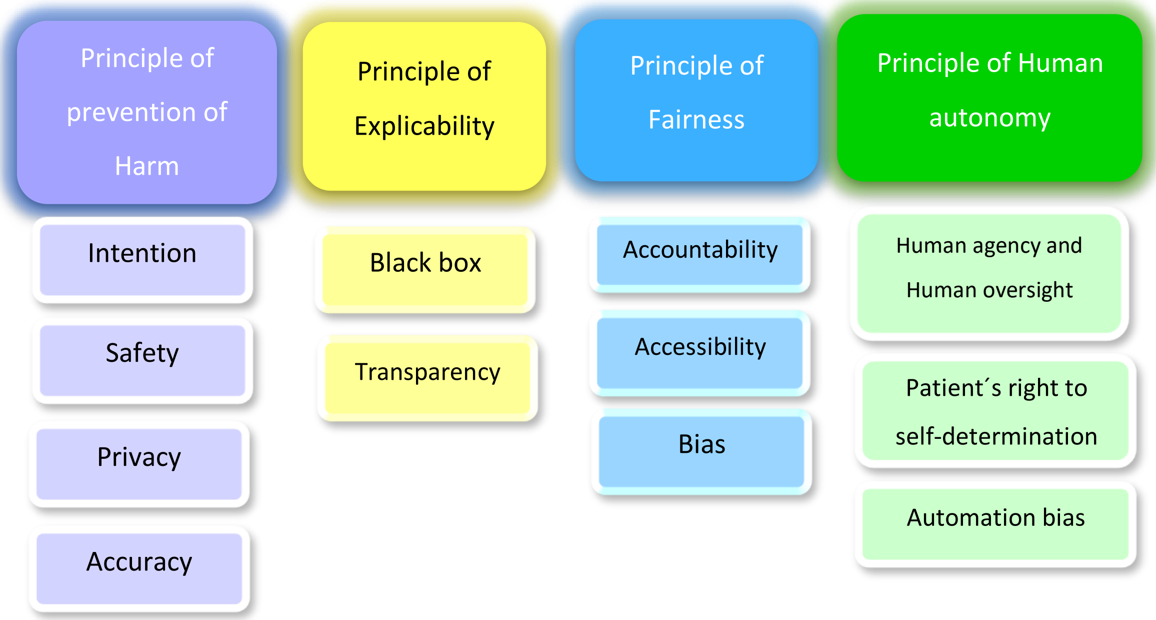

Principle of prevention of Harm
Intention
Safety
Privacy
Accuracy
Principle of Explicability
Black box
Transparency
Principle of Fairness
Accountability
Accessibility
Bias
Principle of Human autonomy
Human agency and Human oversight
Patient´s right to self-determination
Automation bias

Figure 5. Themes and Categories

The danger of mass unemployment that is traditionally often mentioned along with the development of AI, especially in media, was only brought up by Brady and Neri (2020) in the context of radiology and the profession of radiologist. Their conclusion is that when AI takes charge of more routine tasks, the work of radiologists will develop and focus on more complicated areas of the job that require cognitive thinking and other human abilities. Instead of fearing whether AI will replace radiologists, one should realise that “radiologists who use AI will in future replace radiologists who do not.” (Brady and Neri 2020.) That is why it is not a realistic threat or an ethical issue.

6.1.1 Intention

The intention and motive behind the development of diagnostic AI system can become an ethical issue because it is possible to programme AI to perform in unethical ways. Instead of aiming for best possible care and positive impact on patient outcome through improved prevention or diagnosis for instance, the driving motive behind the design of AI can be based on profits for manufacturers and investors. (Pesapane, Volonté, Codari &

Sardanelli 2018.) There is also a significant difference between the moral code that guides the work of healthcare staff and developers. The work of health care professionals is guided by strong code of medical ethics, but system developers do not have a similar code of working ethics, and therefore they are not required to put patient´s interest first. (Carter et al. 2019.)

According to Brady and Neri (2020), “the potential exists for radiologists and others involved in AI development to find themselves in conflict-of-interest situations if and when commercial decisions are being made regarding purchasing and use of AI tools.” For that reason, it is critical that stakeholders from the healthcare setting, especially the ones making purchasing decisions on new systems, require to know what kind of policy is leading the development, which goals are set for algorithms and which values drive them (Carter et al. 2019).

6.1.2 Safety and privacy

Sensitivity and value of health data create risks around safety and privacy. That is why data ownership, consent for use of data, and data protection and privacy are critical issues. (Brady & Neri 2020; Carter et al. 2019.) In particular, the health data that is used to train diagnostic AI systems is both, very valuable and very sensitive. Because the system uses this data to produce output, also that output data is sensitive. (Pesapane et al. 2018.) Using health data creates a risk for data breaches and harm. (Carter et al. 2019; Brady & Neri 2020; Pesapane et al. 2018) Moreover, capitalizing health data unethically can harm patients and common good (Geis et al. 2019). The source of health data used for training the system can be ethically questionable, which creates an ethical issue concerning safety (Pesapane et al 2018). The output of systems that are created with low quality data, and besides that are not maintained and updated scrupulously, can compromise patient outcomes with possible serious consequences (Arambula and Bur 2019).

Recording private health data into the system is a requirement for a diagnostic AI system to function (Pesapane et al 2018). But when an AI system uses detailed health data, individuals can in some cases be quite easily recognised from the provided output, which creates a threat to privacy. (Ahmad et al. 2019; Arambula and Bur 2019; Pesapane et al. 2018; Ienca and Ignatiadis 2020; Brady and Neri 2020;). In case of radiology for instance, image data can easily “capture something of person´s essence” (Jaremko et al. 2019), or private neural data may expose people so that they can be identified or re-identified by using AI methods (Ienca and Ignatiadis 2020).

Complicated data governance and protection bring challenges to companies using diagnostic AI systems (Ahmad, Stoyanov & Lovat 2019). Diagnostic AI systems are trained by using sensitive health data, and they produce an output by processing that data into another form of vulnerable data. That creates a risk for privacy for the people whose data is being used to train the system. (Brady and Neri 2020; Starke, De Clercq, Borgwardt & Elger 2020; Jaremko et al. 2019; Pesapane et al. 2018).) According to Jaremko et al. (2019), privacy data breaches can cause harm such as, discrimination, humiliation, or increased cost of insurance for instance.

For that reason, it is important to get a consent from people whose private health data is being used as training data. But the way algorithms work and how the system´s output is being used later makes it difficult to be precise on where, how, by whom and for what purposes that data might be used for during its life cycle (Ahmad et al. 2019; Jaremko et al. 2019). Patients cannot really give an explicit consent to use their data to train algorithm, unless they get explicit information on who has the access to the data. They also need to know if there is secondary use of it and to which degree the data is being anonymised. In addition, they should be notified if the data is commercialised at some point. (Ahmad et al 2019; Starke et al. 2020.) Another ethical problem here is that once the data is in the algorithm, it might be impossible to delete in the future even though the person whose data it is, would request for it (Ahmad et al. 2019). Starke et al. (2020) also raise an important ethical question concerning consent and patient´s right to self-determination in context of using diagnostic AI tools in care of especially vulnerable patients. Could discussing complicated AI systems and algorithms with a vulnerable group of people, suffering for example from psychotic or paranoid symptoms, cause psychological stress and worsen their situation (Starke et al. 2020)?

Tackling ethical issues of safety and privacy requires broad collaboration between different stakeholders, including clinicians and patients, in the development process of AI. (Carter et al. 2019; Starke et al 2020). Also, well planned quality assurance processes are needed if AI is integrated into clinical practice. To protect privacy and safety different continents have already special regulations for personal and health data protection. To mention a few, in Europe it is called GDPR - The General Data Protection Regulation, in USA HIPAA - The Health Insurance Portability and Accountability Act of 1996, and in Australia the Australian Privacy Principles found in the Privacy Act 1988. (Carter et al. 2019.)

6.1.3 Accuracy

Accuracy of diagnostic AI tools creates an important part of ethical discussion and is closely connected to the safety of the system. To prevent harm caused to patients, AI systems need to provide accurate diagnostic results. The way AI systems work and are built creates an accuracy- related risk for ethics. If diagnostic AI systems are misused by users, or they are technically malfunctioning and consequently produce incorrect output, they are likely to cause harm (Jaremko et al. 2019). To ensure accuracy of a diagnostic AI system there is a need for well-defined standards the AI system follows and procedures to ensure that they are being followed. (Pesapane et al. 2018.) According to Starke et al. (2020), accuracy can also have a significant positive ethical effect on certain patient groups that have suffered from lack of attention and resources more than others. More precise diagnoses and better treatments might convince policymakers to allocate more budget for example on mental health, and as a consequence, psychiatric patients would be empowered. (Starke et al. 2020.)

One risky scenario for accuracy is training data that might be technically outdated fast, for example in imaging, which then affects the accuracy of the output (Brady and Neri 2020). Another notable issue is that the algorithm defines the disease in one certain way although significant variation exists on the conceptual norms of diseases in different regions, even between different hospitals (Grote and Berens 2019). This can seriously affect the accuracy of a diagnostic AI system, and it is not an easy dilemma to solve due to the way the algorithms work. Diagnostic AI systems can also skew the amount of positive or negative results. The increased number of false positive results causes harm as resources are wasted into unnecessary examinations and other procedures. The increased number of false negative results in turn means failing at diagnosis. (Jaremko et al. 2019).

6.2 Principle of explicability

Principle of explicability theme contained only two categories that are Black box and Transparency (see Figure 5. Themes and Categories). Although both categories discuss the same subject, a difference in the point of view could still be distinguished. Matters under Black box category concern the opacity in the way the algorithms work, and the ethical issues generated by the unpredictable and unexplainable decision-making process of diagnostic AI systems. Matters under Transparency category bring up ethical issues that can be found in the contradictions between transparency and accuracy and in the way humans and machines think and make decisions. Also, issues concerning the trust of patients and healthcare professionals in decisions that are not fully explainable are included under the category of Transparecy.

6.2.1 Black box

One of the issues that causes most uncertainty and questions around ethics of AI is its complex decision-making process. An algorithm functions in a way that cannot fully be explained, and this is generally referred as “black box”. It makes the decision-making unpredictable. (Pesapane et al. 2018; Arambula and Bur 2019; Geis et al. 2019; Heinrichs and Eickhoff 2019.) Because of the black box, AI system´s output does not answer to the common need of transparency, and it can violate information rights in terms of the requirement to appropriately inform patients on their care and examinations. The lack of evidence also makes us doubt things although they seem to be true. (Heinrichs and Eickhoff 2019.)

Can the diagnosis be trusted and patients and providers be comfortable with using systems despite the black box dilemma, ask Arambula and Bur (2019). Medical science and practice have been strongly relying on evidence-based medicine throughout the history until present. According to Brady and Neri (2020), expecting patients and doctors now to trust results that cannot be fully explained is taking us away from the evidence-based medicine. Carter et al. (2019) make a strong statement that the use of non-explainable AI should be prohibited in healthcare owing to high ethical and medicolegal requirements of the healthcare field.

6.2.2 Transparency

Current algorithms that are less explainable are also often proved to be more accurate, which indicates that demanding transparency on the AI system might harm getting the best performance out of it (Carter et al. 2019; Ahmad et al. 2019; Grote and Berens 2019; Heinrichs & Eickhoff 2019). It is a big challenge that despite accurate output diagnostic AI systems do not provide understanding on how they came to the final decisions (Heinrichs & Eickhoff 2019). Results from Diagnostic AI systems are often proved to be more accurate compared to human professionals, but this benefit comes with uncertainty (Grote and Berens 2019). An important question is whether we must choose either accuracy or explicability, or whether we can have them both? (Carter et al. 2019).

The way physicians and AI systems are trained is very different, and distinct is the way of reasoning of a human and a machine. In case there is a disagreement between the physician´s opinion and system output, there is a dilemma what to do because of the opacity in the algorithm´s process. (Grote and Berens 2019.) Heinrichs and Eickhoff (2019) point out that the highly accurate results that AI systems provide may be undervalued by professionals due to the opacity of decision-making and in case the results contradict with their own professional experience.

It is important to be as transparent as possible on how algorithms come to a decision to build trust among patients and healthcare professionals (Geis et al. 2019). On the other hand, too much transparency on diagnostic AI systems processes, especially to people that are not direct stakeholders, can also compromise privacy (Geis et al. 2019). When While AI systems are doing part of the tasks in diagnostics, clinicians will have more time to talk with patients. Still, they might not be able to explain decisions in detail, as we are commonly used to. (Carter at al. 2019.) It is also important to understand how the system came to its conclusion and trace the process in case something goes wrong (Geis et al. 2019). In general, there is a need for clearly defined expectations for explicability (Carter at al. 2019).

6.3 Principle of Fairness

Three categories were formed under the Principle of Fairness theme (see Figure 5. Themes and Categories). The first category, Accountability, discusses the responsibility issues in case a diagnostic AI system generates flawed output. The second category, Accessibility, brings up ethical points in respect to equal opportunity to access the benefits of diagnostic AI systems. The third category, Bias, is a broad topic which forms a big part of the discussion around ethics of diagnostic AI system.

6.3.1 Accountability

It is obvious that if a diagnosis that an AI system produces is incorrect, it causes harm, not only directly to the patient, but also other kind of harm, mainly wasting valuable financial and human resources (see chapter 6.2.3 Accuracy). Besides the consequences and harm, also responsibility issues are important in this situation.

When using diagnostic AI and AI in general, the attribution of accountability is very complicated in a situation where something goes wrong. Who is responsible if the output the algorithm produces is flawed? Is it the clinician using the system, the healthcare organization that purchased the system and demanded their clinicians to use it? Or is the responsibility on the organization who created the system, or on the system developers who created and coded the algorithm? (Grote and Berens 2019; Carter et al. 2019; Heinrichs and Eickhoff 2019; Ahmad et al. 2019; Pesapane et al. 2018; Brady and Neri 2020.) The problem is that each of these stakeholders have contributed to the act that ended up causing harm, but neither of them has the full blame (Grote and Berens 2019).

It is to be expected that clinicians will decline to take responsibility on decisions they cannot control or explain (Carter et al. 2019). According to Heinrichs and Eickhoff (2019), it is not possible to assign responsibility for a problem that is formed by a black box to an individual, especially not to an individual who is a system user. The reality that the code of medical ethics leads the work of health care professionals, but system developers are not required to put patient´s interest first, is primarily an issue of intention, but it also plays a role when discussing accountability.

To tackle this problem, there is a need for management of accountability issues, machine-human co-operation, and "peer" disagreement situations (Carter et al. 2019). Although it is likely that an inaccurate algorithm affects more patients than a mistake made by a physician, still also humans make mistakes. A model of shared responsibility, in which competent healthcare professionals review the results of diagnostic AI systems, could be an answer to this problem. (Starke et al 2020.) Carter et al. (2019) state that "trust in healthcare system will require at least some public accountability about the use of AI in those systems."

6.3.2 Accessibility

One important ethical issue with diagnostic AI is that not everyone has the same opportunity to enjoy the benefits of these systems. The distribution of diagnostic AI systems is uneven mainly because of lack of resources. (Brady and Neri 2020; Arambula and Bur 2019; Starke et al. 2020; Geis et al. 2019.) According to Arambula and Bur (2019), there are patient groups from certain socioeconomic backgrounds that receive inferior care because the health facilities accessible for them cannot use AI due to lack of financial resources and trained professionals with adequate skills to use the systems. This can concern patients in some countries, regions, or subgroups of society (Brady and Neri 2020). The reverse point of view to accessibility and ethics is that bringing new skills to an environment where no one has them before can be the actual value of a diagnostic AI system (Pesapane et al 2018).

Other reasons inhibiting equal access to the benefits of diagnostic AI systems can be, for instance, limited availability of the technology (Starke et al 2020), limited and insufficient access to big data, limited computing power for deployment of the system (Brady and Neri 2020), and resources to manage complex AI systems (Geis et al. 2019). It is important take into account that the lack of resources is not the only reason that can deny the access to the benefits of a diagnostic AI system. There are contraindications, such as claustrophobia, that can prevent a patient from reaping the benefits of a system. (Starke et al 2020.) The fact that resources granted for healthcare are insufficient is a big challenge globally. For that reason, another important ethical question concerning accessibility is whether increased diagnostic certainty justifies allocation of limited financial resources to expensive exams, such as MRI? (Starke et al. 2020).

6.3.3 Bias

The existing health data is automatically biased as it is primarily collected from people who have access to healthcare and whose health data is historically being collected. This rules out many other individuals and groups of people who do not belong to this group, due to gender, age, sexual orientation, ethnic, social, environmental, or economic factors, or other type of disadvantage they have. (Arambula and Bur 2019; Carter et al. 2019; Pesapane et al. 2018; Geis et al. 2019; Starke, De Clercq & Elger 2021; Brady and Neri 2020; Jaremko et al. 2019; Ienca and Ignatiadis 2020; Ahmad et al. 2019.) People that are not represented in the system´s training data are in threat of receiving substandard care (Arambula and Bur 2019), in addition using biased training data reinforces discriminatory practices even further (Ienca and Ignatiadis 2020; Jaremko et al. 2019).

The human biases that system developers have, conscious or unconscious, are easily built inside algorithms and affect the system´s decision-making and output (Pesapane et al. 2018; Stark et al. 2021; Carter et al. 2019). Research data also often over-represents positive results, leaving negative studies under-reported, which then skews the system´s output when used as training data (Brady and Neri 2021). The bias in healthcare is somewhere so deep in the structure that it is very difficult to find, which worsens the situation (Starke et al 2021). Some groups, e.g., racial minorities are more impacted by flawed system output due to bias (Grote and Berens 2019; Ienca and Ignatiadis 2020).

Using health data that is already biased harms the system output, its safety, accuracy, and fairness, and therefore it is a key concern (Starke et al. 2021).

Ethnicity is one factor that puts people in an unequal position, and healthcare delivery varies by ethnicity (Pesapane et al. 2019). This is not only an ethical problem of accessibility (see Chapter 6.1.2 Accessibility) but it also distorts the system´s output, making it an ethical problem of fairness. When diagnostic AI systems are trained with insufficient ethnical variety, systems will give biased outcomes (Pesapane et al. 2019; Jaremko et al. 2019; Ienca and Ignatiadis 2020). Hidden bias in training data can result in systematically skewed output for some groups and lead to unfair treatment (Starke et al. 2021).

In addition to moulding health data, structural racism also has an influence on the kind of care people will receive. For instance, there is a common discriminative belief that due to physiological differences, black people feel a lower level of pain compared to white people. Because of that mistaken belief, black people have been systematically untreated for pain for decades. Also, “black US-Americans are more likely to be diagnosed with schizophrenia when presenting certain symptoms, whereas Caucasians with same symptoms are diagnosed with mood disorders or depression.” The formation of this distortion has been influenced by stereotypes, ethnicity of clinicians, and under-diagnosis of other psychiatric diseases. Using this mistaken information to train an algorithm would result in overdiagnosis of schizophrenia on black people, and the system´s output would further consolidate discriminative practises. (Starke et al. 2021.) In case of mammography, it is possible that the algorithm performs distinctively depending on the woman´s breast tissue or sociodemographic group (Carter et al. 2019). Also, image recognition systems are prone to develop bias that cause disadvantage for racial minorities. This can be seen, for instance, with AI systems that detect skin diseases, and more accurately on a person with a light skin colour compared to a person with a darker skin colour. (Grote and Berens 2019.) This kind of systematic biases are particularly dangerous in diagnostic context.

Another source of bias in medical data is gender. According to Starke et al. (2021), clinical examinations are primarily carried out with male participants. For example, heart attack symptoms on women are often missed because they differ from symptoms men have in the same situation (Starke et al. 2021). Technology can also play a part in the formation of bias in AI systems. For instance, different scanning techniques used to collect data and possible comorbidities the patient has, may result in bias in diagnostic AI systems used in radiology. (Geist et al. 2019.) People with disabilities and individual differences in their bodies, such as deviant neurocognitive features, are easily at risk to be discriminated through the output of a diagnostic AI system (Ienca and Ignatiadis 2020). Starke at al. (2021) point out that a correct diagnosis often requires data on gender and ethnicity, so excluding that information from training data would not be a solution for avoiding discrimination of vulnerable minorities. For instance, systemic lupus erythematosus, a rheumatic autoimmune disease is much heavier on women than men and its incidence is much higher among African, Asian and Hispanic ethnicity people, compared to other ethnicities (Starke et al. 2021).

Training data, the way the algorithm uses that data, and bias introduced by developers of the system are factors to take into account when building ethical diagnostic AI systems. Also, because of the well-known issue of bias, close attention should be paid to the potential for systematically different cohort of people when building diagnostic AI. (Carter et al. 2019.) As Starke et al. (2021) underline: “Developing the system to maximise the benefits for the majority justifies overlooking the needs of vulnerable minorities”.

6.4 Principle of human autonomy

The three categories that were formed under the Principle of human autonomy theme are Human agency and oversight, Patient´s right to self-determination, and Automation bias (see Figure 5. Themes and categories). The Human agency and oversight category discussed using the AI system as a co-operational tool and the system user having enough knowledge on how the system works to maintain human agency. The machine and human brains work in very different ways, and the human oversight and human cognitive thinking are necessary to obtain the best possible care for a patient. The Patient´s right to self-determination category looks at the issue from the patient´s point of view. Automation bias category concerns the scenario where the AI system-driven output might lead physicians to make decisions based on that output rather than based on their professional experience, which might then cause harm to patients. Although the automation bias could have been placed under the Principle of fairness with other biases (see Figure 5. Themes and Categories), the decision was made to connect it to the Principle of human autonomy, as this topic has so much to do with the system users´ awareness of this risk.

6.4.1 Human agency and oversight

When talking about medicine, care practices are often very intricate as they do not depend on only one or two fixed factors. For this reason, it is possible that correct diagnosis and best care practices are controversial. (Pesapane et al. 2018). People have different personal values, they have different socioeconomic status, and they come from distinct cultural backgrounds. These things affect the choice of right care practices. For example, glossectomy is a good operation for someone who values long life span, but bad for someone who values culinary experiences very high and feel that they bring them reason to enjoy life. (Arambula and Bur 2019.) Despite black box issues (see. Chapter 6.2.1 Black box), human agents still need to assess the output of a diagnostic AI system and decide against it if necessary (Starke et al. 2020). Unlike physicians, a trained algorithm is not flexible enough to account for conceptual changes. For example, psychiatric conditions are dependent on the societal context, such as change of eating habits or hours spent on using smart devices, so even tested and approved systems might require overhauling and retraining in addition to human oversight. (Starke et al. 2020.)

Omissions errors easily occur when working in fast environments, such as radiology image reading, when people disregard the failure of an AI tool (Geis et al 2019). If diagnostic AI systems are proven to repeatedly outperform human professionals, should they substitute human decision-making or still just assist healthcare professionals, ask Ienca and Ignatiadis (2020). On the other hand, although decisions made by diagnostic AI systems would have been demonstrated to be best standard of care (Ienca and Ignatiadis 2020), an erroneous diagnosis is particularly worrisome if the system´s decisions are easily accepted by clinicians or if the diagnostic process is fully automated, Starke et al. (2020) point out.

Like with assisted driving, although a machine makes decisions, clinicians need to take charge as backup by checking the recommendations and comparing them to their own professional experience. Disagreements between the system output and the physician´s opinion could be taken to consultation with another physician. Instead of using AI tools for automated diagnoses, they can be seen as assistive tools aiming at improving certainty, and they can serve as a second opinion to confirm the clinician´s opinion. (Starke et al. 2020.) For instance, considering how difficult it is to diagnose schizophrenia and how much disagreement there is among experts, using a diagnostic algorithm to support decision-making could increase the likelihood of patients receiving correct diagnosis and adequate treatment. (Starke et al 2020.) But how can clinicians secure their autonomy when using diagnostic AI systems, ask Arambula and Bur (2019)? Regulations are needed to ensure that system users have a sufficient level of understanding of the system to use it safely (Arambula and Bur 2019).

6.4.2 Patient´s right to self-determination

Patient´s right to self-determination means that patients have the right to make their own decisions without being pressured or influenced from the outside. In the context of using diagnostic AI systems, it means that patients have the right to know and decide if AI is being used in their diagnosis. There are patients that value human providers of care high and for that reason want to refuse from the usage of AI tools in their care. Other reasons for declining the use of AI tools might be fear, lack of trust, and suspiciousness towards new technology. This creates an ethical problem if clinicians are convinced that using the AI system would improve the outcomes of the care. (Arambula and Bur 2019.) According to Brady and Neri (2020), patient´s right to self-determination needs to be paramount when deciding whether using AI or not in their care. To tackle this ethical issue, physicians need to be able to give enough information on AI technology, its risks, and its benefits to patients (Arambula and Bur 2019).

6.4.3 Automation bias

Automation bias refers to the human tendency to approve the outcome produced by a computer, although it would be erroneous (Geis et al. 2019; Carter et al. 2019). Human and AI system decision-making processes are different, as the human decisions are based on experience and expertise, whereas the AI system´s decision-making is based on objective evidence provided by its training dataset. Clinical autonomy can be threatened as system output might affect decisions. (Arambula and Bur 2019.) According to Geis et al. (2019), automation bias generates errors of omission and commission.

When a system user follows the flawed recommendations of an AI system and ignores other evidence and their own experience, harm might be caused to the patient (Geis et al. 2019). Grote and Berens (2019) refer to this situation as AI challenging the epistemic authority of clinicians which then leads them to make decisions defensively. Especially busy working environments predispose clinicians to automation bias, claim Arambula and Bur (2019). For example, research is showing decrease in diagnostic accuracy, when clinicians view erroneous imaging data produced by a machine. To avoid automation bias, a diagnostic system should not be over-automated. Clinicians also need training so that they are able to understand automation bias and recognise situations that expose them to this ethical risk. (Carter et al. 2019.)

Unlike all the other articles that were selected to this review, Heinrichs and Eickhoff (2019) do not bring up automation bias. On the contrary, they refer to psychological bias of self-centrism, which makes people rely more on their own perceptions than quantitative confidence ratings into which the recommendations of AI systems could be incorporated (Heinrichs and Eickhoff 2019).

7 Discussion

The findings of this review show that there are many ethical issues to consider when using AI in healthcare diagnostics. The content analysis led the reviewer back to the theoretical framework of this review as the ethical issues recognised from the reviewed articles quite naturally fell under the same main themes of the four ethical principles of

Ethics guidelines for trustworthy AI defined by the European Commission (2019: 14). The realization of how everything comes together to the same point where it all kind of started was interesting.

The findings point out clearly that ethical issues regarding diagnostic AI systems concern all the stakeholders around the systems, including the system developers and the companies that produce diagnostic AI tools, the organizations that purchase the systems and take them into service, the clinicians who use the systems to diagnose patients, and the patients themselves. Also, the society can be considered a stakeholder as an ethical and trustworthy diagnostic AI system can have a bigger impact than just at the individual level. The level of responsibility considering ethical issues is minor for patients compared to, for example, system developers and system users, but patients are the ones who most concretely experience the consequences. The code of medical ethics leads the work of healthcare professionals, but system developers are not required to put patient´s interest first (Carter et al 2019). This is primarily an ethical issue of intention and the motive behind building diagnostic AI, but it also plays a role when discussing accountability. The evidence points out the importance and necessity of an ethical code and requirements for its deployment and governance for the field of AI as it already exists in healthcare and obligates people working in healthcare professions. This kind of regulation would improve the consideration and understanding of ethical issues within the people who develop AI solutions and other stakeholders, ultimately resulting in ethically sustainable AI systems.

7.1 Ethical issues of developing diagnostic AI

European Commission (2019: 13) underlines how important it is to remember the role of a human as moral operator when developing and using AI in order to maintain and respect human dignity. A machine cannot act as a moral agent, a human must take that role, and it is up to the human to teach the system to work in an ethical manner and take care of its updating and surveillance. Only ethical humans can build ethical AI.

It is crucial from the ethical point of view to consider what kind of data is used to train algorithms. The benefits that diagnostic AI systems provide should not discriminate anyone, and to avoid that, training data should represent different groups of people, paying special attention to potentially vulnerable groups such as children, women, ethnic minorities, disabled people etc. (European Commission 2019: 11). Arambula and Bur (2019) ask a relevant question of what companies who develop diagnostic AI systems can actually do to ensure equal and non-discriminative access to benefits of their system.

The reviewed evidence repeats once and again how problematic it is that the existing health data is already biased historically until present because of biased people´s mindset that is the fruit of the patriarchal, unequal society we live in (Ahmad et al. 2019; Arambula and Bur 2019; Brady and Neri 2020; Carter et al. 2019; Geis et al. 2019; Ienca and Ignatiadis 2020; Jaremko et al. 2019; Pesapane et al. 2018; Starke et al. 2021). Clinical utility cannot be regarded as the most important criterion when assessing diagnostic AI systems, because in doing so, it falsely leads us to justify pursuing maximal benefit for the majority and override the needs of minorities (Starke et al. 2021). This is a major ethical issue that needs to be kept in mind when developing diagnostic or any medical AI systems. Ethics guidelines for trustworthy AI (European Commission 2019: 14, 19) endorse careful assessment of the AI system´s impact on the fundamental rights of people at the very beginning of the development process.

Ethics has always been and continues to be a difficult theme as it is not unambiguous. There are always many points of views to ethical issues and so many things to understand and consider. Whereas the most important ethical issues can be divided into groups, they still coexist in practice and are often conflicting. For example, the transparency and explicability of the process leading to the diagnosis are an important part of ethical AI, but the same applies to more accurate diagnostics that at present achieve better patient outcomes using less explainable diagnostic AI systems. Pesapane et al. (2018) talk about finding a balance in the controversy of accuracy and privacy, obtaining a better diagnostic outcome by using more personal health data and still maintaining the person´s privacy. According to Geis et al. (2019), too much transparency on diagnostic AI systems processes can involve a risk on privacy. This creates controversy between privacy and transparency.

Because of these common discrepancies between ethical issues concerning diagnostic AI, it is even more critical to consider ethics thoroughly when building the systems. In order to recognise all the ethical issues and their different dimensions, it is necessary to involve professionals of the medical field in question as well as experts on AI ethics in the system development process.

7.2 Ethical issues of using the diagnostic AI system

It is not only the development of diagnostic AI systems that raise ethical issues, but also their usage. System users must have a proper level of understanding and knowledge on the system, how to use it and how it works, and how it produces its output, despite the opacity of the algorithm´s decision-making process referred to as the black box dilemma (Arambula and Bur 2019; Starke et al. 2020). Clinicians should also be able to explain all this in a comprehensible way to their patients, so that the patients can give their informed consent on using AI system in their diagnostics process (Arambula and Bur 2019; Heinrichs and Eickhoff 2019). That is why it is relevant to consider whether the black box dilemma inside medical AI systems has too much conflict with the common requirement to appropriately inform patients on their care. By using a diagnostic AI system to assist them, clinicians save time on the actual diagnosis process; consequently they have more time to talk with their patients personally, which for many is an important value as well as an ethical perspective to good care. However, the way algorithm works as a black box precludes them from explaining in detail the diagnostic decisions and which things have led to this particular diagnosis, like people are used to hearing from their clinician (Heinrichs and Eickhoff 2019).

The 2021 Coordinated Plan on Artificial Intelligence (European Commission 2021) underlines that the final decision should always involve human oversight and the evidence of this review agrees with this. Healthcare professionals using diagnostic AI systems and algorithms are trained in distinct ways, and also their reasoning processes differ greatly. In case of "peer"-disagreement, this poses a problem for the clinician, because the AI system does not explain why and how it decided something due to the black box dilemma. (Grote & Berens 2019.) However, it was highlighted in many references used for building the theoretical background of this review that diagnoses produced by AI are not worse than those made by humans; on the contrary, they have often proved even more accurate (Arieno et al. 2019; Bohr & Memarzadeh 2020; McDougall 2018; Miller & Brown 2018). Nevertheless, it is important to recognise that this creates another complex ethical issue that requires reflection. Besides that, automation bias, the human tendency to let a decision made by a machine surpass their own judgement based on experience and knowledge, is another important ethical issue for the system users to be aware of. (Carter et al 2019; Geis et al. 2019). Heinrichs and Eickhoff (2019) present an opposite perspective pointing out that because of opacity of algorithms decision-making process, healthcare professionals might undervalue highly accurate results provided by diagnostic AI systems if they contradict with their experience. Arieno et al. (2019) say that since the output produced by a diagnostic AI system is based on data, the results are always evidence-based. Brady and Neri (2020) do not see this in similar way but instead claim that expecting patients and clinicians to trust results that are impossible to fully explain, is taking us away from the evidence-based medicine.

The findings showed that the current diagnostic AI systems have proved higher level of accuracy when being less transparent (Ahmad et al. 2019; Carter et al. 2019; Grote and Berens 2019; Heinrichs & Eickhoff 2019). The evidence discussed this matter from a few different angles. First, in terms of diagnostic accuracy, it would be an ethical issue for a clinician to decide not to use a diagnostic AI tool at the request of a patient. The ethical dilemma concerns whether to act according to the patient´s request that obviously is based on the patient´s right to self-determination that must be valued very high, or act against it, or try to persuade the patient to change his/her mind to provide the best possible care. (Arambula & Bur 2019.) Secondly, it is to be noted that the algorithm easily fails when moved to another setting (Carter et al. 2019).

Machines and humans both make mistakes. Starke et al. (2020) propose a model of shared responsibility in which competent healthcare professionals review the results of a diagnostic AI system. This would help to build patients´ trust on diagnostic AI systems. But will this invalidate the benefit that diagnostic AI systems provide saving the clinician´s time on routine tasks and leaving them with more time to concentrate on other, more specialised duties where human presence is more valuable?

There is a clear disagreement and difference in opinions about the role of diagnostic AI systems. Starke et al. (2020) express a strong opinion against diagnostic AI tools used as independent, or leading agents in diagnostic decision-making. Instead, they should be used as assistive tools to improve certainty and as second opinion to confirm the judgement of a clinician (Starke et al. 2020). Rusanen and Lappi (2018) agree with Starke et al. (2020) opinion. Ienca and Ignatiadis (2020) in turn challenge this view by appealing to the proved records of AI systems outperforming human clinicians in accuracy of diagnosis, which according to them provides the evidence that diagnostic decisions made by AI systems could be considered as the best practice of care. In their later article, Starke et al. (2021) have a less strict approach, and they suggest that "instead of aiming at a supposedly objective truth, outcome-based therapeutic usefulness should serve as the guiding principle for assessing machine learning applications in medicine".

7.3 Limitations and potential bias

All literature reviews have their limitations and potential for bias, and it is important consider and evaluate them. The purpose of this chapter is to examine in general which factors in the process of making a review may have caused bias in the results. (Stolt et al. 2016: 32-33.)

Careful planning and structural implementation of the review strategy were strengths of this review. The review question was clearly and carefully formulated; the exclusion and inclusion criteria and the search strategy were designed to support the review question. The search of evidence was conducted in a well-planned and structural manner to avoid bias in the process of selecting the evidence, and proper tools were used to assess the quality of the selected articles to ensure good quality of the gathered information. However, the quality assessment was realised by one person only, which may impair the reliability of the review. Contrary to general scoping review conventions, grey literature was not included in the search of evidence. For that reason, some interesting and relevant information about the topic might have been left out from this review. The fact that this review was done by one author, may have also limited the search of evidence and lack of peers assessing the evidence introduces risk for bias. Besides that, the author had no previous experience on doing a scoping review. Help of expert librarian was used to choose the appropriate databases, to find proper search terms, and to realise the searches. This brings reliability to the search of evidence.

The available evidence on the topic set its limitations to the review. Most of the research discusses the topic of ethics of AI or ethics of medical AI in general terms, and evidence that concentrated on ethics of AI in healthcare diagnostics setting was very limited. If there had been several authors writing this review, the search could have been less exclusive to evidence specifically discussing ethics of AI in medical diagnostics setting. Also, grey literature could have been included into the search. All the evidence that was selected for this review, were review articles, editorials and other similar design of text and opinion. The search of evidence that was done in 2021 revealed that empirical research on this topic is basically non existing. After that some empirical research has been published, but it is not included into this review. The fact that the writing process was prolonged so that the review is published more than a year after the search of evidence was conducted may have impacted the accuracy of the review.

7.4 Ethical questions

The responsible conduct of research, introduced by Finnish Advisory Board on Research Integrity (2012), was followed during this review process. The research of the evidence, its charting and analysis, and reporting of the results were conducted meticulously with honesty and accuracy. Other authors´ sources have been referenced consistently and appropriately, following the referencing guidelines of Metropolia University of Applied Sciences that require distinguishing anyone else´s text from the author’s own text. The originality of this review has been verified using Turnit’s plagiarism detection system. As a scoping review of already published literature, it was not necessary to apply for a separate permission for the research. No sources of funding or other relevant interests have been involved in making this scoping review.

8 Conclusions

ETENE (2001) defines that ethics helps us to make choices in life and guide us to evaluate our actions and the reasons behind them. It is easy to apply this to diagnostic AI and see the value and importance of understanding, considering, and acting on ethical issues when building and using diagnostic AI systems. It is very difficult for a diagnostic AI system to be fully ethical. This relates to the fact that many ethical issues are interconnected, and many times they are also in conflict with each other.

Despite the challenge of the task, it is possible to take control over the ethical issues of diagnostic AI systems. First, it requires a multi-professional team including system developers and other representatives of the company who is building the system, healthcare professionals representing that specific field of medicine in which the diagnostic system is meant to be used in, and an expert of ethics of AI. Sharing knowledge between these stakeholders, also decreases the possibility of personal bias of system developers transferring into the algorithm. Second, at the beginning of the process there's a need for a thorough evaluation of the impact the diagnostic AI system will have on the fundamental rights of people. Third, the quality of the training data must be assessed cautiously to avoid bias. After that, the system users need to be trained properly on how to use the system. It is also necessary to make sure they have an adequate level of knowledge and understanding on the technical side of the system and how the algorithm works and makes decisions. Also, it is important to provide the system users with a sufficient level of information about the possible ethical issues concerning the usage of a diagnostic AI system. System users need all this knowledge to be able to maintain human oversight and agency over the machine, so that they can make decisions on whether the diagnostic AI is being used or not. They also need to be capable of evaluating the output of the system. Furthermore, they have to be able to explain to their patients at a sufficient level and in a comprehensible way the functioning of the system and the output system provides in order to protect the patients' right for self-determination. Moreover, to assure the safety of the diagnostic AI system, it must be maintained and updated regularly. Also, if the system is transferred to another setting, it must be ensured that the training data represents people in that area.

Finally, it is very important that specific rules and regulations are created to ensure that diagnostic and any medical AI is being developed and used ethically. Companies that develop diagnostic AI should have proper procedures on the governance of AI ethics. Also, regulations should guide the system users to work with diagnostic AI tools in an ethical way. Ethics discussion must be an ongoing and fixed part of the development process of new diagnostic AI systems. There is definitely a need for laws, general rules and regulations to guide the ethics of AI.

The author´s recommendation is that a multi-professional team, as described above, should be established right at the beginning of the planning process of a new medical AI system. It is obvious that not all companies have ethics experts among them, and that is why it would be beneficial to use external consultants with expertise in ethics of AI.

The search of evidence for this review revealed that empirical research on the topic of ethics of diagnostic/medical AI is practically totally non existent. The author finds this as one of the main findings on this topic. In conclusion, in addition to well-defined ethical guidelines and instructions on ethical AI governance, empirical research is needed for further development of the ethics of diagnostic AI. Empirical research could be done, for example, in two steps: first, by studying how the governance of AI ethics is implemented in companies that develop diagnostic AI systems and in healthcare organizations where they are being used; and second, assessing the effectiveness of the measures in place. This could help understand which measures or policies are useful in practice, and as a consequence, it would facilitate the creation of general guidelines for the ethical use of AI in medical environment.

Reviewed articles

1. Ahmad, O., Stoyanov, D. and Lovat, L. (2020). Barriers and pitfalls for artificial intelligence in gastroenterology: Ethical and regulatory issues. *Techniques and Innovations in Gastrointestinal Endoscopy.* 22(2). 80-84.

2. Arambula, A. and Bur, A. (2020). Ethical Considerations in the Advent of Artificial Intelligence in Otolaryngology. *Otolaryngology–Head and Neck Surgery - SAGE Journals* 162(1), 38-39.

3. Brad, AP. and Neri, E. (2020). Artificial Intelligence in Radiology-Ethical Considerations. *Diagnostics*. Basel. 10(4), 231.

4. Carter, S., Rogers, W., Win, KT., Frazer, H., Richards, B. and Houssami, N. (2020). The ethical, legal and social implications of using artificial intelligence systems in breast cancer care. *Breast*. 49, 25-32.

5. Geis, J., Brady, A., Wu, C., Spencer, J., Ranschaert, E., Jaremko, J., Langer, S., Kitts, A., Birch, J., Shields, W., van den Hoven van Genderen, R., Kotter, E., Gichoya, J., Cook, T., Morgan, M., Tang, A., Safdar, N., and Kohli, M.(2019). Ethics of Artificial Intelligence in Radiology: Summary of the Joint European and North American Multisociety Statement. *Journal of the American College of Radiology*. 16(11) 1516-1521.

6. Grote, T. and Berens, P. (2020). On the ethics of algorithmic decision-making in healthcare. *Journal of Medical Ethics.* 46(3), 205-211.

7. Heinrichs, B. and Eickhoff, S. (2019). Your Evidence? Machine learning algorithms for medical diagnosis and prediction. *Human Brain Mapping.* 41(6), 1435-1444.

8. Ienca, M. and Ignatiadis, K. (2020). Artificial Intelligence in Clinical Neuroscience: Methodological and Ethical Challenges. *AJOB Neuroscience.* 11. 77-87.

9. Jaremko, J, Azar, M., Bromwich, R., Lum, A., Alicia Cheong, LH., Gibert, M., Laviolette, F., Gray, B., Reinhold, C., Cicero, M., Chong, J., Shaw, J., Rybicki, FJ., Hurrell, C., Lee, E. and Tang, A. Canadian Association of Radiologists (CAR) Artificial

Intelligence Working Group. (2019). Canadian Association of Radiologists White Paper on Ethical and Legal Issues Related to Artificial Intelligence in Radiology. *The Canadian Association of Radiologists Journal.* 70(2), 107-118.

10. Pesapane, F., Volonté, C., Codari, M. and Sardanelli, F. (2018). Artificial intelligence as a medical device in radiology: ethical and regulatory issues in Europe and the United States. *Insights Imaging.* 9(5), 745-753.

11. Starke, G., De Clercq, E. and Elger, B. (2021). Towards a pragmatist dealing with algorithmic bias in medical machine learning. *Medicine, Health Care and Philosphy.* 24(3), 341-349.

12. Starke, G., De Clercq, E., Borgwardt, S., and Elger, B. (2021). Computing schizophrenia: Ethical challenges for machine learning in psychiatry. *Psychological Medicine,* 51(15), 2515-2521.

References

Abàmoff, M., Tobey, D., and Char, D. (2020). Lessons Learned About Autonomous AI: Finding a safe, efficacious, and ethical path through the development process. *American Journal of Ophthalmology,* 241, 134-142.

Acosta, S., Garza. T., Hsu, H. and Goodson, P. (2020). Assessing quality in systematic literature reviews: A study a novice rater training*. SAGE Open*, 19(3)

Anom, B., (2020). Ethics of Big Data and Artificial Intelligence in Medicine. *Ethics, MedicineMedicine, and Public Health,* 15(4), 100568

Arieno, A., Chan, A. and Destounis, S. (2019). A Review of the Role of Augmented Intelligence in Breast Imaging: From Automated Breast Density Assessment to Risk Stratification*. American Journal of Roentgenology,* 212(2), 259-270.

Arksey, H. And O´Malley, L. (2005). Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology*null, 8(1), pp. 19-32.

Aromataris E, Fernandez R, Godfrey C, Holly C, Khalil H,. and Tungpunkom P. (2020) Chapter 10: Umbrella Reviews*. In: Aromataris E, Munn Z (Editors). JBI Manual for Evidence Synthesis*. <https://doi.org/10.46658/JBIMES-20-11>

Baethge, C., Goldberg-Wood, S. And Mertens, S. (2019). SANRA—a scale for the quality assessment of narrative review articles. *Research Integrity and Peer Review*, 4(1), 5.

Bartneck C, Lütge C, Wagner A, Welsh S (2020). *An Introduction to Ethics in Robotics and AI.* Springer.

Bartoletti, I. (2019). AI in Healthcare: Ethical and Privacy Challenges. *In: Riaño, D., Wilk, S., ten Teije, A. (eds) Artificial Intelligence in Medicine. AIME 2019. Lecture Notes in Computer Science, vol* 11526. Springer, Cham. 7-10. Bartoletti, I. (2019). *AI in Healthcare: Ethical and Privacy Challenges*. Artificial Intelligence in Medicine. AIME.

Bengtsson, M. (2016). How to plan and perform a qualitative study using content analysis. *NursingPlus Open,* 2, 8-14.

Bitkina, O., Kim, H. and Park, J. (2020). Usability and user experience of medical devices: An overview of the current state, analysis methodologies, and future challenges. *International Journal of Industrial Ergonomics*. 76.

Bohr, A. and Memarzadeh, K. (2020). *The rise of artificial intelligence in healthcare applications.* Artificial Intelligence in Healthcare, 25-60.

Borenstein, J., Howard, A. (2020) Emerging challenges in AI and the need for AI ethics education. *AI Ethics* 1, 61–65

Cambridge Dictionary. <https://dictionary.cambridge.org/dictionary/english/ethic > Read 13.3.2021.

Davenport, T. and Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future healthcare journal,* 6(2), 94-98.

De Fauw, J., Ledsam, JR., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., Askham, H., Glorot, X., O'Donoghue, B., Visentin, D., van den Driessche, G., Lakshminarayanan, B., Meyer, C., Mackinder, F., Bouton, S., Ayoub, K., Chopra, R., King, D., Karthikesalingam, A., Hughes, CO., Raine, R., Hughes, J., Sim, DA., Egan, C., Tufail, A., Montgomery, H., Hassabis, D., Rees, G., Back, T., Khaw, PT., Suleyman, M., Cornebise, J., Keane, PA. and, Ronneberger O. (2018). Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature Medicine,* 24(9), 1342-1350.

EIT Health (2019). A*I and health: Reaping rewards while addressing ethical concerns*. <https://eithealth.eu/view/ai-and-health-reaping-rewards-while-addressing-ethical-concerns/>

Erlingsson, C. And Brysiewicz, P. (2017). A hands-on guide to doing content analysis. African *Journal of Emergency Medicine*, 7(3), 93-99.

ETENE - The National Advisory Board on Social Welfare and Health Care Ethics. (2001). *Terveydenhuollon yhteinen arvopohja, yhteiset tavoitteet ja periaatteet*. https://etene.fi/documents/1429646/1559098/ETENE-julkaisuja+1+Terveydenhuollon+yhteinen+arvopohja%2C+yhteiset+tavoitteet+ja+periaatteet.pdf/4de20e99-c65a-4002-9e98-79a4941b4468/ETENE-julkaisuja+1+Terveydenhuollon+yhteinen+arvopohja%2C+yhteiset+tavoitteet+ja+periaatteet.pdf Read 20.2.2021.

European Commission, AI HLEG. (2019). *8.4.2019. Ethics Guidelines for Trustworthy AI*. <https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html>

Geras, K., Mann, R. and Moy, L. (2019). Artificial Intelligence for Mammography and Digital Breast Tomosynthesis: Current Con-cepts and Future Perspectives. *Radiology,* 293(2),. 246-259.

Gerke, S., Minssen, T. and Cohen, G. (2020). Chapter 12 - Ethical and legal challenges of artificial intelligence-driven healthcare. *In: A. BOHR and K. MEMARZADEH, eds, Artificial Intelligence in Healthcare.* Academic Press, 295-336.

ISPN – University of Bern (2009). STROBE Statement. <https://www.strobe-statement.org/index.php?id=strobe-home> Read 10.4.2021

JBI – Joanna Briggs Institute (2017). *The Joanna Briggs Institute Critical Appraisal tools for use in JBI Systematic Reviews - Checklist for Text and Opinion.* <https://jbi.global/sites/default/files/2019-05/JBI_Critical_Appraisal-Checklist_for_Text_and_Opinion2017_0.pdf>

Keskinbora (2019). Medical ethics considerations on artificial intelligence. *Journal of Clinical Neuroscience*, 64(6), 277-282.

Lee, E. Torous, J., De Choudhurym, M., Depp, C. Graham, S., Kim, H., Paulus, M., Krystal, J. And Jeste, D. (2021). Artificial Intelligence for Mental Healthcare: Clinical

Applications, Barriers, Facilitators, and Artificial Wisdom., *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging,* 6(9), 856-864.

Leino-Kilpi, H. and Välimäki, M. (2014). Etiikka hoitotyössä. 8. uud. p. edn. Helsinki: Sanoma Pro.

Levac, D., Colquhoun, H. aAnd O´Brien, K. (2010). Scoping studies: advancing the methodology. *Implementation Science*, 5, 69.

McDougall, R. (2019). Computer knows best? The need for value-flexibility in medical AI. *Journal of medical ethics,* 45(3), 156-160.

Mehta, N. And Devarakonda, M. (2018). Machine learning, natural language programming, and electronic health records: The next step in the artificial intelligence journey? *Journal of Allergy and Clinical Immunology*, 141(6), 2019-2021.

Miller, D. and Brown, E. (2018). Artificial Intelligence in Medical Practice: The Question to the Answer? *The American Journal of Medicine,* 131(2), 129-133.

Murdoch University. Systematic Review – Research Guide. *Using PICO or PICo.* <https://libguides.murdoch.edu.au/systematic/PICO > Read 21.12.2020.

Newman-Toker, D., Schaffer, A., Yu-Moe, C., Nassery, N., Saber Tehrani, A., Clements, G., Wang, Z., Zhu, Y., Fanai, M. aAnd Siegal, D. (2019). Serious misdiagnosis-related harms in malpractice claims: The “Big Three” – vascular events, infections, and cancers. *Diagnosis*, 6(3), 227-240.

Peters, M., Godfrey, C., Mcinerney, P., Munn, Z., Tricco, A. and Khalil, H. (2020) Chapter 11: Scoping Reviews (2020 version). *In: Aromataris E, Munn Z (Editors). JBI Manual for Evidence Synthesis*

Rigby, M. (2019). Ethical Dimensions of Using Artificial Intelligence in Health Care. *AMA Journal of Ethics*. 21(2), 121-124.

Rusanen, A-M. and Lappi, O. (2018). Tekoäly ihmisen kognitiivisena avustajana: Kysymys tiedollisista riskeistä. <https://vm.fi/documents/10623/10841416/Rusanen-Lappi-kognitiivinen-avustaja.pdf/b6d168dd-c79e-59ee-81a2-50ead1c63aaf/Rusanen-Lappi-kog-nitiivinen-avustaja.pdf> .

Stickdorn, M., Hormess, M.E., Lawrence, A. and Schneider, J. (, 2018). *This is service design doing: applying service design thinking in the real world.* O'Reilly Media, Inc.

Stolt, M., Axelin, A. and Suhonen, R. (2016). *Kirjallisuuskatsaus hoitotieteessä*. 2nd edition edn. Grano Oy.

Suomen Koodikoulu (201989). Johdatus tekoälyyn.< https://finna.fi/L1Record/aoe.2>

Topol, E. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25, 44–56.

Toronto, C. and Remington, R. (2020). *Step-by-Step Guide to Conducting an Integrative review.* Cham: Springer International Publishing.

Tutkimuseettinen neuvottelukunta. (2012). *Hyvä tieteellinen käytäntö ja sen loukkausepäilyjen käsitteleminen Suomessa.* 2012. <https://www.tenk.fi/sites/tenk.fi/files/HTK_ohje_2012.pdf> Read 16.1.2022.

Véliz, C. (2019). Three things digital ethics can learn from medical ethics. *Nature Electronics,* 2, 1-3

Data charting

Reference	Country	Aim an purpose	Design	Data and methods	Main results	Quality assessment
Ienca, M. & Ignatiadis, K. (2020) Artificial intelligence in clinical neuroscience: Methodological and ethical challenges	Switzerland	To provide an overview of current AI-driven approaches to clinical neuroscience and an assessment of the associated key methodological and ethical challenges. To discuss which ethical principles are primarily affected by AI approaches to human neuroscience, and what normative safeguards should be enforced in this domain	Review article	Review didn´t have description of data and methods. Yet, references from recent studies from fields of neuroscience, neuroethics, neuroengineering and artificial intelligence were used thoroughly to present and justify the data and findings.	AI holds great potential for human neuroscience, inter alia in diagnostic prediction and diagnostics in neuroscience, but to succeed to help individuals and healthcare in large scale there are critical ethical and methodological challenges to be addressed. Accelerating innovation of AI for the benefit of people in need is also an ethical point given the global burden of neurological and psychiatric disorders. Article highlights five ethical challeges that require attention in neuroscienece context: scientific and clinical validity (blackbox dilemma), accountability , risk of neurodiscrimination, agency and neuroprivacy.	12/12 JBI
Brady & Neri (2020) Artificial intelligence in radiology - Ethical considerations	Switzerland	To discuss some technological drawbacks of AI, certain related ethical issues, and to address potential solutions.	Review article	Review didn´t have description of data and methods. Yet references from recent studies concerning ethics of AI and using AI in radiology were moderately widely used to present and justify the content.	The main challenge and key to success is in anticipation of what may go wrong or could be abused with rapidly evolving AI solutions and to take actions against possible negative outcomes ideally before they happen. It is not enough as a radiologist to know how to use AI systems but also to understand how to implement new technology ethically. Aim should be to develop human- and environment-centered AI (instead of tech-centered). Article presents 4 themes of ethical issues: resource inequality (fair use and access to AI tools must be monitored carefully), liability (who is liable for bad outcomes? doctor, software devoler or the hospital?), conflicts of interest (radiologist making decisions regarding purchase and use of AI tools) and workforce disruption (danger of mass unemployment - radiologist who use AI will replace the ones who don´t .	11/12 JBI

Ahmad et al. (2019) Barriers and pitfalls for AI in gastroenterology: Ethical and regulatory issues	UK	To provide an overview and explore some potential solutions for ethical challenges of usin AI in gastroenterology.	Review article	Review didn´t have description of data and methods. Yet references from recent studies concerning ethics of AI and using AI in gastroenterology were moderately widely used to present and justify the content.	Future success of using AI in gastroenterology relies heavily on the professionals´ ability to carefully consider and address ethical challenges. Large data sets are needed for training and later fine-tuning and calibration of algorithms. Data governance and privacy issues arise form this. Article presents a view that to get the benefits that AI systems can offer to us, most likely we need to reconsider the traditional models of fully informed consent. It suggests a "broad consent" type of policy, where people may concent to secundary use of their data without knowing explicitly all future usage, but still with assurance of responsible and safe useage of their data. Efforts are needed on international level to plan satndards for data privacy, storage, access and security. Without clear rules adoption of innovative solutions might peter out. About issue of autonomy of AI article states that conditional automation where human only interfieres when result is positive or indeterminate is possible in reality of healthcare field. Article present transparency and bias as challenges for AI systems but also states that using AI could ultimately help overcome human prejudice.	10/12 JBI
Starke et al. (2020) Computing schizophrenia: ethical challenges for machine learning in psychiatry	UK	To address the ethical challenges concerning psychiatric apllication of AI early on before they develop more and are taken into clinical practice. To demonstrate that any categorical rejection of the use of AI in psychiatry would be ethically wrong given its potential benefits but that careful evaluation is needed.	editorial	Review didn´t have description of data and methods. Yet, references from recent studies from fields of psychiatry, neurology and bioethics concerning artificial intelligence were used thoroughly to present and justify the data and findings.	Currently there is no established AI apllication in psych.clinical practice and existing applications lack indepth ethical analysis. Different types of ML can raise different ethical challenges in psychiatry. AI tools for psychiatric diagnostics can serve as automated second opinion. This can increase likelihood of patient receiving correct diagnosis and adequate treatment and get it without delays. Ethical challenges include: how to protect sensitive data? erroneus diagnosis causes direct harm but also human clinicians make errors. AI system is not flexible as human to account contextual changes that affect psychiatric conditions (like stress and eating habits). Can using AI affect on patient´s trust to clinician? Can it affect vulnerable patients causing psychological stress and put them in danger (psychosis and paranoid symptoms)? Although AI system can take over some tasks, human needs to remain in charge as a backup and if needed decide against AI. Possibility with disagreement to consultate other clinicians. As a result of AI system´s outcome people might need to take expensive tests such as MRI. is this justified? Any new technique needs to prove its cost-effectiveness. More precise diagnoses might convince policymakers to give more money to mental health. Systematic biases are ethical challenge and to avoid them appropriate supervision strategies for data must be developed.	9/12 JBI

Carter et al. (2019) The ethical, legal and social implications of using artificial intelligence systems in breast cancer care	Australia	To map and discuss ethical issues affecting breast cancer care and possible solutions to those challenges.	Narrative review	Review didn´t have description of data and methods. Yet, references from recent studies from fields breast cancer care, mammography and using AI reading screenings were used moderately widely and well to present and justify the data and findings.	Ai has potential to produce good and bad outcomes. Although AI itself is value neutral, algorithms still encode values either explicitly or implicitly. Example: AI can perform differently with different breast tissue on women from different sociodemographic groups. Data includes conscious and unconsious biases introduced by system developers. Black box issue of explainability. Outcome should be possible to explain but less explainable algorithms seem to be more accurate in clinical practice. Is it possible to have accuracy and explainability together in some way? For this clear expectations for explainability are needed as well as welll established quality assurance process. Deliberate design choices can assure that AI deliver more benefits than harm (case overdiagnosis) and systems learn to discriminate between clinically significant and overdiagnosed breast cancers. Feeding biased data into ML systems produces systematically biased outputs. Human choices can skew AI systems to work discriminatory or exploitatice ways. HC and evidence-based madicine are already biased agains disadvantaged groups because for instance under-representation in the evidence base. Explicit human choise are needed to stop these bias to transfer into AI system. Transferring system that works in one place requires re-training of algorithm with data from the new cohort and environment. Management of human-machine disagreement and delegation of reponsibility for decisions and errors are needed	12/12 JBI
Jaremko et al. (2019) Canadian association of radiologists white paper on ethical and legal issues related to AI in radiology	Canada	To summarize and key issues and to provide a framework for study of legal and ethical issues related to AI in medical imaging, related to patient data, algorithms, practice and opportunities from Canadian perspective.	White paper	White paper didn´t have description of data and methods. Yet, references from recent studies from fields radiliology, using AI in imaging and data privacy regulations were used moderately well to present and justify the data and findings.	Standardized implicit consent for appropriate secondary use of health care data (publicly-funded at least) is crucial for AI innovation and development. The changes in concent policies are happening but to change explicit individual consent requires a guarantee of anonymity, minimal risk associated with data sharing, impracticality of explicit consent and crucially a trusted data custodian. Anonymization of data is crucial but challenging. Article highlights the role of institutions implementing diagnostic AI systems acting as responsible data custodians. Liability issues depend on the level how human and AI system work together and how autonomously system comes to conclusions.	11/12 JBI

Heinrichs & Eickhoff (2019) Your evidence? Machine learning algorithms for medical diagnosis and prediction	Germany	The aim is to address some critical issues concerning using ML for healthcare diagnostics and prediction.	Research article	Review didn´t have description of data and methods. Yet, references from recent studies concerning using AI in medical diagnostics in different relevant fields was thoroughly used to justify and support claims and findings.	It is critically notable that there is generally an inverse relation between the portential accuracy and performance of ML algorithms on one hand and their interpretability on the other (results interpretability and model interpretability). Systems that give the most accurant results are the least transparent ones (black box issue, especially with DL). Lack of evidence makes an assertion suspicious although we had a strong feeling it might be true. Article highlights two interconnected issues: 1. epistemic opacity is at oods with a common desire to understand and potentially undermines information rights. 2. who is responsible in case of failure? Article claims that compatibility with discursive practice is the essential point from ethical perspective. AI systems´ outcomes must include discursive elements or points of contact for linking them to other information and enabeling assessing the information based on evidence. Article calls for empirical investigation on different target groups: what type of opacity people are willing to accept in medical testing and what does it depend on (AI system ordering taking a pill or surgery),	12/12 JBI
Pesapane et al (2018) AI as a medical device in radiology: ethical and regulatory issues in Europe and The United States	Italy	To draw a clear picture of the state of AI regulation in context of medical device and to consider issues of accountability both legally and ethically.	Review article	The description of data and methods was missing but the literature concerning AI regulations of medical devices in Europe and US, using AI in radiology and its ethics were very thoroughly used to justify the arguments of the article.	Using AI in tools to assist radiologists and perform radiological reading has many benefits but also many ethical issues. Accountability is a big issue because of black box issue and unpredictability. AI doesn´t think like humans, but it processess all the different kind of possibilities that exists, which is so much that human cannot process that amount of data. It is also designed to develop and learn from its experiences. In case something goes wrong, who is to be held responsible because it has been impossible to forsee what will happen? The development of AI systems for radiology should follow core ethical principles that have guided field of medicine through history: beneficience and respect for patients. "In the context of evidence-based medicine, the best external evidence has to be combined with patient´s preferences and values". This means that yes AI system can analyse dataand give output but phycisian´s/radiologist´s role is to bring that human touch and take care of quality assurance and - improvement, communication of findings, education, policy-making and many more tasks. Ethical and legal responsibility of the decision-making in radiology will remain on humans and complicated cases should be handled in multidiciplinary boards.	11/12 JBI

Grote & Berens (2019) On the ethics of alhgorithmic decision-making in healthcare	Germany	To lay grounds for further ethical reflection of the opportunities and pitfalls of ML for enhancing decision-making in healthcare. The aim is to demonstrate that the deployment of ML in medicine goes hand in hand with trade-offs on the epistemic and normative level which might cause many ethically non-beneficial effects.	Extended essay	Essey didn´t have description of data and methods. Yet, references concerning ethical issues of using AI in decision-making in HC were used moderately widely and very thoroughly to present and justify the data and findings.	ML solutions can improve accuracy of diagnostics, but it comes at the expense of opacity when trying to assess the reliability of given diganosis. This essay questions the comparability of accuracy of AI and clinicians as in reality clinicians use many kind of evidence to come to conclusion. In case of peer-disagreement it is possible that clinicians make "defensive decisions" in favor of AI because in case of being wrong they can be personally held accountible for it, unlike AI system. Although, explainability issues of AI are currently under work, they will most likely remain in some form. Because of opacity of decision-making the patient doesn´t get information on how the result was achieved and because of this she might not be able to give her consent to treatment. Protecting people´s dignity and autonomy is crucial. For the challenge of attribution of accountability essay suggests possibility to implement less individualistic notion of reponsibility (distributed or collective) because of various stakeholders in the process. Training algorithms with more divere set of data and validating algorithms for different subpopulations can resolve problem of bias and discrimination, but issues of data provacy and sharing must be considered carefully. Another ethical issue the essay raises is the problem of "normative alignment" that means a situation where values of other institutions, countries and world regions might shift to another institution with the training data the system is using. Authors claim that as AI is part of tech industry lead by computer science departments (where as evidence-based medicine is in healthcare field) the engagement of the whole industry entails ethical problem of it own.	12/12 JBI (100%)
Arambula & Bur (2020) Ethical considerations in the advent of AI in otolaryngology	USA	To present ethical questions concerning using AI in otolatyngology and advocating empathic approach to patient care when evaluating these AI tools.	Comment	The description of data and methods was missing but the comment did include relevant references to justify its key points.	Ethical issues associated with AI in healthcare can be monitored through 4 pillars of medical ethics: autonomy (informed consent, patient privacy), beneficience ("good" doesn´t mean same for every patient, but individual factors affect this), nonmalefience (development, validity and safety of programs are in key role) and justice (fair treatment and distribution). Although, the comment is not wide it does raise interesting points and examples to the theme.	10/12 JBI

Geis et al. (2019) Ethics of AI in radiology: summary of the joint European and North American multisociety statement	Europe, USA, Canada	The aim is to inform a common interpretation of the ethical issues related to using AI in radiology and to inspire radiology AI´s builders and users to enhance radiology´s intelligence in humane ways to promote just and beneficial outcomes while avoiding harm to those who expect the radiology community to do right for them.	Condensed summary of an international multisociety statement	The paper didn´t include description of data and methods, but it used recent studies of the using AI in radiology and its ethical challenges moderately widely and thoroughly used it and the expertise of the authors to justify the results and claims.	Radiologists´ understanding of ethical issues and their response to them shift constantly. That´s why it is their moral responsibility to to consider the ethics of how they use sensitive patient data and how they operate and build AI systems helping with decision-making. Users of AI system must understand how it works. Ethical issues concern; ethics of data (privacy, informed consent, protection, ownership, objectivity, transparency, accuracy), ethics of algorithms and trained models (decision-making, biased data sets, fairness and equality that are responsibility of humans!,appropriate level of transparency vs. opacity of the outcome, safety, accountability, liability, explainability vs. black box), ethics of practice (automation bias = humans favoring decisions generated by a computer, commission errors, liability issues, fair access to technology). AI developers need to have the same standard as phycicians of "do not harm". The paper urges radiology community to develop codes of conduct of ethics and practice for AI that promote use that helps and creates good and blocks opposite and usage for financial gain. It also states that reconsideration of ethical issues on this topic must be done continuously as new ethical issues rise while technology develops and our appreciation of these issues change over time when we see these systems in practice and learn more about their impact, benefits and challenges in a long run.	11/12 JBI
Starke et al. (2021) Towards a pragmatic dealing with algorithmic bias in medical ML	Switzerland	The aim is to guide readers (people working with medical AI tools) to relate the issue of bias through pragmatism and use the outcome these tools provide as guiding principal to assess them.	Bulletin of the WHO	Review article	One of the key ethical challenges with medical AI tools is training them woth biased data and because of that replication and reinforcement of existing discriminatory practices. Some of these bias can be traced but in case of some it is impossible. Training AI systems with already biased data and the reasons hidden deep into algorithms can lead to medical practices being naturally even more discriminatory and more difficult to address. On the other hand in some cases e.g. gender and ethinicity are relevant information. Distinguishing between these different situations is challenging but very imortant. Better transparency of the system makes it easier to detect bias in algorithm and correct them. The writers argue that in context of medical ML accurate diagnostis and treatment are in priority over explainability. "It is useful because it´s true" "it is true because it´s useful". The issue here is how to measure and monitor clinical utility. The biggest problem here is that it could lead to ignoring the needs of minorities when maximizing the benefits of the mayority and this is why algorithmic fairness must be carefully considered. Ethically it is right to prioritize the ones who are most vulnerable. Ground truth in case of medical AI is probably unobtainable. Medical AI systems require empirical testing regarding fairness to recognize and assess the possible bias and to know which information to use and how in training algorithm.	12/12 JBI

Quality Assessment (JBI 2017)

Brady & Neri 2020	2	2	2	2	2	1	11
Ienca & Ignatiadis 2020	2	2	2	2	2	2	12
Ahmad et al 2019	2	2	1	2	1	2	10
Starke et al. 2020	2	2	2	2	2	2	12
Carter et al. 2019	2	2	2	2	2	2	12
Heinrichs & Eickhoff 2019	2	2	2	2	2	2	12
Jaremko et al. 2019	2	2	2	2	1	2	11
Grote & Berens 2020	2	2	2	2	2	2	12
Geis et al. 2019	2	2	2	2	2	1	11
Starke et al. 2021	2	2	2	2	2	2	12
Arambula & Bur 2020	2	1	2	1	2	2	10
Pesapane et al. 2018	2	2	1	2	2	2	11

1. Is the source of opinion clearly identified	Yes	2
2. Does the source of opinion have standing in the field of expertise?	No	0
3. Are the interests of the relevant population the central focus of the opinion?	Unclear	1
4. Is the stated position the result of an analytical process, and is there logic in the opinion expressed?	Not applicable	
5. Is there reference to the extant literature?		
6. Is any incongruence with the literature/sources logically defended?		