Impact of Bias in AI and Human Decision-Making: A Healthcare Perspective

A Master Thesis for MSc Philosophy of Science, Technology, & Society

Student: Kirty Bol

University of Twente, Faculty of Behavioural, Management, and Social Sciences
Enschede, The Netherlands

Date: July 2024

Supervisor: YJ Erden
Second reader: Jan Peter Bergen

Table of Contents

Summary

The development of different AI tools within various fields is going rapidly in the past years, one of the fields included is medicine. Within this field, the AI tools can have a large impact on the way that healthcare professionals work and how patients experience their visit to the physician. Especially when the decision is made to use AI as additional tool for the healthcare professional in the decision-making process, for example when it comes to a patients' treatment or when to discharge.

Important to notice in the decision-making process is the presence of possible bias, from either the healthcare professional or the AI. In this case, I chose to focus on the following kinds of bias: cognitive bias, clinical bias, or computational bias. These different forms of bias can be connected to either human decision-making or AI decision-making and the impacts that could be considered within these domains. Research shows that different forms of bias come with different impacts on human decision-making and AI decision-making, and that the bias can be positive, negative, or neutral. Mostly the bias is present without being noticed by the person that must deal with it, which can be the healthcare professional that talks with the patient or the programmer that is programming the algorithm for the AI tool, among others.

I used case studies on AI for decision-making in acute medicine, AI as co-pilot in healthcare, and Clinical Decision Support System (CDSS) for assisting with diagnosis to show examples of how AI (tools) are used in practice or what the applications in the (near) future will be. The studies showcase the influence when it comes to bias in human decision-making with the usage of an AI tool as addition in the assessment of the patient. But the research does highlight that a good foundation and good data quality are important to build a robust and useful AI tool.

It is important to be aware of the impacts that come along with the presence of bias and that the impacts of bias are context dependent. Therefore, "What might the impacts of bias in AI and human decision-making within healthcare?" can be answered by arguing that the

impact of bias in healthcare can be very diverse in nature and that it is recommended to conduct more research on the topic.

1. Introduction

There are a lot of different applications that can be considered when we start talking about artificial intelligence (AI), but in this case, the focus will be on AI decision-making specifically in the field of healthcare. I chose this based on what I read about the use of AI in assessing when a patient can be discharged from the Intensive Care Unit (ICU). This technology is currently used in the Amsterdam University Medical Centre (UMC) (Elbers & Thoral, 2022). For now, the AI is used as an addition in making the call to discharge patients, while the doctor makes the final decision. Having the patient stay for too long on the ICU can be expensive yet discharging them too soon can worsen their health (Klugt, 2022). The idea is that the AI can help to find the right moment for discharge. However, there is a possibility that in the future the patients discharge would be decided by the AI alone, based on the patient's electronic file, without any input from a doctor in this decision. For now, the AI would help in the decision-making of the healthcare professional and provide the doctor with additional information for the decision-making process in a way that it is beneficial for the patient. Mostly because the AI has access to a large dataset that the AI can easier navigate through the present data than the healthcare professional, that could help to decide the best moment to discharge a patient from the ICU. What was not mentioned within the article, was the possible presence of bias within the AI algorithm. It was also not mentioned if a potential bias has an impact on the healthcare professional's decision-making.

Attention has been devoted to the aspect of bias in combination with AI when it comes to healthcare decision-making, however the attention is limited when it comes to the different shapes that the bias could take in this instance. Within this thesis, I research the impacts of bias in the AI or the human decision-making process. When the AI tool would be used on a great(er) scale, both healthcare professionals and patients experience the dependency on AI as it leaves less room to argue what they think. Even though in most cases it will be argued that the doctor will have the final say, it is possible that in the future this will not always be the case anymore. This highlights the current discussion on the use of AI within healthcare, as shown by the article mentioned earlier and made me curious to explore what relations there might be between the two concepts.

In the scope of this thesis, I will argue that the moment we allow an AI to decide about a patient's health through the patients discharge plans/medical file on its own, there is a chance that the more human aspects of medicine, like care and interaction, will be lost. The aim of the present work is to examine the different aspects next to the human aspects, i.e. care and interaction, that must be considered before deciding that an AI can take over some part of decision-making in healthcare-related instances without human interference (van Wynsberghe, 2022).

Issues that will be considered are privacy, bias, vulnerability, and transparency. Privacy is about who has access to data and how is the data handled. In this case, the research focuses on the healthcare professional, the patient, the AI, and the programmer as the people who have access to the data. Where the healthcare professional has access to all data in an anonymized way, the patient has access to their own data, and the programmer has access to the data that is needed to build the AI or algorithm. With bias, I consider prejudice towards a person or ethnicity that can be positive, negative, or neutral. I will elaborate on bias in chapter 2 and provide the definition that will be used throughout the thesis. Vulnerability comes to play when we consider a patients' medical condition and there could be an influence through their ethnical background. Or when the decision of care would be made on incomplete or incorrect patient data. When we talk about transparency in this case, it is about being transparent towards the patient on what data is used and what happens with the data when the AI is used for decision-making.

Furthermore, AI should be considered an emerging technology, which means that it is a technology where the development and/or practical applications are not yet fully realized and there is no long-term impact study available. Hence the reason that vulnerability and transparency come into play, especially when it comes to altering the outcome of the AI and making it more fitting to the patient's medical condition. It would not be reasonable for a healthcare professional to change the AI decision when it is not clear what data is used and what steps have been taken to come to this decision. The decision will probably also not highlight whether a patient's ethnical background has an influence on the recommended care and could harm the patient if the background is not correctly included. This also applies when the AI analyses data before providing the healthcare professional with different options, as there needs to be a certain transparency on the data that is used in the process.

The uncertainty that arises with the topic of AI decision-making in healthcare plays an important role within this thesis. On one side, the more human aspects within healthcare and their importance will be considered. What will the role of the doctors become when the decision-making lies with the AI and how will it impact the frequency and number of interactions between doctor and patient? There are various case studies that can be connected to the topic which showcase how AI can be used for decision-making within healthcare. The case studies are used to open for discussion on the way that we look at decision-making within healthcare and the changes that might take place in the (near) future. This will be the other side of the research and discussion where the AI is considered as tool to be used within healthcare.

1.1. Research Question

In this thesis, my research question is: "What are some impacts of bias in AI and human decision-making within healthcare?". I have chosen to go for impacts plural, as bias is a broad concept and I expect there to not be one simple impact to be found. It could be possible that there are several impacts of bias to be found and that there is a difference of impact whether we look at AI decision-making or human decision-making. That is why I divide this question into sub-questions:

- The first sub-question: how can bias be defined, and what are the different variants of bias that will be included within this research?
- The second sub-question: how do we define decision-making, and what are the differences between human decision-making and AI decision-making?
- The third sub-question: what can real-life case studies stell us about the role of AI in healthcare decision-making processes?

1.2. Framework and Approach

The main approach that I will be using is evidence-based decision-making after establishing a definition to bias and decision-making. I will use these definitions through the thesis. The arguments I will make will be based on what impacts of bias there will be on human decision-making when we will allow AI to make final decisions in healthcare. The research question will be used as the common thread through the chapters and topics. I will conduct my research through the views of philosophers that have written about unconscious bias and decision-making, such as Szaniawski (1980), Saul (2013), and Brownstein (2020). In addition, I will provide a definition on how I see bias and decision-making within the scope of this thesis. Another important aspect I consider within this topic is how an understanding of ethics within technology is changing based on the development of AI and the different ways that AI is implemented in healthcare. I will not be using a specific ethical theory and therefore talk about ethics through transparency, accountability, and responsibility. In addition to this, I will use the lenses of privacy and vulnerability as described earlier this chapter. These lenses will specifically be used when we look at the real-life cases based on the research that is done through the rest of the thesis. In other words, the different values that are noted and research will serve as the lens through which I will consider the tools and cases that are used in real-life.

One key question I will consider is about responsibility in the context of AI decision-making. This includes research on where the responsibility for decisions made by AI lies and who should be held accountable for the technology. Does it lie with the programmers that programmed the AI or with the organization that has decided to procure the technology and use it within their systems? This thesis will focus on the responsibility that lies with the organization that decides to use the AI for decision-making purposes. I made this decision as the organization plays an important role when it comes to the human aspect within healthcare. The organizations have the responsibility to implement the AI systems in their processes and procedures, the impacts on patient care outcomes, and the influence on the ethical dimensions of healthcare decision-making. Therefore, it is important to understand the organizational responsibility before implementing AI technologies and ensuring that there is clear accountability for the AI outcomes. Responsibility will mainly be an ethical consideration that I will use to conduct my research and will therefore be a limited used concept in the thesis itself. It will be used explicitly when we talk about decision-making in chapter 3 and within the overall conclusion of the thesis.

1.3.Structure of the thesis

To answer the research question of this thesis, I will first look at the different components that I have included in the question. This paper is divided in a total of six chapters with the main components being bias and decision-making, which will be covered in chapter two and three. In these two chapters, I will start by defining what I mean with bias and addressing cognitive bias, clinical bias, and computational bias. For chapter three, this will include a definition on decision-making and the meaning for both human decision-making and AI decision-making. These sub-categories will include definitions and the relations/similarities between them. In chapter four, I will present three case studies to make the research in chapter two and three more tangible and showcase what the impacts could look like in real life cases. The used cases will be: AI for decision-making in acute medicine, AI as co-pilot in healthcare, and CDSS for assisting with diagnosis. After this base has been established, I will answer the main question in chapter five by combining the impacts that can be concluded from the research and case studies together. Answering the research question will also include possible ways to reduce the negative bias and therefore the impacts of bias on decision-making. Negative bias as reducing all bias will be very unlikely to achieve. I conclude by summarizing what the impact of bias is on human decision-making and AI decision-making, and possible ways that AI might be able to help reduce cognitive/clinical bias within healthcare without AI having an impact on the quality of care for the patient. Ifnew biases come up during the research, I will make sure to include them within the conclusion and elaborate on how they fit in in comparison with the one I already included.

2. Bias

People tend to have biases about things, people, or situations without necessarily noticing them. Within natural sciences, the aim is to avoid bias as much as possible, and this should be manageable to a certain level considering most experiments are conducted in a controlled environment. The influence on bias when an experiment is conducted in a controlled environment is that the accuracy of the results improves due to the lack of external influence and relationship between different variables becomes clearer. There is less chance for research bias this way, as all the steps can be traced back to a certain point with the extra layer of accuracy. However, research implies that it is necessary to critically examine philosophical bias rather than simply avoiding it (Andersen et al., 2019). But how is bias defined within this thesis? Here, I will focus on cognitive bias, clinical bias, and computational bias.

2.1. What is bias?

Having a judgement about someone or something without really realising where it is coming from is a typical experience for many people. We can also call this having a bias towards a person or situation, but what is bias? I want to start by highlighting that there is no simple answer to the question of what bias is, but I will use this chapter to introduce the definition that will be used within my thesis when talking about bias to have a coherent understanding between author and reader. I will especially include the main forms of bias related to the field of healthcare; due to the scope of this thesis, no other forms of bias will be included here.

Bias can affect the way that we perceive, evaluate, or interact with people from different backgrounds or groups, according to Jennifer Saul (2013, p. 40). This could be done consciously but also unconsciously based on associated tasks concerning the group or event. Additionally, Hilbert (2012) mentions that bias is formed based on previous experiences. However, Saul highlights that "one might unconsciously associate groups with different flavours of ice cream without this having any negative effects" (2013, p. 40). This gives the input that bias does not by default need to be a negative thing but can have a neutral, or even positive, influence as well.

The definition that I therefore will use for bias is as follows: bias is having a judgement (positive, negative, or neutral) about a person, situation, or thing based on perceived knowledge and/or previous experience(s) whether it is conscious or not. However, there are situations that bias can be hypothetical, i.e. you are speculatively biased against a situation while not having any experience with it. Important to note with this definition is the fact that the bias by default is not marked as positive, negative, or neutral, but that this must be evaluated per instance. With this definition, I combine the definition as given by Hilbert and the way that Saul defines bias. However, my definition does not imply that there is only one form of bias. An overview of some different types of bias that are discussed in this research can be found in the remainder of this chapter.

2.2.Cognitive bias

Human capacity for memory, learning and reason can be achieved through experiences and knowledge gained from various resources, cultural influences, and personal beliefs. The combination of these different inputs can lead to cognitive bias. "Cognitive bias is a preference for or against a person or a group of people. It is the filter through which a person evaluates the world, though they are unaware" (Rauf et al., 2022, p. 89). However, Hilbert (2012) argues that since the brain only has a limited space for processing information, this can make us implement shortcuts in our brains based on our past experiences and knowledge. Hilbert writes,

> "These simple but often effective approximations make us use a representative case instead of the specific one (representativeness). They also make us work whatever first comes to mind (availability), and based on our first thoughts, it turns out that the subsequent mental search process is limited (adjustment and anchoring)" (Hilbert, 2012, p. 212).

It is argued that relying on the representativeness of the first thing that comes to mind may lead to a limited judgment or behavior, fostering inaccuracies and misunderstandings. In some sense, it could be said that judgement or behaviour comes down to choices that we will make based on the information that is present in our minds. This limitation may lead to having a cognitive bias towards something or someone. However, there is a certain accountability in the cognitive bias that we must consider when we have a judgement based on the presented information from earlier experiences. We must think about what the consequences could be after the judgement is shown.

Cognitive bias and implicit bias both include deviations from decision-making processes, but they are not the same. The difference can be found in the way that cognitive bias is a broader concept that covers various biases in decision-making, while implicit bias specifically refers to the bias located in unconscious associations (Holroyd et al., 2017). Therefore, implicit bias is the one that can have influence without our awareness when it comes to behaviours towards someone or something. The implicit bias that people have is derived from the fact that we have a bias which can be reasoned from the shortcuts that we make in our brain based on past experiences and knowledge (Amodio & Mendoza, 2010; Balakrishnan et al., 2018). Understanding the biases enables for a critical evaluation of thought and actions, developing greater awareness in our decision-making processes. The moment that we unconsciously act on a bias, we are talking about implicit bias, and this specifically relates to unconscious attitudes that influence our behavior. Both forms of bias show the possible limitations and complexities in the human decision-making process. By acknowledging the presence of bias, we are better prepared to make fair and informed decisions throughout life.

Brownstein et al. (2020) argues that it is crucial to remember that research into implicit measures began by recognizing that explicit attitudes consistently predict behavior within a small to medium range. This can be supported by recent analyses (Cameron et al., 2012, Kurdi et al., 2019), which however contrasts with Hilbert in the way that the brain makes shortcuts based on past experiences and knowledge. There is an understanding that individual's judgements or behaviours can be deduced from the attitudes formed through their life experiences. This can lead to cognitive bias when presented with a case or event that can be related to a representative case present in the mind. According to Brownstein et al., the mind holds not only past cases but also parts of social judgement and action related to such cases (2020, p. 296). The cognitive bias can therefore be influenced by judgements that are presented by a group or the environment that the person is exposed to on a somewhat regular basis. Someone who lives in environment A can form a completely different bias about an event than someone who is from environment B. The bias as formed by person A can be negative based on a bad experience that they have encountered around them. Simultaneously, person B may have a positive bias as the event can be related to a helpful event. However, it is important to notice that a person is not only shaped by the environment they live in. The person can make choices and can have a different experience, based on other factors and variables. The example only highlights that the choice can be influenced by the environment the person is used to and that there is an uncertainty present as to whether this is the case with the person's choice in the situation. A difficult to resolve uncertainty is to say that everyone from the same environment will have the same positive, negative, or neutral bias. Or put differently, people with the same background or from the same environment can experience different biases.

From the above it seems clear that cognitive bias is present in every person, and the moment that we acknowledge this, we can better examine our thought patterns and investigate our assumptions towards different events. However, I do add that not all cognitive biases may be accessible, as things could be so much taken for granted that a person will not see it as a bias or as something that needs to be analysed. This combined could eventually lead to refining our understanding of the world around us when we are faced with a decision.

2.3. Clinical bias

Cognitive bias is present in every human, and therefore the world of healthcare, and healthcare professionals are also affected by it. A systematic review from 2020 suggests that "cognitive biases were associated with diagnostic inaccuracies in 36.5%–77% of case scenarios" (Rauf et al., 2022, p. 89). This shows that the presence of cognitive bias in healthcare plays a significant role. However, this is not necessarily cognitive bias alone, as healthcare professionals will base their judgements not only on their past experiences but also on textbook knowledge and the results of medical research. Campesi et al. add,

> "Researchers and health professionals should be focused on the person and not solely on the disease, considering psychological health and social events and how they can contribute to the prevention, medicine, and treatments; biological aspects are largely dependent on interactions with environments" (Campesi et al., 2021, p. 11).

This implies that there are cases where the healthcare professional does not always consider the person and, can give advice based on their (textbook) knowledge and experiences with other patients. "In addition, researchers and health professionals should acquire the awareness of implicit biases, which could help to elevate the care through mitigation of personal biases and how to apply intersectionality" (Campesi et al., 2021, p. 11). The moment that healthcare professionals are aware of possible biases, they may be able to minimise the influence of bias on the professional-patient relationship and start to include the personal circumstances of the patient. The human aspects within healthcare, i.e. care and interaction, are important in including the patient rather than only textbook knowledge and medical data and will improve when the awareness about implicit bias is created.

The decisions that a healthcare professional needs to make can still be affected by bias. "Both positive and negative biases impact clinical decision-making; however, negative bias is of particular concern as it can lead to poor patient care and worse outcomes" (Rauf et al., 2022, p. 89). We should keep an eye on both negative and positive bias, but how can you know beforehand if the bias will turn out to be positive or negative? The answer would be that you cannot know for sure and that the best option would be to avoid clinical bias in the best way possible. I argue that this is the best option, as you would not have to consider the nature of the bias in advance. Though this may be hard to achieve: "Healthcare providers are trained to leave their prejudices at the 'patient's door'; however, they must first be aware of those biases" (Rauf et al., 2022, p. 90). The first step would be to find a way to become aware of the biases that could arise and learn to recognise that the mind is forming (or has already formed) a bias. Once the healthcare professional starts recognising them, curiosity might arise about why they are occurring in the first place (Rauf et al., 2022, p. 94). The added benefit of curiosity could be that the healthcare professional will hopefully try to find a way to minimise the interference of clinical bias.

Clinical bias is not the only form of bias I will be researching within this thesis, the last form is computational bias. It is possible for humans to recognize that they are having a bias, within the next part I will explain how computational bias comes to play. Both clinical bias and computational bias play a role when it comes to decision-making within healthcare.

2.4. Computational bias

Bias can also be found in technology and software. “Every software is biased by the decisions made by its programmers and by the very algorithms used as its building blocks” (Kudless, 2023, p.266), which means that there is probably no software to be found that is bias free. In case of technology and software, we talk about computational bias that has an influence on the decision-making by the technology. There are various factors that could lead to the computational bias as “programmers build tools and their associated algorithms into applications based on the target user and the programmers’ own skills and background” (Kudless, 2023, p. 266). The programming of the software is not based on neutral knowledge and is affected by what the programmers already know. Knowledge that the programmer has about what has worked in the past or what satisfied the target users/client.

On AI text-to-image generation, Kudless argues that “beyond the biases of the algorithms, the programmers of these models need to make explicit decisions about numerous other factors that affect the types of images that can be generated” (p. 268). The same would apply for programmers that make programs for use within healthcare. Indeed, it seems clear that computational bias is present in every domain where technology is used for decision-making. Important to notice is that bias is integrated in not only biological ones, but also i.e. digital neural networks. “The training data, algorithms, and users themselves all carry inherent biases that will never completely disappear” (Kudless, 2023, p. 277). Or in other words the connections that are made within a system can be compared to the way that connections are made in the brain and therefore the chance of computational bias completely disappearing is low. Computational bias will influence the decision-making within technology, even if we would manage to completely build the software in a neutral way.

In the book “Algorithms of oppression”, Noble (2018) conducts research to the way that there is bias present within the internet and claims that there is a need for awareness of why the bias exists and who benefits from this. Some might believe that the internet is neutral and that every idea and activity get an equal chance to appear when a search engine is used. However, Noble highlights the fact that Google effectively blocks sites that compete with services from Google to make sure that their properties are at the top of the search list. Which contributes to the search bias as research is showing that 71,33% of the people choose one of the search options that is on the first page, or at most on the second page, of the suggestions (Petrescu, 2021). This research was conducted in 2021, and there are more recent speculations but unfortunately there are no official numbers and Google never publicly discloses the exact percentage of people not clicking further than the first page. This contributes to the search bias and Noble arguing: “there are several cases that demonstrate how racism and sexism are part of the architecture and language of technology” (Noble, 2018, p. 9). These cases that are specific to various groups, such as black women and girls, and low-scholar people, highlight the topics of racism and sexism, which are examples of bias that can, for example, arise on Google. If one company has the power to decide what comes up with a search, it means that they control what people come across first and make it part of the technology’s architecture.

In addition to the research on clinical decision-making, I argue that machine errors come with similar consequences as human errors. A good example of machine error consequences can be found in Noble arguing that “Discrimination is also embedded in computer code and, increasingly, in artificial intelligence technologies that we are reliant on, by choice or not.” (Noble, 2018, p. 1). We are reliant on the technologies as if we trust that they come without a bias and can help us with the medical challenges that require different sources to get to a solution. However, as bias is infiltrated into AI technologies, the technologies can continue discriminating and endangers the quality of care provided.

Addressing computational bias involves not only improving the technical aspects of algorithms but also critically examining the data used for training, identifying potential sources of bias, and implementing strategies to mitigate these biases to ensure fairer outcomes from computational systems. These points can be considered important when talking about decision-making and the difference between AI and human decision-making.

3. Decision-making

During my research on the topic of my thesis, I have seen that the topic is strongly related to decision theory. Important to note is that the research on decision theory is generally not from a philosophical view. However, as I investigated decision theory more, I found various similarities between decision theory and the philosophy of decision making. That is why I will use the base of decision theory to conduct my research on AI decision-making within this chapter and the differences with human decision-making.

3.1.AI decision-making

With the research on bias, a relation with decision-making became clear and, in this case, I relate it to AI decision-making. The research on AI decision-making becomes clearer by highlighting the differences with human decision-making, which is why this chapter includes parts of the decision-making that are different between humans and AI. While researching the changing world of AI, it is necessary to recognize that AI is intertwined with machine learning (ML). Important to understand this interweaving is to have a definition of both AI and ML. Additionally, research on AI shows that the boundaries between AI and ML might become more difficult to grasp as many authors in philosophical literature tend to use the two terms as if they are the same.

A definition for AI is that it is a field within computer science that focuses on creating systems and tools that can perform on a human-like intelligence. Examples of tasks that require this intelligence are recognizing patterns, decision-making, and solving complex issues to high levels. For high levels, I am referring to the fact that humans can perform these tasks to a more advanced level than an animal could. Animals could learn to execute certain tasks but will probably do this on a lower level than a human would. Which is why I am talking about a human-like intelligence for the AI to perform on. The AI systems can learn, in some instances reason, and solve problems, which relates to mimicking human functions with the difference that the AI performs them more efficient or accurate. ML is a sub-category of AI that allows systems and computers to learn from data and make predictions or decisions without explicitly being programmed to do so. The ML algorithms learn from the data, identify patterns, and make predictions or decisions without human interference. Similarities between AI and ML are that both can aim to replicate cognitive human abilities, both use algorithms to process data and makes decisions, and both can be used within applications in different fields including healthcare. It is not necessary the case that human cognition is the goal, but that it is more of an inspiration of what is aspired to achieve with the AI. Besides similarities, there are some differences between the two. As mentioned above, ML is a sub-category of AI, and this includes the fact that AI has a broader range of techniques and approaches than ML including language processing and robotics. Another difference is the way that the systems learn, ML focuses on learning from data and AI allows for rule-based systems or other approaches additionally to learning from data. Within this thesis I will use ML and AI both, but as two separate terms. I do this to show that most of the AI that is used within medicine is mostly just ML based on training data from patients and health related research. Both ML and AI can have the same goal when it comes to showcasing cognitive human abilities, with the difference in the way of learning and their complexity.

Within this section, I will navigate through the interchanging use of AI and ML to show even more how connected and comparable the two are and I will introduce the definition that I will use when talking about AI decision-making. This allows for research that challenges us to think about the nature of intelligence, learning and the philosophical implications of machines that attempt to mimic human cognitive processes. Comparisons between human decision-making and AI decision-making will therefore also be present in this part of the thesis. Eventually, the lines between artificial intelligence and human intelligence will start to blur slightly.

AI has a different way of information processing than a human being, yet they share some things in common. For instance, AI and humans both make shortcuts during information processing. The difference with AI can be found in the fact that the AI is able to easily analyse big amounts of data and see the patterns within very large dataset. This way it can "utilize the data in a predictive or prescriptive sense" (Giuffrida, 2019, p. 441). The predictive or prescriptive sense could be considered somewhat akin to the shortcuts in the human brain. According to Schnapp et al. (2018) there is increasing evidence that shows how mental shortcuts in information processing can contribute to diagnostic errors. Berthet (2020) uses this within his review on the impact of cognitive bias on decision-making and shows that the shortcuts that are made during the processing of new information are present in the creation of bias in the brain. This processed information is used in the decision-making process, and it is not argued that this cannot happen with information processing within AI. The difficulty in recognizing where bias is coming from mainly lies in the fact that recognition of the bias implies that there has been a bias in the process. For AI, this difficulty lies in where the output or decision is exactly from. It might be possible to get to the origin by reverse engineering the entire decision-making process that the AI system has relied on while going through the given data (Giuffrida, 2019, p. 441). However, various researchers (Gillespie, 2014; Polack, 2020) point out that the large datasets and dynamic algorithms make it nearly impossible to reverse engineer the process. Which I argue relates to the awareness that is needed to admit that bias is present within the mental shortcuts that are created during information processing. I have addressed this awareness when talking about bias and the way shortcuts are created in our brain based on past experiences and knowledge.

If in the future researchers would argue that it is possible to do reverse engineering, there needs to be a certainty about the data at the start and the way this has been processed. Additionally, there needs to be assurance that there are no decisions made while processing the original data. A way to achieve such certainty would be by questioning the truth in the data by using various sources to compare what is written/said. A truth in the way that the data has been checked and a conclusion could be made that the used data is objective. For AI decision-making it will mean that there must be a starting point from before the actual AI is built, and from where the system and healthcare professional could start to make decisions.

The relationship between AI and healthcare can be made when we, for example, take a closer look at the electronic health record (EHR) systems that are “rapidly and pervasively adopted within healthcare systems” (Giordano et al., 2021, p. 2) in different countries. The EHR is a digital paper chart of a patient that could include their medical history and it is mostly stored and accessible through a secure electronic system. Important with the EHRs is who has access to the information and how this can be regulated when AI would need access to it, besides the healthcare professional. EHRs are designed in a way to provide the healthcare professional with a centralized database of patient data and offer several benefits over paper records. For example, by being accessible from different places and being able to combine information automatically in a clear way. The EHR contribute to the decision-making process within healthcare and can be seen as a tool that a healthcare professional can use to decide on a patient's treatment.

There are various application areas where AI has potential in the aspects of clinical decision-making processes, of which one of the application areas is risk stratification. Risk stratification can be defined as a technique to use a patient's health status and other factors to systematically categorize them (Dera, 2019, p. 22). Numerous different tools can be included in clinical decision-making and with the addition of AI, the number of tools has been increasing. Other application areas are "patient outcome optimization, early warning of acute decompensation, potential bias in ML, and future medical training" (Giordano et al., 2021, p. 3). With additional research, it could be possible that there are more applications to be discovered. For the scope of this thesis, I limited my research to the different applications as described by Giordano et al. with the additional perspectives of other researchers.

The upcoming sections on AI decision-making will elaborate on three specific application areas: risk stratification, patient outcome optimization, and potential bias in machine learning (ML). I will provide a detailed exploration and explanation of each of these areas. These topics have been chosen based on the direct connection to human decision-making and the application's role within the domain of healthcare. In addition, these topics have to do with the human aspects, responsibility, and transparency towards the patients.

The application of risk stratification can help identify high-risk patients and optimize preoperative decisions through categorising the patients' health and other medical factors. Different methods, e.g. including subjective data (Dera, 2019, p.26), can be used to identify the risks that could influence decision-making, but research shows that the methods should be used cautious as many are too broad or lack precision on the patient level (Dera, 2019; Giordano et al, 2021). A lack in precision can have negative impacts on the human aspects, i.e. care and interaction, in the way that the method is not patient specific and result in the wrong conclusion/care. Another downside of these methods could be that a trained physician is required to review the records and assess the risks (Giordano et al, 2021, p. 4) to keep the responsibility of providing the right care. This would mean that either all healthcare professionals need to get educated on all the possible risks or that the trained physician always needs to be available to the healthcare professional when needed.

It is possible to use ML within the risk stratification, “In perioperative (period of a patient's surgical procedure) medicine, ML can maximize the benefits of technology to provide safe, timely, and affordable healthcare.” (Giordano et al, 2021, p. 4). ML has a certain way of obtaining new knowledge and this is mostly done through the continuous training process within the system. The program needs to be able to make rapid changes when there is a new reasonable recommendation to be rendered. However, a disadvantage of this approach is that the outcome is not patient-specific, by which I mean personalised, and based on the underlying data that is used to develop the ML. This generalization might cause inaccuracies of specific health conditions, resulting in individual patient outcomes varying from the broader predictions as applied to patients with the same health issue. The use of ML is mostly interesting to implement when the system comes up with recommendations that fit a patient’s profile best and give the possible outcomes. The healthcare professional would still have the final say, as “Computer assistance can only facilitate the work of physicians, not replace it” (Meskó, 2017, p. 129). A more patient-specific approach would allow for better individual care to meet the unique needs of the patient and could theoretically benefit more patients. This also allows for more transparency and less vulnerability towards the patient as it can clearly be shown what patient specific condition a recommendation is made on.

The moment that different patients are seen as the same, can result in a kind of generalisation. Generalizing patient care brings this research to the possible bias that can occur in ML. An example of this bias can be deducted from the work of Weber et al. (2017) where they researched filtering for patients with “complete” EHRs. One of the results was the introduction of bias towards older patients, in particular female patients, as they more often had their file labelled as complete. This way the data excluded a large portion of the population that where not older patients mostly identified as female, which supports the occurrence of possible bias in ML. A way to avoid this would be to argue that fully complete data is essential and that the usage of filters, as in the research of Weber. et al (2017), can minimize the risk of missing data. These filters can help to include more files as they show which data is needed and enough to fulfil the "complete data" description in different scenarios. However, there is still an importance in addressing biases within EHR data management. And the definition of a "complete EHR" could be changed into an as complete as possible EHR, as it is not certain that you know every detail of a patient and can cover everything in an electronic file. Especially when the visual appearance of a patient could play a role in the follow-up steps of care.

The example on using "complete" EHRs shows a harmful, i.e. the misdiagnosing of certain patient groups, outcome of bias that occurs when general data is used instead of specifically looking at what criteria need to be fulfilled to get a complete overview. General data outcome can lead to unintended bias towards a patient and can pass without notice, as the advice might be the right one in the bigger part of the patients' cases.

The aspect of patient outcome optimization is added as "optimization is vital to the clinical decision-making process and the ensuing patient care" (Giordano et al., 2021, p. 4). Which argues that optimization benefits the quality of patient care. I argue that the benefit will arise when the optimization is done in a way that applies to all patients individually and is based on data from patients with different ethnic backgrounds and genders. I chose these categories specifically as they can be illustrative in the way that the bigger group of people can imagine what to include or what is meant when talked about ethnic background and gender. This could be done by adding criteria to the available data that it should consist of percentages of patients with different ethnic backgrounds and genders based on percentages in the population. The limitations of this proposal, however, can be found in the available data as it is not guaranteed that the health data is available for all kinds of patients.

A limited data set could result in not being able to add data in the same percentages as the people present in the overall population.

The use of EHRs within healthcare systems in various places has “created vast repositories of personalized data sets that are perfectly fitted for AI to examine, develop, and predict upon” (Giordano et al., 2021, p. 7). The ideal fit of the EHRs for AI contributes towards risk stratification, limiting bias in ML, and patient outcome optimization as discussed. However, I argue that this perfect fit will only apply when the EHRs are all complete and can show bias towards specific groups of patients. “Artificial intelligence may produce entirely new solutions for tackling global health issues” (Mésko, 2019, p.3) therefore enhancing patient care for the better. The introduction of new solutions could contribute to the completion of the EHRs and the way that patient care takes shape. The most important is still the patient; thus, the way that AI produces solutions to use for their care should be centred around them, and we must not forget this when we introduce AI in patient care.

Reflecting on the relation between AI decision-making and human decision-making, I argue that AI's remarkable capabilities to analyse vast datasets and distinguish patterns can be compared to the cognitive shortcuts human brains use. However, I do admit that the comparison is more a metaphor than actual being true, which is related to the argument by Erden on the fact that the comparison is by analogy rather than by necessity (Erden, 2021, p.25). In other words, identifying the exact origins of AI decision-making can be as complex as retracing the steps of our own thought processes. Problems to consider with this complexity are that what counts for the human brain cannot by default be used for the AI. The analogy can be used, while the origin in AI decision-making can be found in the used data. For the human decision-making, experience and environment are additional fields to consider besides the data. Both have a beginning and somewhere started to create shortcuts based on the processed information to keep it, among other things, easily accessible.

4. The impacts through case studies

I present three practical examples on the way that AI is used within healthcare and the combination with decision-making in medicine. In this chapter, I show the examples and explain the ethical issues within these technologies that can be related to the impacts of bias on decision-making.

These cases are added to create the connection between the theoretical part of using AI in healthcare and how this takes shape in real-life. All three cases are implemented in the past five years and are a good representation of what is currently possible with AI/ML in the field of healthcare and medicine. Besides these three, there are many other examples and cases available. However, I have chosen these specific ones because they seem to cover the variety of applications in which AI could be used in healthcare.

The focus while looking at these cases will be what type of bias there could be with the specific application, but also what the consequences are/would be for the healthcare professional or the patient. The consequences will be reflected in what I would think is beneficial in the specific case and what the influence would be on the (present) bias.

4.1.AI for decision-making in acute medicine

As I established earlier within this thesis, AI can be used for different applications, of which one could be decision-making in acute medicine. Acute medicine can be described as the hospital's speciality that is concerned with the “diagnosis and treatment of adult patients with urgent medical needs” (SAM, 2023). It can be argued that these patients need a good and fast diagnosis that fits their medical condition best. Physicians listen to the conditions as described by the patient and connect this to a medical diagnosis and care. The physicians mostly also conduct additional tests and combine this with their own senses of what is noticed about the patient. For years, this has been done by the physicians but there are ways that AI could contribute to this process and make the time invested by the physicians alone shorter. For the computer programmer that could develop the AI, it might seem that the only thing needed to make the diagnosis is an indication of the assumed infection, an algorithm for detection with boundaries, and an overview of the treatment rules as known by the physicians (Lynn, 2019, p. 2). However, expert physicians will argue differently as they know that the protocols are not true to the exact complexity of acute medicine.

A benefit of using AI would be that it can analyse more relations between complex data within a patient's file. For example, when conditions would enhance/contradict each other or that gender/ethnic background could play a role in the final diagnosis. Which is related to the possible bias that could occur in ML when the AI is filtering for "complete" EHRs (Weber et al., 2017), I argue that this can be prevented if the filters would be to look at the conditions and the patients' gender/ethnic background. On the other side, a disadvantage would be that the decision-making process could become some sort of black-boxed process when it is done through AI. With the black boxing, a vulnerability towards to patient could arise as the steps to the decision cannot be seen and a decision made on the incorrect/incomplete patient data is harder to be recognized. A way to prevent this would be to design the system with transparency as a key point to see which steps the AI has taken to come to a decision. With transparency, it is important to consider the patients' privacy as the used data should only be accessible after the decision-making process to the healthcare professional and, if needed, the patient. Next to that, the communication of the AI cannot be less than the communication from a human. Mostly because of the human aspects that play an important role within healthcare, such as care and interaction with the patient.

I argue that the usage of AI in addition to the work of the healthcare professional could be beneficial as it saves the time for both patient and healthcare professional to come to the diagnosis and needed care. The AI can go over the patient's conditions and known research to narrow down faster which different diagnoses could be connected to the patient's medical condition. The healthcare professional should still have the final say and oversight, otherwise, it could be that the AI approach will not be beneficial.

With this, I argue that the AI decision-making process should be transparent to at least the healthcare professional as they can trace back on what grounds certain diagnoses or recommendations have been made. The healthcare professional will know where the certainty could be in the base of the decision-making process. Making the process transparent for patients would require more work as they do not, in general, have the medical background that professionals do. It is important to first establish the base and details of transparency for the AI decision-making before the decision can be made to realize the AI decision-making in acute medicine. If it is decided that the healthcare professional is the only person to know the steps involving the patient's data throughout the decision-making process, we could guarantee patient privacy.

It is difficult to immediately say that the medical decision that is made by either healthcare professional or AI is the wrong one until days later if complications arise, there is recovery failure, or the patient’s health declines considerably (Lynn, 2019, p. 3). The chance that the AI can quickly respond to the complications that arise will always be less fast as when the professionals see what happens with the patient and that the healthcare professional right away can act on what is happening. Which is different from when the information is conducted from research or a dataset, where the AI will act faster than the health professional. For the AI, I argue that it is needed to get more information on the current state of the patient and on the history between diagnosis and the complications. There is a certain danger of the potential delay that can be caused by the AI that will become clear when a patient getting treatment by the AI fails to recover compared to a patient treated by a healthcare professional alone (Lynn, 2019, p. 4). As mentioned before, a way to prevent this is by making the AI decision-making process transparent and making sure that the AI never operates without the supervision/oversight of a physician. The usage of AI in the diagnosis decision-making could contribute to eliminating the bias that comes from a healthcare professional while treating a patient and getting to a diagnosis. AI tool usage would influence the education of health professionals in the way that they are extra trained in the fields of empathy, comfort, counselling, and end-of-life care (Wynsberghe, 2022)**.** These are the fields that cannot be integrated on a personal level into the AI system.

4.2. AI as co-pilot in healthcare

Another noticeable example of using AI in healthcare can be found in the Amsterdam University Medical Center (UMC), where they research the usage of AI as a co-pilot in the medical teams within their hospital. During a symposium in 2023, they made clear that “the AI stays in hands of the patients and the medical professionals” (Asselbergs, 2023). This shows that they know it is not possible for AI to work without human supervision and supports the argument made in the acute medicine case. The study they are conducting is in collaboration with the Mayo Clinic Platform, the Mayo Clinic Platform focuses on improving availability of care to everyone and connecting new technologies that create opportunities and approaches to change the way care is provided. However, this works when the new technologies can show to improve care for patients with different genders and ethnic backgrounds. Which can be hard as the used patient data is anonymous and therefore might not show which genders and ethnicities are included. This platform works on combining anonymous patient records from all around the world to develop AI models that are more accurate. Mayo clinic platform claims that the patient records are anonymous, but there is no way to check this from the outside which leaves room to doubt whether all the patient data is indeed anonymous. Within the scope of this thesis, I will not conduct further research on whether the data is a hundred percent anonymous and leave a note that this should be checked within the span of further research on the topic. It would contribute to the privacy of the patients if the data is indeed anonymous. The fact that they use not only data from one hospital or country contributes to making the AI model more accurate and can be used with more confidence on different patients. Another benefit of this approach is the chance of bias in ML being reduced, as the data will consist of more complete EHRs.

Halamka, from the Mayo Clinic Forum, talks about the results that are already visible from the usage of AI in analysing the results of colonoscopy where, in their research, a doctor that makes use of AI comes with a better result than a doctor that does not use AI. I argue that the results can be different in other research and therefore underline that this is the result from the Mayo Clinic forum that is connected to this specific project. However, these results do not prove that AI on its own should take over the analysis of the results. I argue that AI in this sense is also more of an addition to the healthcare professional to get better, or worse, results collaboratively. For this to work, Halamka adds that “the AI should be transparent, reliable, and consistent” (Asselbergs, 2023) and this is also what I argue, before the consideration can be made that the AI could be of good value within healthcare. The chance of AI making things worse can be found in wrong usage of data or when there is a lack in transparency in the AI to see the reasoning.

Not only the healthcare professional could benefit from the usage of AI within healthcare, but also the patient. Especially when it comes to digital support about the condition they are in and monitoring the symptoms they experience. I argue that this could even lead to the patient getting more familiar with their condition and making it easier for them to know when they should contact their healthcare professional because of complications or other health-related matters. It can be beneficial for both patient and healthcare professional (Asselbergs, 2023). A downside is when the patient believes they are more familiar with their condition and they for example misinterpret a signal that shows their condition getting worse with a signal that shows improvement of treatment.

Within this case, I would not necessarily argue that the bias from the healthcare professional is changed drastically as the healthcare professional does not necessarily need to do something with the results that come out of the AI system. The AI is an addition that the healthcare professional could use when they see fit. In case the healthcare professional does choose to act on the results from the AI, I argue that there can be a decrease of bias in the process, and that the impact of bias on the decision-making becomes smaller.

4.3. CDSS for assisting with diagnosis.

Another example is the case about Clinical Decision Support Systems (CDSS) that can assist with diagnosis. CDSS is based on machine learning and can assist with diagnostic decisions (Lysaght et al., 2019). Furthermore, treatment outcomes can be forecast with CDSS to a certain point, based on the known data in the machine learning dataset. The way that Lysaght et al. (2019) describe CDSS, is the way that the system monitors information that is entered into the EHR continuously and that this information is analysed in combination with relevant data that is connected to the EHR. As the system has continuous access to new information, the system is continuously updated with the newly added information and therefore able to make more relevant diagnoses that are close to the diagnosis made by the physician.

The CDSS works based on an algorithm that is programmed based on clinical guidelines and (published) medical research. However, it will not include the human factor of seeing how a patient responds or what else there might be happening that could influence the patient's medical condition. Therefore, it is important to notice that the system needs to be transparent to the healthcare professional and that the accountability for the decision made by the system should be found not with the CDSS but rather with the healthcare professional or the programmer that has built the algorithm with perhaps little to no medical knowledge. Research from Jiang et al. suggests that machine learning can be almost as accurate as a healthcare professional with the addition that the CDSS can come to the diagnosis much faster than the healthcare professional on its own (Jiang et al., 2017). CDSS could possibly reduce the professionals bias when they compare a patient to similar earlier cases that have been experienced/treated by the healthcare professional.

The CDSS can be a good addition to making the work of the healthcare professional less time-consuming by taking over the comparison of a patient's medical condition to the information that is known to get to a diagnosis that fits the unique case of the individual patient. This is because the CDSS can get access to the information that the healthcare professional might be unaware of or information that is newly published after a fair amount of review on the research. Still, it is important that the system is transparent, and that the information included in the CDSS is limited to the amount of bias that is included from the data. The systems transparency should therefore extend to the data that is used and the outcomes of the system that can be used within the decision-making process of the healthcare professional. However, the transparency should not go so far that the patient' data is visible to everyone as this would comprise the patients' privacy.

5. The impacts of bias

The impacts of bias on decision-making have been researched from the view of human decision-making and AI decision-making. Within this thesis, I have argued that there are different impacts to be considered within both categories. Therefore, my answer to the question will be that the impact of bias is context dependent and can have an influence on the decision-making in healthcare. I will further elaborate on this in the following chapter.

5.1.Impact of bias in human decision-making

The impact of bias on human decision-making can be divided in different areas, as the research shows that there are different impacts. There are various ways in which bias can influence perceptions, judgements, and actions of people. For example, on how we interact with people that have a different (ethnic) background from ours. The way we interact can be formed through past experiences and change the way we speak and/or act to people. I argued before that bias does not by default need to be negative and the impact on the human decision-making can therefore still be positive, negative, or even neutral. The bias is somewhat a “filter through which a person evaluates the world, though they are unaware" (Rauf et al., 2022, p. 89), which shows that the impact of the bias is not always noticeable.

There are ways to possibly reduce negative bias in human decision-making, such as awareness training, promoting diversity, and inclusive policy (Holroyd, 2015). The awareness training can help to make people more aware of the bias that they (naturally) have and to make the impact smaller by being conscious about potential bias. Promoting diversity by including people with different backgrounds within either healthcare or the programmers that create AI applications as addition to the healthcare professional’s work. The inclusion of people from different backgrounds would mostly be beneficial within the healthcare professionals that are included, as the professionals can share their experiences and learn from each other how they address different situations. Healthcare is already relative diverse, which is why highlighting this is important and should be consider in all stages of the innovations in healthcare. The professional can still make choices on their own, but there is the possibility that the experience, and the connected bias, contributes to a decision they might make. However, we need to remember that not everyone makes decisions based on the same knowledge/experience and it is important to consider their values and beliefs. A person can give different worth to a behavior or object and their assumptions about health can have a distinctive image. The inclusive policy can be achieved by creating an overall policy that has allowed different people to express their opinion on it and include their feedback to make the policy more inclusive.

I argue that acknowledging and reducing bias in human decision-making is important for encouraging fairness and reasonable outcomes. The strategies as described earlier can help address the impact of bias on human decision-making and reduce the effect of the bias.

5.2. Impact of bias on AI decision-making

Besides the different impacts that can be noted about human decision-making, there are several impacts to discuss on AI decision-making. However, the approach toward bias on AI decision-making requires a different approach including consideration of (training) data quality, algorithmic transparency, and continuous monitoring. Making sure that the data quality is good, and the dataset includes data from people with different genders/ethnic backgrounds can result in partially limiting the bias within the AI. This can also help reduce the bias that the healthcare professional has based on research that only includes a small dataset that focuses on complete EHRs. The same complete EHRs that research from Weber et al. (2017) earlier showed that can form a negative bias towards older patients, and mostly female. Mainly as their research shows that these are the EHRs that are mostly not complete and therefore not included in the dataset. When we have algorithmic transparency, the chance of algorithmic bias itself becomes smaller, and we can start to focus on the bias that might be coming from the training data.

Which is why it is important to notice that there are multiple impacts of bias on AI decision-making that can be described. I argued two ways to reduce the bias in AI through considering the data quality and by creating algorithmic transparency, but there are more possible remedies to be considered. Such as risk stratification, possible bias in ML, and patient outcome optimization. As mentioned before, these could be viewed as methods to reduce bias by increasing awareness of potential biases. Risk stratification, potential bias in ML, and patient outcome optimization are related in AI decision-making. In contrast, they do not play the same role in human decision-making, which underlines their relevance in the context of AI. Risk stratification mostly leads to a general outcome on the underlying data that is used to develop the ML, and this could lead to bias on the form of the treatment. The impact on the patients' care is that the treatment might not be completely matching the need of the individual as the patient might differ from the general picture the ML has of a patient with similar health conditions. The research field of healthcare ethics focuses on this issue and other healthcare related decisions. Generalisation can promote efficiency and consistency in healthcare, but it also raises ethical considerations. Patients are all unique in the combination of their health conditions and environmental influences, which can influence the patients' response to treatment. A general approach within healthcare is therefore not ethical and allows for more research to tackle this broader issue in healthcare. For the scope of this thesis, I argue that the individual should be considered before the consideration to use the general approach.

The possible bias in ML can partly be reduced by the considering the quality of data. Another way to possibly reduce the bias is by letting the system learn how a disease could be treated by considering a persons' ethnic background and/or gender. In other words, the impact of bias in ML is visible when general data is used instead of specific patient data based on their gender and/or ethnic background. It could be that in most cases the general outcome can lead to unintended bias towards the patient and go by unnoticed as it is the right outcome in most of the cases.

Other ways to reduce the impact of bias and contributing to more just outcomes could be done through ethical AI development, AI as support, diversity in (software) teams, and extensive testing procedures. Ethical AI development can be achieved by including different starting points of data and including rules in the system to align with certain values as decided upon by, for example, the programmers. Which leads us to the diversity in (software) teams, to make sure that there is no bias included in the team that is going to develop the AI, or ML. A way to achieve this is by including people that have different ethnic backgrounds and genders, and by allowing everyone to voice their opinion. This variety of opinions allows for discussion to find common grounds to include in the system.

Furthermore, it is important that the AI should act as a tool to provide the healthcare professional with information rather than to make decisions on its own, as illustrated in the case from the Amsterdam UMC. This way potential biases in the AI decision-making can still be filtered out by human decision-making, which does leave room for potential bias of the person. This highlights the importance to include extensive testing procedures before the AI is used in practice. The goal of these tests is to exclude more potential biases from the program and making the AI as neutral as possible. I argue that it will be hard, perhaps impossible, to make the system hundred percent bias free and the AI neutral, but who knows how close the system could get in the (near) future.

6. Discussion & Conclusion

Bias within healthcare has many different forms and within this thesis I included a part of these different forms. I do admit that including every possible way of bias within the span of this thesis is nearly impossible, which is why I limited myself to the ones that I consider to be most familiar to the greater audience. The bias as explained within this thesis is related to the notion of decision-making, both from humans and through AI in the field of healthcare. Research on bias shows a relation between decision-making processes with noteworthy impacts differently in human decision-making and AI decision-making. Acknowledging, identifying, and understanding these impacts is important in developing ways to reduce bias and introduce fair outcomes. Within healthcare especially, it is important to understand and categorize impacts regarding human decision-making and AI decision-making to create customized interventions. One of them is the use of AI, or specifically ML, in healthcare that I argue should be used as an additional tool. A tool that is transparent and helps the patient or healthcare professional to get all the information to make the best decision instead of simply saying what should be decided based on known research. Research that might only cover a small part of the population or does not include data from people with different backgrounds. When considering the transparency of the tool, it is crucial to prioritize patient privacy. This is mainly achieved by ensuring that the patient's data is only accessible to the AI system and the healthcare professional. The AI accesses the data for decision-making processes, while the healthcare professional accesses the data to review the AI's outcome accuracy and provide better advice to the patient.

This project started with the research question: “What can be impacts of bias in AI and human decision-making within healthcare?”, which I can now (partially) answer by arguing that the impact of bias in AI and human decision-making within healthcare can be very diverse in nature. Most of the time the impact of the bias is context dependent and can therefore not be put in one box. By making the distinction between human decision-making and AI decision-making, it allows for a focused approach that can help reduce bias and increase fairness within both contexts. During the research it became clear that the two contexts are related on different aspects within the field of healthcare, which is another reason why the research started by defining bias in general before looking into decision-making. By looking into both AI and human decision-making, I argue that the way to reduce the impact of bias would be a combination of different approaches. When we will start to include AI-generated information within the decision-making process of the person, we are able to reduce the persons' own bias. However, this does mean that the bias included within the AI can start to play a bigger role and could lead to a different result than expected. Which is why it is important to start with a certainty before creating the AI system by considering the data quality and allowing for algorithmic transparency. Furthermore, transparency of the process that the AI follows is important to consider, as this contributes to retracing a part of the steps that have led to the decision. Especially when there are uncertainties from the healthcare professional or the patient. In addition to this, I considered values to consider the ethics that is connected to the usage of AI tools in decision-making. These values are accountability, responsibility, and vulnerability and they could each individually be connected to different impacts of bias. The impact of accountability on bias is visible when we identify the bias that is present and to ensure transparency and fairness in the outcome. When the developer or healthcare professional is accountable, we also know who is responsible in case something goes wrong and use this to improve further developments. Which is connected to the impact of responsibility on bias and why I have used the values combined to research the impacts of bias in AI and human decision-making. Additionally, I explored the relationship with vulnerability that comes to play when decisions about care are based on incomplete or incorrect patient data. I will elaborate on vulnerability in the next paragraph.

The addition of the case studies provided a more tangible dimension to the research, as they showcase real life scenarios to the theory of the topic. Having AI as an additional tool in the decision-making process of the healthcare professional can be seen as a positive way to reduce the human bias, while still arguing that the AI should consist of a broad dataset that could be considered of good quality; with good quality being defined as data that covers patients with different ethnic backgrounds, ages, and genders to avoid the AI having a negative bias towards certain patients or patient groups. This contributes to less vulnerability when it comes to patients with different backgrounds that perhaps are less heard when smaller datasets are used. To review the case studies, I used privacy as a lens to determine who needs access to patient data and how to best protect the patient's privacy. The main conclusion I would make about privacy is that data access should be restricted to only those individuals or systems necessary for optimal patient care.

Further research on this topic could focus on new developments within the field of decision-making tools and AI systems, and additionally could consider the application of these tools and systems in other field, beyond medicine. There are things to consider when using decision-making tools in different fields that differ from the use in medicine, such as the data that is used and the importance of difference in ethnic background/gender. Another possibility for further research can be found in the fact that there are other considerations about the way that bias plays a role or the significance that there is a potential bias. It could be argued that there are fields where the presence of bias does not play as large a role in the process, for example when someone was in violation and there is a law that leads to the consequence of the violation. Indeed, there is a certain discussion that could be started when we talk about bias in the field of law, as it mostly will go further than simply looking at the law and including the person in question. Which could lead to potential bias from the judge or law enforcement officer based on the person’s ethnic background and/or gender, such as in the field of healthcare.

Another possibility for further research could be to focus on the origin of bias, which can be done by focussing on the historical aspects rather than on the innovative parts such as AI tools. I have included a small part about the creating a foundation when I talked about certainty, but there is still far more to be researched than I have touched upon. Research on certainty within bias could be an interesting study as one could question whether there can be bias when there is certainty or if there is certainty when bias is present. I expect that this provides another fascinating avenue for further research.

Literature

Andersen, F., Anjum, R. L., & Rocca, E. (2019). Philosophical bias is the one bias that science cannot avoid. *Elife, 8*, e44929.

Amodio, D. M., & Mendoza, S. A. (2010). *Implicit intergroup bias: Cognitive, affective, and motivational underpinnings.* In B. Gawronski & B. K. Payne (Eds.), Handbook of implicit social cognition (pp. 353–374). New York: Guilford.

Asselbergs, F. (2023, November 3). *Artificial Intelligence als copiloot in de zorg.* Amsterdam UMC. https://www.amsterdamumc.org/nl/vandaag/artificial-intelligence-als-copiloot-in-de-zorg.htm

Balakrishnan, K., Boss, E. F., & Chang, D. (2018, May 4). *Cognitive and implicit bias as barriers to optimal patient management.* AAO-HNS Bulletin. https://bulletin.entnet.org/home/article/21247268/cognitive-and-implicit-bias-as-barriers-to-optimal-patient-management#:~:text=Cognitive%20biases%20generally%20apply%20to,%2C%20gender%2C%20and%20socioeconomic%20status.

Berthet, V. (2022). The impact of cognitive biases on professionals' decision-making: A review of four occupational areas. *Frontiers in psychology,* 12, 802439.

Brownstein, M., Madva, A., & Gawronski, B. (2020). Understanding implicit bias: Putting the criticism into perspective. *Pacific Philosophical Quarterly,* 101(2), 276-307.

Cameron, C. D., Brown-Iannuzzi, J. L., & Payne, B. K. (2012). Sequential priming measures of implicit social cognition: A meta-analysis of associations with behavior and explicit attitudes. Personality and Social Psychology Review, 16(4), 330-350.

Campesi, I., Montella, A., Seghieri, G., & Franconi, F. (2021). The person's care requires a sex and gender approach. *Journal of Clinical Medicine,* 10(20), 4770.

Croskerry, P. (2003). The importance of cognitive errors in diagnosis and strategies to minimize them. *Academic medicine, 78*(8), 775-780.

Croskerry, P., Singhal, G., & Mamede, S. (2013). Cognitive debiasing 1: origins of bias and theory of debiasing. *BMJ quality & safety.*

Dera, J. D. (2019). Risk stratification: a two-step process for identifying your sickest patients. *Family practice management,* 26(3), 21-26.

Descartes, R. (1641). *Discourse on Method and Meditations on first philosophy.*

Elbers, P., & Thoral, P. (2022, August 25). *Wereldprimeur: Artificial intelligence ondersteunt artsen bij beslissing ontslag IC-patiënt.* https://www.amsterdamumc.org/nl/vandaag/wereldprimeur-artificial-intelligence-ondersteunt-artsen-bij-beslissing-ontslag-ic-patient.htm

Erden, Y. J. (2021). Is the brain a digital computer? Rethinking a binary question. *Think, 20*(57), 23-37.

Gillespie, T. (2014). The relevance of algorithms. *Media technologies: Essays on communication, materiality, and society,* 167(2014), 167.

Giordano, C., Brennan, M., Mohamed, B., Rashidi, P., Modave, F., & Tighe, P. (2021). Accessing artificial intelligence for clinical decision-making. *Frontiers in digital health,* 3, 645232.

Giuffrida, I. (2019). Liability for AI decision-making: some legal and ethical considerations. *Fordham L. Rev.,* 88, 439.

Graber, M. L., Kissam, S., Payne, V. L., Meyer, A. N., Sorensen, A., Lenfestey, N., ... & Singh, H. (2012). Cognitive interventions to reduce diagnostic error: a narrative review. *BMJ quality & safety, 21(7)*, 535-557.

Hilbert, M. (2012). Toward a synthesis of cognitive biases: how noisy information processing can bias human decision making. *Psychological bulletin,* 138(2), 211.

Holroyd, J. (2015). Implicit bias, awareness and imperfect cognitions. Consciousness and cognition, 33, 511-523.

Holroyd, J., Scaife, R., & Stafford, T. (2017). What is implicit bias? Philosophy Compass, 12(10). https://doi.org/10.1111/phc3.12437

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., ... & Wang, Y. (2017). Artificial intelligence in healthcare: past, present and future. *Stroke and vascular neurology,* 2(4).

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I* (pp. 99-127).

Klugt, G. van der. (2022, August 25). *AI helps Dutch doctors discharge intensive care patients*. Techzine Europe. https://www.techzine.eu/news/analytics/86965/ai-helps-dutch-doctors-discharge-intensive-care-patients/?utm_source=dlvr.it

Kudless, A. (2023). Hierarchies of bias in artificial intelligence architecture: Collective, computational, and cognitive. *International Journal of Architectural Computing,* 14780771231170272.

Kurdi, B., Seitchik, A. E., Axt, J. R., Carroll, T. J., Karapetyan, A., Kaushik, N., ... & Banaji, M. R. (2019). Relationship between the Implicit Association Test and intergroup behavior: A meta-analysis. American psychologist, 74(5), 569.

Lynn, L. A. (2019). Artificial intelligence systems for complex decision-making in acute care medicine: a review. *Patient safety in Surgery,* 13(1), 6

Lysaght, T., Lim, H. Y., Xafis, V., & Ngiam, K. Y. (2019). AI-assisted decision-making in healthcare: the application of an ethics framework for big data in health and research. *Asian Bioethics Review,* 11, 299-314.

Mayo Clinic Platform. (2023, March 17). About us - Mayo Clinic Platform. https://www.mayo clinicplatform.org/about/

Meskó, B. (2017), *the guide to the future of medicine (Technology AND the human touch)*, ISBN 978-963-08-9802-7

Meskó, B. (2019). The real era of the art of medicine begins with artificial intelligence. *Journal of medical Internet research,* 21(11), e16295.

Morewedge, C. K., Yoon, H., Scopelliti, I., Symborski, C., Korris, J., and Kassam, K. S. (2015). Debiasing decisions: improved decision making with a single training intervention. *Policy Insights Behav. Brain Sci.* 2, 129–140. doi: 10.1177/2372732215600886

Noble, S.U. (2018), *Algorithms of Oppression: how search engines reinforce racism*, ISBN: 9781479837243

Petrescu, P. (2021, March 31). *Google Organic Click-Through Rates in 2014.* Moz. https://moz.com/blog/google-organic-click-through-rates-in-2014#:~:text=On%20average%2C%2071.33%25%20of%20searches,10%20account%20for%20only%203.73%25

Polack, P. (2020). Beyond algorithmic reformism: Forward engineering the designs of algorithmic systems. *Big Data & Society*, *7*(1), 2053951720913064.

Rauf, I., Hartmann, A., Koumtchev, A., Khan, S. A., & Kashyap, R. (2022). Conscious and Unconscious Bias: The Hidden Pandemic of Biases in Healthcare Exacerbated by COVID-19. *HCA Healthcare Journal of Medicine,* 3(3), 89.

Saul, J. (2013). *Implicit bias, stereotype threat, and women in philosophy. Women in philosophy: What needs to change*, 39-60.

Schnapp, B. H., Sun, J. E., Kim, J. L., Strayer, R. J., & Shah, K. H. (2018). Cognitive error in an academic emergency department. *Diagnosis,* 5(3), 135-142.

Sellier, A. L., Scopelliti, I., and Morewedge, C. K. (2019). Debiasing training improves decision making in the field. *Psychol. Sci.* 30, 1371–1379. doi:10.1177/0956797619861429

SAM. (2023, May 1). Training in Acute Medicine - Society for Acute Medicine. Acute Medicine. https://www.acutemedicine.org.uk/training-in-acute-medicine /#:~:text=Acute%20Medicine%20(or%20Acute%20Internal,patients%20with %20urgent%20medical%20needs.

Szaniawski, K. (1980). Philosophy of decision making. *Acta Psychologica,* 45(1-3), 327-341.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. science, 185(4157), 1124-1131.

van Wynsberghe, A. (2022). Social robots and the risks to reciprocity. AI & SOCIETY, 37(2), 479-485.

Weber, G. M., Adams, W. G., Bernstam, E. V., Bickel, J. P., Fox, K. P., Marsolo, K., ... & Mandl, K. D. (2017). Biases introduced by filtering electronic health records for patients with "complete data". *Journal of the American Medical Informatics Association*, 24(6), 1134-1141.