Stellenbosch
UNIVERSITY
IYUNIVESITHI
UNIVERSITEIT

The Ethics of Artificial Intelligence in Healthcare Settings

by
Dr Adetayo Emmanuel Obasa

Thesis presented in fulfilment of the requirements for the degree of
Master of Philosophy in the Faculty of
Arts and Social Sciences at Stellenbosch University

Supervisor: Dr Andrea Palk

December 2023

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third-party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

December 2023

Acknowledgements

I am incredibly grateful to Dr Andrea Palk, my supervisor, for her invaluable guidance and advice throughout this project. Without her expertise, this day would not have been possible. I also want to thank Prof. Anton van Niekerk for his unwavering support during difficult times.

I want to thank my wife, Zimvo Obasa and our newborn daughter Oluwapelumi for their unwavering support, motivation and encouragement. The latter kept me awake to continue studying when it was required. I also appreciate the Redeemed Christian Church of God's Christian community's prayers and moral support.

I sincerely appreciate the Faculty of Medicine and Health Sciences, Stellenbosch University, for their financial support with my school fees. With the faculty's financial support, my project was a success.

Finally, I want to thank my parents for believing in me and supporting me throughout my years of studying. I am forever grateful for God's wisdom, protection and guidance!

"I have set the Lord always before me:
Because he is at my right hand,
I shall not be shaken."
Psalm 16:8

Dedication

"My help comes from the LORD, the maker of heaven and earth" (Ps 121: 02). I dedicate the thesis to Yeshua Hamashiach

Abstract

Artificial Intelligence (AI) has the potential to transform and revolutionise the healthcare industry. More specifically, it stands to improve patient outcomes, reduce costs, and increase productivity by providing customised and precise solutions. AI applications range from mental health to diagnosis, treatment, nursing, and hospital management. However, there are ethical concerns and obstacles that must be addressed, such as bias, data privacy, regulatory compliance, and various other ethical considerations. In this thesis, I critically apply the Principlist framework to the abovementioned issues, with the aim of incorporating AI into healthcare in a way that fosters dignity, solidarity, and trust in healthcare and technology. In a medical context, trust is critical because patients have no choice but to put themselves in the hands of healthcare practitioners who have the specialist knowledge they need. The relationship between patients and healthcare practitioners is thus one of dependency or asymmetry, whereby patients must assume that the healthcare practitioner has their best interests at heart.

After applying the Principlist framework, I then use the three influential ethical theories: consequentialism, deontology and virtue ethics, to consider the notion of morally competent AI in 'robot' form. This includes a critical consideration of human interactions with autonomous robots and some of the concerns elicited in this regard. Finally, I propose the ethics of responsibility, first introduced by Max Weber and subsequently developed by numerous influential thinkers, as a potential framework to address the ethical, legal, and social implications of AI in healthcare. I also explore the historical development of ethics of responsibility to gain valuable insight and apply these insights to AI in Healthcare. I conclude with some recommendations and insights that may be valuable for policymakers, practitioners, and the public in navigating the ethical challenges and ensuring the responsible and beneficial use of AI in healthcare settings.

Contents

Declaration *2*

Acknowledgements *3*

Dedication *4*

Abstract *5*

Chapter 1 .... *9*

1.1 Brief Introduction and Background 9

1.2 Aims and Objectives 9

1.3 Chapter layout 10

Chapter 2 .... *12*

2.1 Background, Definition, and terminology 12

2.2 Defining artificial intelligence 12

2.3 How can AI be used in healthcare? 13

2.4 Robots in healthcare 15

2.5 Areas of AI that pose ethical risks 16

2.5.1 Machine Learning (ML) 16

2.5.2 Deep Learning (DL) 19

2.6 Big data in healthcare 19

2.7 Conclusion 20

Chapter 3 .... *22*

3.1 Overview 22

3.2 A Principlist Framework approach for ethical AI application in healthcare 22

3.3 Respect for Autonomy in Healthcare 23

3.4 Non-maleficence 25

3.4.1 Harms and the responsibility gap in AI 28

3.5 Beneficence 29

3.6 Justice 31

3.7 Trust in AI Systems 33

3.8 Conclusion 35

Chapter 4 .... *36*

4.1 Overview 36

4.2 ‘Moral’ Robots 36

4.3 “Moral’ robots in healthcare 39

4.4 Human interaction with autonomous robots 41

4.5 Developing morally competent and ethical robots for healthcare contexts 42

4.6 Deontological considerations 42

4.7 Consequentialism 45

4.8 Virtue Ethics 49

4.9 Conclusion 52

Chapter 5 .... *53*

5.1 Overview 53

5.2 Historical Development of Ethics of Responsibility 53

5.3 Application of these perspectives on the ethics of responsibility 56

5.4 Phronesis and Ethics of Responsibility 58

5.5 The possibility of failure 61

5.6 An Ethics of Responsibility as an ethics of futurity.. 61

5.7 The future of AI in healthcare.. 62

5.8 Potential recommendations 62

5.9 Conclusion 64

5.10 Concluding remarks. 65

References.. *66*

List of figures

Figure 1: Shows the Flow chart of the types of machine learning algorithms……………………………….15

Figure 2: Examples of big data sources in healthcare……………………………………………………………….18

Chapter 1

1.1 Brief Introduction and Background

The field of Artificial Intelligence (AI) is currently undergoing rapid development and has increasing application throughout various sectors of society. In the healthcare industry, the abundance of data obtained from medical research and healthcare, coupled with the integration of technology, has paved the way for an exciting prospect - the effective utilization of these data through artificial intelligence (AI). Experts believe that AI has the potential to analyse and extract valuable insights from these data, leading to ground breaking discoveries in the field, with implications for improved diagnoses and treatments, while reducing costs and increasing productivity. (1) AI has already been utilized in multiple areas including medical image analysis, oncology scanning, disease diagnosis, forecasting the success of dental implant cases, and, more recently, during the COVID-19 pandemic to triage patients in the Intensive Care Unit (ICU) etc. (2,3)

The World Health Organisation (WHO) has, however, recommended caution in adopting innovative technologies, such as AI. (4) Successfully translating such technologies into clinical and research contexts requires rigorous evaluation and expert supervision, to avoid errors and harm to patients. Moreover, transparency and public engagement are crucial to foster trust in AI. In this regard, the WHO proposes that unambiguous evidence of benefit be measured, before widespread use in routine healthcare contexts, to ensure patient safety and protection, primarily. (5) This requires addressing concerns about safety, threat to autonomy, data privacy, regulatory compliance increased healthcare costs, inequity, data-source bias, transparency, and accountability regarding the implementation of artificial intelligence and other ethical considerations. (3 6,7) Failure to adequately address these concerns could result in AI exacerbating existing inequities in society. (9) It is crucial to examine these challenges to ensure that AI is used ethically and effectively in healthcare.

1.2 Aims and Objectives

I contend that the successful and ethical implementation of AI in healthcare clearly necessitates attention. In this thesis, therefore, my aim is to consider the various ethical concerns, regarding the translation of AI into healthcare contexts, by drawing on various ethical theories and perspectives. I also aim to address some of these concerns by developing a framework for the ethical implementation and responsible use of AI in healthcare.

Relatedly, my research objectives are:

1. to provide relevant definitions and outline the increasing relevance of AI in healthcare.
2. to conceptualise the ethical concerns related to use of AI in healthcare.
3. to examine and fully consider these ethical challenges to ensure that AI is used ethically and effectively in healthcare.
4. to address ethical concerns related to AI in healthcare using the Principlist framework.
5. to consider how three influential ethical theories - consequentialism, deontology, and virtue ethics – can inform the development of morally competent AI robots.
6. to suggest an ethics of responsibility as a way of complementing existing ethical approaches, and one which is particularly relevant for the application of AI in healthcare contexts.
7. to make future recommendations based on the proposed ethical framework.

1.3 Chapter layout

In the second chapter I begin by defining Artificial Intelligence and considering how AI and robots are currently used in healthcare and their future potential. I then introduce the AI components that pose ethical risks, namely, Deep Learning, Machine Learning, and the generation of Big Data, and Robots. I also briefly touch on the ethical implications of AI in healthcare which will be further expanded in Chapter Three.

In the third chapter, I delve into the ethical implications of AI in healthcare. I begin by introducing the concept of an ethical framework and providing a short justification for the suitability of Beauchamp and Childress' Principlist framework for ensuring ethical application of AI in healthcare. This is followed by a more substantial discussion of the Principlist framework and a deeper discussion of how it can be applied to address the ethical issues arising from AI in healthcare. Next, I explore how conceptions of morality and ethics relate to AI. I analyze the ethical issues involved in the use of AI in healthcare and the potential impact on patients and healthcare providers. I also examine ethical considerations in the design, development, and implementation of AI in healthcare. Furthermore, I discuss the importance of trust in AI in healthcare and how it relates to ethical considerations. The introduction of AI in healthcare creates a responsibility gap, which, without sufficient oversight and accountability, risks leaving healthcare providers and patients potentially vulnerable. I consider the ethical implications of the responsibility gap and propose ways to address them, including the need for transparency, accountability, and ethical oversight in the design and implementation of AI in healthcare.

In the fourth chapter I consider the possibility of effectively training AI systems to be sensitive to human moral norms, values, and ethics. The focus of this chapter is to highlight the necessity for AI systems to be programmed to 'make decisions' and 'behave' in a manner that reflects and is sensitive to human values and principles such as privacy, fairness, non-discrimination, and human dignity. AI systems that are not programmed or 'trained' without deep consideration of these principles and values may reinforce biases and other unintended consequences such as human rights and privacy violations. To assist this discussion, I consider how influential ethical theories, such as Utilitarianism, deontology or Kantian ethics, and virtue ethics, can guide the design and programming of AI systems in healthcare. Each approach has its limitations and challenges, and a combination of these frameworks may be necessary to address the complex ethical issues in AI development.

In the fifth chapter, I propose the ethics of responsibility as a framework to consider when developing a morally sensitive robot. I propose this framework as a dual construct, encompassing both forward and backward responsibilities. The main aim of this framework is to support developers and healthcare professionals in taking accountability for the outcomes of their decisions. This includes assuming proactive responsibilities to shape their behaviour and prevent future harms, as well as accepting retrospective responsibilities for any harm that has already occurred. Thereafter, I offer suggestions that could be beneficial to decision-makers, professionals, and the public when it comes to handling ethical dilemmas and promoting the ethical and advantageous application of AI in healthcare environments.

Chapter 2

2.1 Background, Definition, and terminology

The field of Artificial Intelligence (AI) has experienced radical change in recent years and is now in a position of considerable, and increasing, global relevance and interest. This field is, however, a vast one that defies easy definition, because its scope is continually changing due to rapid development in this area. In this chapter I start by briefly defining artificial intelligence before moving on to the focus of the chapter, and thesis, which is the application of AI in healthcare. After this overview, which introduces some of the ethical concerns that will be discussed in subsequent chapters, I then narrow the focus to explain the specific areas of artificial intelligence systems that elicit ethical concern, namely deep learning and machine learning as well as the risks associated with the ability to generate large datasets. I conclude by discussing another area where AI makes a significant contribution to healthcare, namely through medical robots.

2.2 Defining artificial intelligence

Artificial intelligence (AI) can be broadly defined as the simulation of intellectual processes usually associated with intelligent human cognition, such as learning, decision-making, troubleshooting, problem-solving, executing tasks and self-correction. (10–12) AI is a vast field that encompasses various subfields, including Machine Learning (ML), which involves the development of computer algorithms[1] that are designed with the capacity to learn and improve automatically through experience. (2,13) However, the terms AI and ML have fuzzy boundaries which are heavily debated in the literature. Kaplan and Haenlein define AI as “a system’s ability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation.” (14) Poole and Mackworth (2010) define the field of AI as focused on “the synthesis and analysis of computational agents that act intelligently.” (15) Here we can understand an agent as an entity capable of action, and an intelligent agent as an entity whose “actions are appropriate for its circumstances and its goal…is flexible to changing environments and changing goals…. [can] learn from experience, [and] makes appropriate choices given its perceptual and computational limitations.” (15) While the development of AI systems has numerous applications in various industries (2,13), what is of interest for the focus of this thesis, is its application in health care.

[1] An algorithm is a procedure used for solving a problem or performing a computation task. In AI, an algorithm enables a computer to learn from data and make decisions without explicit programming. AI algorithms in healthcare assist in radiographic image interpretation and skeletal age determination, improving diagnostic accuracy and efficiency.

2.3 How can AI be used in healthcare?

The introduction of AI into the healthcare system has the potential to transform how healthcare is delivered. (16) AI can help healthcare professionals make better decisions, improve patient outcomes, and increase healthcare delivery efficiency. More specifically, machine learning and natural (human) language processing[2] are two AI technologies that can be used to analyze large amounts of data and extract meaningful insights that can be used to improve clinical decision-making and patient care. (16) AI systems also hold significant potential in aiding early disease diagnose in healthcare settings. For example, AI-driven technologies are trained to analyse medical images to diagnose and identify specific diseases, including being able to differentiate between benign and malignant tumours. (17) In addition, AI-enabled microscopes can scan for harmful microorganisms in blood or fluid samples and monitor viral transmission patterns in real-time, quicker, and more efficiently than manual scanning. In Low- and Middle-Income Countries (LMICs), for example, AI has been used to assist in the detection of tuberculosis by scanning for symptoms and signs of tuberculosis, X-ray scanning, and interpreting staining images, which allows for early identification of the disease. (18,19)

During the COVID-19 pandemic, AI technologies were used to assist decision-making about prioritisation and allocating scarce resources. While, the Sequential Organ Failure Assessment (SOFA) had been used for decades before the pandemic, (20) to determine the severity of illness, analyse, and predict mortality, to help guide the allocation of Intensive Care Unit (ICU) beds, it is not an AI-powered system. An AI-powered version called "DeepSOFA" has, however, been developed, and used in some settings to help lessen the burden associated with the use of the manual SOFA's score. (21) During the peak of the pandemic, most countries lacked sufficient ventilators and bed capacity, which resulted in high mortality levels, with the worst affected countries being overwhelmed due to limited resources. The DeepSOFA technology powered with AI technologies has considerable potential to assist in identifying patients who will benefit most from bed and ventilator allocations, as well as those who no longer require ventilator support or further interventions. Beyond patient care, AI can also be used to optimize administrative tasks to ease the workload of healthcare professionals and prevent unnecessary hospital visits or remissions. However, the aim of AI is not intended to replace human physicians but to enable them to more accurately diagnose and better treat patients. (22)

[2] Natural Language Processing (NLP) is a field of computer science, specifically within artificial intelligence (AI), that focuses on empowering computers to comprehend and interpret text and spoken language like human beings.

The use of health-related AI technologies has also extended beyond the healthcare environment because these technologies are now easily, and in many cases, affordably, acquired and can be used without involvement from the healthcare sector. Wearable technologies on the body, such as smart watches, insulin pumps, electroencephalogram devices and activity trackers, enable patients without clinical training to self-manage their health. (23) These wearables create opportunities to efficiently monitor and capture data to predict health risks in real-time. It has been estimated that by the end of 2025, 1.5 billion wearable units will have been purchased annually. (24) While healthy individuals using these wearables can generate information to detect health risks and improve wellbeing, for those with health conditions, these devices could be used to suggest medications/prescriptions and improve treatment when necessary. (17,25)

In medicine, AI was first introduced in the 1970s when medical expert systems— based on Bayesian statistics and decision theory — were used to diagnose and recommend glaucoma and infectious disease treatments. (26,27) Considerable progress made in artificial neural networks, hybrid intelligent systems, and Bayesian networks in the late 1990s has scaled up bioinformatics research. As a result, the uptake of Medical Artificial Intelligence (MAI) has expanded. (26,27) It has been projected that global investment in MAI will reach $6.6 billion by 2021; therefore, it is anticipated that AI implementation in healthcare can help save $150 billion in costs by 2026. (28) Medicine thrives on innovation – and as the examples mentioned above have shown, the primary aim of rolling out AI in the health sector is to perform autonomous procedures, relieve clinicians from mundane roles, assist with precision medicine and provide more efficient and accurate results.

While the incorporation of AI in healthcare is a much-needed innovative approach to better service delivery within the healthcare industry, medical practice is already fraught with ethical dilemmas and challenges. Insofar as every decision in healthcare has ethical implications, such decisions must be morally and ethically justified, or at least, justifiable. From a technical perspective, AI would have to either suggest or make predictions about the best possible treatment outcomes. This has ethical implications insofar as there is a potential for misdiagnosis by way of flaws in the programming of such systems, such as algorithmic bias, or errors that might occur as part of the learning process. Where there is resultant harm or loss of life, the use of AI systems may generate gaps in accountability or responsibility. There are also concerns about privacy and the security of the large scale generation of data sets, with the use of AI in healthcare contexts, which may threaten patients, physicians, manufacturers, regulatory bodies and other stakeholders. Addressing these gaps and concerns would improve both trust in, and trustworthiness of, AI, whereas failure to address them could have detrimental effects on the use and success of AI in the healthcare sector.

In this regard, while the introduction of AI in the healthcare sector is primarily aimed at improving service delivery within the industry, (29) the impact it will have on the healthcare sector as a whole, and on patient well-being, will depend on how it is developed, applied, and regulated. This includes identifying, understanding, and fully addressing the ethical concerns raised by the application of AI. These concerns include: (1) obtaining consent to store and use data (2), ensuring adequate attention is paid to safety and the need for transparency, (3) algorithmic fairness and awareness of algorithmic biases[3] (4) data security and privacy (5) dignity and solidarity and (6) trust. (30) All these challenges pose a significant threat to the success of AI in healthcare if adequate attention to them is not forthcoming. Moreover, failure to address these concerns could have the result of increased inequities and inequalities in society. All these ethical challenges will be discussed fully in chapter two.

2.4 Robots in healthcare

In the 21st century, robots and robotics in healthcare are poised to revolutionise medical practice. A robot can be defined as an autonomous or semi-autonomous machine that has the capacity to act independently of external commands. Robots[4] can make use of artificial intelligence to improve their autonomous functions by way of machine learning, but not all robots have artificial intelligence, as defined above. While machine learning and miniaturisation have contributed to improvements in the design and increased use of robots in healthcare, medical robots are not new to the field. In fact, robots were first introduced 34 years ago when biopsy specimens were obtained by a commercial robot and computed tomography navigation when inserting a probe into the brain. (31,32) Subsequently, robots have been used to perform urological procedures and total hip arthroplasty. (33,34) Healthcare practitioners, and surgeons in particular, remain sceptical about the possibility of fully autonomous robots. (33,34) As a result, developers had to design robots to function as tools for their surgeon operators, rather than being fully autonomous. (31,32)

[3] Algorithmic biases refer to the lack of fairness in the outputs generated by an algorithm. These biases may include age discrimination, gender bias, and racial bias.

[4] Robots are used in healthcare in various forms, such as humanoid robots, animal-like robots, and specialized robotic systems. These robots are equipped with sensors, actuators, and artificial intelligence algorithms to interact with patients, assist healthcare professionals, and perform specific healthcare-related tasks.

Robots are well known for their roles in precision medicine, clinical diagnoses, surgeries, and computer software, in which they are used to accurately analyse radiology images and manipulate surgical instruments through one or more small incisions for multiple procedures. (35) In 2000, the Food and Drug Administration (FDA) in the United States approved a device called da Vinci robotic surgery. The da Vinci robotic device, which has four 'arms', has been used to perform more than 6 million surgeries worldwide, including in Cape Town, South Africa. (35) Both surgeons and patients benefit from robotic-assisted surgery, which is more time-efficient and cost-effective, in comparison to traditional surgery, and which have the added advantage of being able to make smaller incisions, which reduce blood loss. (35)

Robotic endoscopic capsules can be ingested into a patient's digestive system to gather and send diagnostic information back to the operator. Micro-robotics can also be used to deliver therapy such as radiation or medication to a specific site in the body via the blood vessels. Prostheses and robotic limbs are also benefitting from new structures and control systems. In rehabilitation centres, robotic exoskeletons (orthoses) are used to assist paralysed patients in walking and to correct malformations. (36,37) Due to the global shortage of healthcare workers, there are also robotic assistants designed to assist or replace overworked nursing staff with mundane tasks such as digital entries, drawing blood, patient monitoring and moving carts. Beyond healthcare diagnoses, robots are also used to keep the hospital clean using high-intensity ultraviolet (UV) light applied by a robot. (38) Robots in healthcare are becoming more prevalent due to their ability to perform tasks with greater efficiency and accuracy. Amidst the pandemic, the medical field has witnessed the reliability and transformative potential of robotic and AI-powered. These innovations have emerged as significant advancements, poised to revolutionize the discipline and open new frontiers in healthcare.

In the remainder of this chapter, I provide a brief explanation of areas that pose ethical risks such as: Deep Learning (DL), the various categories of Machine Learning (ML), and Big Data in Health care. It is important to understand these concepts as they form the backbone of innovative technologies and solutions in healthcare.

2.5 Areas of AI that pose ethical risks

2.5.1 Machine Learning (ML)

As mentioned above, machine learning is the process whereby systems can independently learn, thereby improving their performance on a task. Rather than needing to be explicitly programmed, large amounts of data are fed into a computer program, which uses statistical algorithms to find patterns and make predictions or decisions based on the input data. This data-driven learning process has a wide range of applications, from detecting spam emails to recognizing speech and images. This is in keeping with the goal of machine learning which is to teach computers to perform tasks that would otherwise require human intelligence. (39) Machine learning has many practical applications in fields such as healthcare, as mentioned in the previous section, machine learning algorithms can be used to predict patient outcomes and diagnose diseases in healthcare.

A concern that has received extensive attention, is the risk of bias in machine learning, which can be introduced during the acquisition of datasets. Bias in medical data can arise in two ways: first are cases in which the data source does not reflect true epidemiology within a given demographic population. An example of this is the entrenched overdiagnosis of schizophrenia in the African American population. (40) This overdiagnosis can be attributed to racial biases, whether conscious or subconscious, in the African American population. (41) Second are cases whereby the algorithms are trained on a data set that does not represent enough members of a certain population – for instance, an algorithm trained mostly on data from older white males. (40) Such an algorithm would have poor predictions, for example, among older black females. (40) If these algorithms trained on biased data sets are introduced in healthcare settings, they can exacerbate disparities within the health care sector. (42,43) Thus, for AI to function efficiently, there is a need for robust data sets to enhance its effectiveness within the healthcare environment. Contemporary ML is a rapidly evolving field, and there are three sub-categories of learning: supervised learning, unsupervised learning and reinforcement learning.

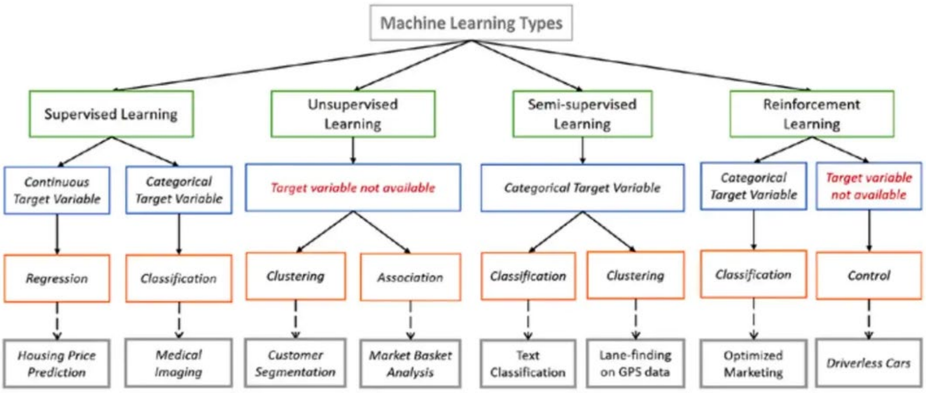

Figure 1: Shows the Flow chart of the types of machine learning algorithms. (Source-https://towardsdatascience.com/types-of-machinelearning-algorithms-you-should-know-953a08248861)

Supervised learning is method centric such as regression and classification. In this type of learning, we have labelled data, which are used to solve classification problems. The process of labelling is important as it helps humans to access each piece of data and label the data (**Figure 1**). In this form of learning, algorithms aim to understand the connections and patterns between input features and the desired output, enabling them to predict output values for new data based on the knowledge gained from past datasets. (44) The outcome of this ML process is called a classifier. A classifier is a software that can interpret the label and automatically predict the label of a new piece of data. (44) **Unsupervised learning** is quite different because the computer is trained with unlabelled data. (**Figure 1**) The methods involved are principal components and analysis, and clustering. Unsupervised learning serves as an exploratory precursor to the supervised learning method. (44) **Semi supervised learning** occupies a middle ground between the two types of machine learning. In this approach, some observations in the dataset are provided with labels, while others lack them. (**Figure 1**) In this type of learning, labelling requires experts and skilled individuals, however, the cost of labelling can be high. In addition to this excessive cost, it also requires experienced human professionals to carry out the task. Semi-supervised algorithms are the most suitable option for constructing a model in situations with insufficient labelled data. (44) **Reinforcement learning** does not focus on labelled data but provides feedback by reinforcing function and can perceive and interpret its environment, take actions, and learn through trial and error. (**Figure 1**) This type of learning allows machines to determine the ideal behaviour within a specific environment and context, to maximize its performance. Simple feedback reward mechanisms are required for agents to assert the acceptable behaviour. (44)

The ethical concern with machine learning occurs when the AI algorithms modify their own behaviour or 'learn' to perform complex task in ways that are not transparent to their developers. This is often referred to as "algorithmic opacity" or "black box" AI, as it can make it difficult to understand how a particular decision or outcome was reached. (12) Beyond transparency, there are multiple ethical challenges that arise from algorithmic opacity or black box AI such as unintended harmful consequences and negative social impact. For effective machine learning, there is a need for vast amounts of data, which include personal and sensitive information, therefore there is a risk of data breach or unauthorised access which could lead to privacy violations. Beyond the data breach these attacks could have serious consequences in situations where hackers manipulate equipment that operates autonomously to conduct malicious acts. To address these challenges, a multifaceted approach to addressing these challenges is needed along with regulation and governance. (45)

2.5.2 Deep Learning (DL)

DL is a type of machine learning where algorithms are arranged in layers, with each layer using the information learned by the previous layer to form a more complex and abstract representation. (46) It is an advanced approach to AI that can be used for tasks such as image recognition. In a deep image recognition algorithm, the first layer would focus on identifying basic colour patterns, while subsequent layers would look for shapes and more intricate combinations of colours and shapes. (47) The top layer would then be able to identify the actual object. DL is currently the most advanced form of AI architecture utilized today. It involves various complex algorithms, with Convolutional Neural Networks being one of the most widely used. These networks apply weights and biases to objects within an image, allowing them to differentiate and classify them. They are often used in tasks like object detection and image classification. Recurrent Neural Networks are another type of DL valuable algorithm in tasks such as speech and voice recognition, time series prediction, and natural language processing. These algorithms can remember sequential data. (47)

Long Short-Term Memory Networks (LSTMs) are DL algorithms capable of learning order dependence in sequence prediction problems. These are commonly used in tasks such as machine translation and language modelling. Generative Adversarial Networks (GANs) work by using two algorithms pitted against each other, each using the other's mistakes as new training data. They are used in digital photo restoration tasks and in creating deep fake videos[5]. Deep Belief Networks (DBNs) are unsupervised DL algorithms that use each layer as a hidden layer for the previous layer and a visible layer for the next layer. They are becoming increasingly popular in the healthcare sector for cancer and disease detection tasks. Recent news reports have raised concerns regarding the dangers of utilizing DL to develop generalized models capable of handling language-based tasks and natural language processing. (48,49) There are numerous examples of Large Language Models (LLMs) including chat bots such as Open AI's ChatGPT, Google's Bard, ErnieBot, ChatSonic and Bing's Chat. (49) However, there are significant risks to consider when utilizing these models: the possibility of bias, lack of understanding, and the risk of misleading outcomes. (49,50)

2.6 Big data in healthcare

Related to the concerns mentioned above are concerns arising due to so-called big data. Big Data refers to large, complex, and diverse datasets. These large complex datasets are generated from

[5] Deep fake is a type of artificial intelligence used to create convincing images, audio and/or videos.

multiple sources and are characterized by velocity in terms of the speed of processing and generating data, according to the demand and various formats of datasets. We could classify these datasets into two forms: unstructured (which contain errors) and structured data (cleaned-up). For unstructured data to be useful in AI healthcare, collected data must be filtered, organized, and validated before analyses. (51) Big data in healthcare, as shown in **Figure 2** shows that data can be gathered from multiple sources. However, these data can be obtained from search engines and social media outlets, and data might not necessarily be obtained from established sources like research studies, and Electronic Health Records (EHR). (51,52) There are critical ethical concerns raised by Big Data in healthcare which includes respecting patient's autonomy via proper adequate consent, ensuring equity, respecting participants' privacy, and questions about data ownership. (53) These ethical concerns will be further discussed in subsequent chapters.

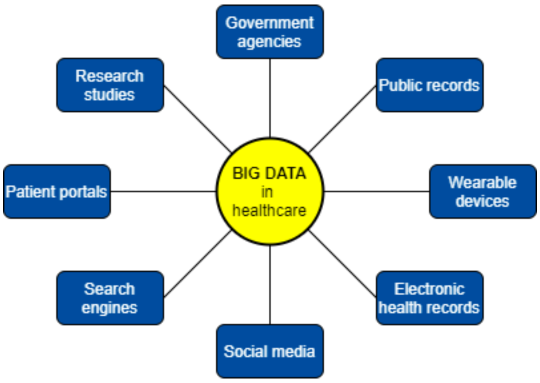

Figure 2: Examples of big data sources in healthcare (51).

2.7 Conclusion

The first part of this chapter introduced the descriptive background of the use of AI and provided an overview of technical terminologies, focusing on their benefits and applications in healthcare. In Africa, in particular, the implementation of AI technology in healthcare systems has immense potential due to the high disease burden, shortages of healthcare professionals, and limited resources in general. However, as highlighted in this chapter, the ethical considerations in the development of AI require stringent attention if we are to fully harness the potential of these technologies and create a more efficient healthcare system. Furthermore, the discussion in this chapter has further revealed the gaps and concerns about privacy and the security of the large scale generation of data sets of AI in healthcare and beyond. Addressing these gaps and concerns would improve both trust in, and trustworthiness of, AI, whereas failure to address them could have detrimental effects on the use of AI in the sector. In the next chapter, I use the Principlist framework to explore ethical considerations related to AI use in healthcare and the importance of collaboration between healthcare professionals and policymakers to ensure responsible and equitable implementation of these technologies.

Chapter 3

3.1 Overview

As explored in the preceding chapter, the advancement and implementation of artificial intelligence (AI) in healthcare has immense potential to improve the efficiency, accuracy, and precision of disease diagnoses. AI technologies, such as machine learning and deep learning algorithms, can analyze vast amounts of medical data, including patient records, laboratory results, and imaging scans, to extract valuable insights and aid in diagnosing various medical conditions. In this chapter, I use the Principlist framework to discuss some of the ethical concerns that arise in the context of using AI in healthcare. I have chosen this framework because it provides a structured and widely recognized approach to addressing ethical issues, relevant to my focus, such as (1) obtaining consent to store and use data (2), ensuring adequate attention is paid to safety and the need for transparency, (3) algorithmic fairness and awareness of algorithmic biases (4) data security and privacy (5) dignity and solidarity and (6) trust in healthcare and technology. (54–57) As mentioned in the previous chapter, if these concerns are not adequately addressed, patients may be misdiagnosed, and AI could cause harm, including the exacerbation of existing inequities in society. It is crucial to examine and fully consider these challenges to ensure that AI is used ethically and effectively in healthcare. This chapter will address the ethical limitations that pose a significant threat to AI in healthcare.

3.2 A Principlist Framework approach for ethical AI application in healthcare

Beauchamp and Childress' principlist framework has dominated both clinical practice and medical research for decades since its inception in 1979. (58) Their framework comprises four ethical principles that can be used to guide and ensure the ethical application of AI in healthcare settings. The principles are **Autonomy,** in the sense of self-governance, and on the basis of "respect for persons"; **Non-maleficence,** which stipulates that medical practice and research should not cause harm to patients and participants; **Beneficence** which includes both harm prevention and removal, where possible, as well as active promotion of wellbeing, and **Justice** – referring to fairness, both in terms of not discriminating but also in terms of ensuring the fair distribution of benefits and burdens. (58) We can also derive additional moral requirements, or rules, from this list of fundamental moral principles by reflecting on these principles. In keeping with the pluralist approach of principlism, these principles, and the rules they generate, are not absolute but rather *prima facie*; they must be fulfilled unless they conflict with a principle or rule that is equally, or more, demanding, in that context.

In keeping with respect for persons and their autonomy, most countries have adopted laws to safeguard and protect individuals' privacy and personal information. In South Africa, the Protection of Personal Information Act (POPIA) was enacted for this very purpose, to protect personal information. (59) However, some countries do not have data protection laws or adequate healthcare frameworks to guide AI in Healthcare and protect personal information. (59) Therefore, it is critically important to have independent regulatory bodies and government structures to enforce rules and ban the use of any technology that violates or does not adhere to regulations. However, such rules should be grounded in core values such as equality, human dignity, justice, non-discrimination, privacy, freedom, participation, and solidarity; accountability should also not be overlooked. Ethical principles and human rights are interlinked; however, rights are legally binding, whilst ethical principles might not be, depending on the context. Human rights law provides a robust legal framework for governments to guide international bodies and private and public entities.

3.3 Respect for Autonomy in Healthcare

Respect for autonomy refers to the importance of self-rule, self-governance, or self-determination (autos, "self," and nomos, "rule"). Put differently, autonomy in this sense refers to: "the ability to develop one's own conception of value and sense of what matters, to [develop] the values that will guide one's actions and decisions, and to make important decisions about one's life according to those values where one sees fit." (60) The concept of autonomy is widely considered to be important across different cultures and is often justified or grounded in appeals to human dignity, referring in its broadest conception to the intrinsic moral worth of persons, or seen as an integral part of common morality. The principle of autonomy in biomedical contexts, translates to the patient having control over their healthcare decisions and information. Health-care providers have access to sensitive unique information and therefore have a special duty to respect and protect this information in accordance with the patient's autonomy. The importance of autonomy in AI ethics, (61) and legal theory cannot be underscored enough. (62)

Patients have the right to make an informed choice regarding medical decisions relevant to them. Respect for autonomy creates the following obligations: informed consent, confidentiality, truth-telling, and effective communication. (58) There are severe ethical challenges and implications for autonomy that might arise with utilizing AI tools for decision-making in healthcare settings. Opaque and biased AI algorithms could potentially undermine patient autonomy, even when there are provisions to respect other requirements such as reliability, transparency and fairness. These challenges are associated with principles and values such as accountability and justice. (63) When considering how to respect autonomy in AI-driven medicine, there is a patient, a machine, and a healthcare practitioner; therefore, a 'patient–doctor' relationship has become a 'patient–machine–doctor' relationship. Patient autonomy and dignity may also be at risk in the healthcare sector due to opaque-biased algorithms. Since algorithms set values and rate various treatments, using AI in medical decisions may re-establish a paternalistic approach. For instance, the algorithms may recommend a medication that increases longevity for a patient even though they would prefer one that lowers suffering. In this way, AI engagement may compromise trust relationships between a doctor and patient as well as joint decision-making. (64)

For the application of AI technology in public health, large datasets are crucial to developing algorithms. One of the most sensitive special areas of personal information about a person is their health information. With algorithm development, this sensitive special personal information of patients will be saved and distributed around the network when using AI systems. The concern about potential data breaches give rise to challenges in privacy, safety, and governance. (65) Beyond this concern, as large datasets are necessary for AI systems to function successfully, patients are also concerned about personal information being collected without their knowledge or agreement. To this end, machine models and algorithms must be designed and overseen with a respect for patient autonomy in mind and must be able to maintain anonymity of patients. (63)

Fully adopting and deploying AI means that decision-making could be transferred to, or shared with, these technologies. AI technologies should therefore be designed to be sensitive to the importance of patient autonomy and human rights. While it is likely that healthcare practitioners' oversight would continue to be required when performing specific procedures and making certain clinical decisions, the shortage of clinicians in certain contexts is such that AI technologies could provide much-required assistance to healthcare practitioners. The presence of a clinician would then mean that AI technology decisions can be restricted or overridden by the clinician. (66) There are ethical limits to autonomy, however. While AI algorithms must be sensitive to the requirements of autonomy, this may also include the support, to a degree, of an individual’s right to make self-regarding choices which permit a degree of human self-harm or permit the individual to assume risks that might lead to harm. (60) For instance, a patient should be allowed to decline a surgical procedure if there are concerns about quality of life after the procedure. Similarly, human beings choose risky activities such as skydiving, ice skating, or car racing which to a certain extent could cause harm.

This matter is not straightforward, however, as in the case of increased use of medical robots, we can imagine future scenarios in which patients might request that robots provide them with assistance in dying. The question here, even in countries in which euthanasia is legal, would be whether an AI-powered robot should ever be given autonomous power to assist in ending a life. Moreover, a further complication would be if the robot was designed or constructed in a country that permits euthanasia but is being used in a country in which it is illegal and regarded as morally unjustifiable in an absolute sense. If the robot carries out this action, in such a context, the action can be categorised as illegal, however, the question then would be who should rightfully be held legally liable and morally responsible? There are a few stakeholders who could potentially be held liable for such an outcome, however, one might, rightfully, hesitate to delegate such a grave decision to machines in the first place. When assigning autonomy to these devices, care should be taken to ensure that the machines with the ability to make life-and-death decisions are limited.

As elaborated by Beauchamp and Childress, and mentioned above, the principle of respect for autonomy can generate several rules to guide action: always tell the truth, respect the privacy of persons, including protecting their confidential information, and, when asked, help others make important, informed, decisions. (67) Similarly, in the context of the use of AI in healthcare, respect for autonomy would then confer the following related responsibilities on AI systems and robots: informed consent, confidentiality, truth-telling and effective communication. Respect for autonomy in the context of AI also involves adopting an appropriate legal framework that protects personal information and confidentiality and ensures informed consent and privacy for data protection as a legal requirement. Storage of special information and processing of personal information should be protected by information or data law to maintain the confidentiality of patients whose information is stored on the technologies, thereby fostering their trust. Hefty penalties and fines should be implemented against any potential breach of these laws.

3.4 Non-maleficence

Medical practice is firmly rooted in the principle of *primum non-nocere* – first to do no harm. Primarily, the duty of a health care practitioner is to alleviate pain and suffering, help maintain quality of life and optimise health and well-being (Beneficence). However, in discharging their duties, they may inadvertently cause harm. This fundamental duty would be extended to applications of AI in healthcare. The potential applications of AI are vast and, as discussed in Chapter 2, have significant potential to benefit society, and more specifically, to help us advance global health systems. However, AI also presents several challenges and potential risks which can lead to intended and unintended harms. (67) One of the main concerns is the potential for AI to perpetuate or amplify existing biases and discrimination in decision-making processes. (67) For example, if a machine learning model is trained on data that contains biases, it can continue to make decisions that reinforce those biases. In a healthcare setting, bias could be defined as a systemic error introduced into study sampling analysis either consciously or subconsciously, making an inference favouring a particular outcome. (47)

Biases imply errors which result in a group being favoured, in some way, over another group which could implicate the outcomes or datasets. (47) Systemic inequities are perpetuated based on biased data sets such as age, race, gender, and ethnicity, which might limit the diagnostic and treatment performance of AI. Of course, data disparities have existed in the healthcare system long before the introduction of AI, and these issues are extremely complicated to resolve. (56) These biases are hidden in the algorithm created by humans, who may be completely unaware of problematic values or implicit biases that might inform their assumptions. (56,68) These errors could have grave consequences for individuals and communities who are unfairly affected. Along with unintentional harms, AI has the potential to be used for malicious purposes, such as in cyberattacks, the spread of misinformation, and to give hackers access to special information. (47) There are also ethical considerations around the development and deployment of AI systems, such as the impact on the patient-doctor relationship, discussed in the previous section, and the balance between the benefits and risks of AI-powered systems. (69) The developers of these algorithms must ensure that diversity of data upon which algorithms are based eliminates tendencies that can encourage prejudice. From an ethical standpoint, ensuring data diversity, understanding various social contexts and the risks posed to fairness by AI biases could help eliminate potential biases. (56,68)

Depending on the circumstances, it is difficult for human healthcare practitioners to decide where benefits end, and harms begin. For example, the justification for a termination of pregnancy (TOP), might be that this is in the interests of the pregnant woman or, as is the case in end-of-life situations, in ones' attempt to alleviate pain and suffering, the outcome may be death. As alluded to in the previous section, we should rightfully consider if it will ever be appropriate for AI to be tasked with such decisions. Alternatively, we could consider another example whereby an AI system which is used to diagnose cancer cells in the human body gives a false negative result for a disease which results in the patient not seeking a second opinion, only to realise that they have been misdiagnosed once the symptoms worsen. Depending on how AI is used in healthcare, it may be necessary to rethink our notions of responsibility and accountability. More specifically, we will need to consider who should be held liable when a poor judgement call, or learning error, involving an AI is made. Given the weight ascribed to the principle of non-maleficence in healthcare, consideration must be given to how AI technologies could be developed with this fundamental requirement in mind. However, it is important that we address these challenges and risks proactively, through ongoing research to inform innovative developments, ethical guidelines, and effective regulations and policies. This will help ensure that AI can be developed and used in a responsible and sustainable way that benefits society.

When it comes to AI use in Africa, we need to consider harm in the context of African moral and cultural worldviews which, while varying extensively, are characterised by insights which differ from other contexts. Thus, in addition to harm to individuals, there is also a need to consider collective harm, to families and communities in general. In African contexts, cultural considerations are crucial. (70) Culture can be interpreted as the manifestation of the unified teaching and understanding of values which direct us to what is considered acceptable and unacceptable conduct. It is the collective mentality of our societies. (71) As such, cultures are the vehicle for common ideas and shared perceptions. While culture should not be reified or uncritically accorded value, it should also not be discounted. When it comes to ethical AI standards, if we agree that cultural values shape them, it is to be expected that there will be differences between various cultural groups. The need to safeguard the range of cultural diversity and its associated particularisms from any attempt to enforce legal or ethical standards from outside entities must not be overlooked. This diversity should form an integral part of the discourse surrounding the ethics of AI. (70,72)

Beyond cultural concerns, there are several unique structural issues that are distinct to ethical AI in healthcare and relevant for a consideration of potential harms. (72) One of the most significant issues is the need for more infrastructure, lack of high-quality data to train AI and regulatory bodies to conduct oversights. (73) While the lack of AI regulatory law, data privacy and protection is seemingly a particular problem for African contexts, given general resource constraints and related challenges, it is actually not unique to Africa; it is a global issue. (74) Lack of regulatory oversight could lead high-income countries or organizations to take advantage of low-to-middle-income countries by providing them with sub-standard or even biased AI technology. (30) Furthermore, there are concerns that AI could promote inequity in healthcare which might lead to unfair targeting of specific groups or individuals based on their race, ethnicity, or politician connections. (30,73) Moreover, an algorithm trained on a biased dataset is inadequate in handling diverse data that represent the wider population. This can lead to the algorithm suggesting treatment plans that harm underrepresented groups. (75) Healthcare systems that have access to individual data, such as ethnicity and political affiliations, may be compromised or influenced for political gain. This opens the door for AI systems to be manipulated or intentionally programmed to target specific groups or individuals based on these characteristics. Additionally, medical resources may be unfairly allocated to individuals who belong to a particular political group or have close connections to influential politicians, while disregarding the needs of other individuals who may be equally or more deserving of those resources.

While there are a number of countries in Africa that have data privacy and protection regulations, individuals living in countries that do not have such regulations could be at risk of serious harms related to breaches or misuse of data. In the absence of privacy laws, an AI algorithm developer may collect, store, and use personal information in ways that violate people's privacy rights, thereby undermining their trust in technology. For example, without proper safeguards, sensitive personal information such as health records, and personal health information can be easily accessed, misused, or leaked. This can lead to identity theft, financial fraud, and other types of harm. The lack of specific legislation means that patients cannot be sure if their personal health information is being used only for the purpose for which it was collected, and not for purposes such as marketing or research. Moreover, when personal information is used for unethical or discriminatory purposes, it can undermine people's ability to control their own lives and participate fully in society. In addition, the absence of privacy and data protection regulations can also create barriers to innovation and competition, for smaller companies. In Africa, it is crucial that privacy laws for AI in healthcare be developed and implemented in a way that protects patient privacy while also allowing for the beneficial use of AI in medicine. This will require a balancing of the potential benefits and risks of AI and the creation of clear, transparent guidelines for its use in healthcare.

3.4.1 Harms and the responsibility gap in AI

Harm can result from acts of omission and commission. In medicine, an error may be defined as a commission or an omission with potentially negative consequences for the patient who had been wrongly diagnosed by a medical practitioner. Errors can be classified into two categories: system errors and Individual errors. (76,77) System errors involve mistakes made by practitioners, due to the system's inadequacies. Examples of system errors include cases in which a shortage of skilled practitioners results in healthcare workers having to work long hours, with resulting impacts on performance, faulty equipment, depleted oxygen supplies and so on. However, medical doctors share responsibilities for these errors with other elements in the healthcare delivery system. Individual errors are different in the sense that the doctor's own lack of knowledge, skills and attentiveness is the primary cause of the error, therefore, the practitioner is responsible for the error. With AI system errors could be shared; however, the question that remains unanswered is with whom? Individual errors attributed to AI systems are also 'their' responsibility, however, this type of error raises a layer of unresolved complexities in the sense that, if an AI system is to be found 'guilty' of a harmful error, who is then the fitting target of blame and liability. (45,78,79) In the case where resulting litigations are not resolved, how would the family or community members seek justice? These are perplexing questions that highlight the need for AI regulations within the healthcare environment.

A more speculative concern warrants consideration. It could be that at some point in the future, robots might be programmed to 'evolve', through machine learning, thereby displaying a kind of moral agency, or, at least, the appearance thereof. However before anything resembling moral agency could ever be ascribed to AI, certain criteria would have to be met. First, AIs must have original self-originating thoughts, reasoning and decision-making capacity to execute their thoughts and act with valid reasons. (45,76,80) Second, we would need to have some way of knowing, or proving that such capacities are self-originating in a way that satisfies an ascription of authentic agency. Beyond this point, AI's decision-making capacity needs to be ethical. The components of ethical decision making are highly complex, however, and include intentionality, attention to relevant values, sensitivity to contextual factors, and awareness of consequences, among many other factors. Humans are regarded as moral agents because of their ability to comprehend and consciously apply these ethical principles to their day-to-day activities. Moreover, key to moral agency is moral responsibility. In the context of morality, AI's lack of consciousness is of significance because moral sense is the core feature of human emotional experience – it guides our approach to complex situations and circumstances, and how we think and feel. Although these systems might behave autonomously and display certain attributes that are characteristic of moral emotions, and of significance for moral judgements, without moral sense, there is no true moral agency. Moreover, if AIs are not moral agents, how will we ascribe legal liability and moral responsibility in cases where there is an error in the system that causes harm, injury or death to human beings. I shall further explore the questions of responsibility and liability of AI in chapter 3.

3.5 Beneficence

Beneficence as a principle refers to the removal or prevention of harm as well as the more proactive obligation to do and promote good or the interests of others. (58) The principle is, however, complex and difficult to unpack because we need to set a boundary between when we are doing good by subjective standards and when we are doing good by an objective standard. (67) Put differently beneficence faces thresholding challenges in defining the scope of our obligations. Both humans and AI machines, as far as they are used in healthcare by humans, have a responsibility to provide beneficial treatment and to avoid or minimise harm. According to Beauchamp and Childress (2013) “the rules of beneficence are to protect and defend the rights of others, prevent harm from occurring to others, remove conditions that will cause damage to others, help persons with disabilities and rescue persons in danger.” (67)

Beneficence has three critical dimensions in patient care: clinical competence, risk-benefit analysis, and paternalism. Clinical competence in medicine is a critical component of the combination of knowledge and skills acquisition during medical training. However, maintaining competence in AIs could pose a challenge to healthcare settings. For example, in South Africa, medical doctors must commit to lifelong learning through continuing professional development (CPD) activities. However, when we consider AI machines, the question that comes to mind is how would these technologies fulfil this requirement? AI systems employ machine learning and part of this process is the freedom to make errors as the system learns and improves. Meaningful human oversight of this process, in tandem with ethical sensitivity would therefore be crucial to maintain the principle of beneficence. As mentioned in the previous section, for AI technologies not to harm patients, they should meet stringent regulatory requirements for safety, accuracy, and efficacy before deployment, in the same way that medical professionals require ongoing supervision in the early parts of their careers, and continued development throughout their careers. The principle of beneficence emphasises the urgent need for a regulatory body that would ensure quality control and quality improvement for the AI systems when necessary. Furthermore, continuous performance evaluation should be constantly carried out to ensure that clinical competence requirements are met and to safeguard against harm.

Risk-benefit analysis in medicine refers to balancing the principles of doing good, or beneficence, and non-maleficence and is critical to achieving net benefits in the healthcare setting. Therefore, regulatory bodies have a more prominent role in ensuring that the potential of AI technologies to cause harm is minimised whilst the benefits are maximised.

According to Beauchamp and Childress, paternalism is defined as "the intentional overriding of one person's known preferences or actions by another person, where the person who overrides justifies the action to benefit or avoid harm to the person whose pretences or actions are overridden." (2013) Paternalism in AI healthcare would therefore refer to decisions made by AI systems that prioritize the well-being and interests of patients over their autonomous wishes. Paternalism can be grouped into two forms, active and passive paternalism. (67) Active paternalism could occur if an AI chose to override a patient’s choice to decline an intervention because the AI judges the intervention to be beneficial whereas, passive paternalism could arise if an AI refuses to perform an intervention or provide treatment for reasons of patient-centre beneficence. (67) Advocates of AI paternalism argue that both forms would be justifiable on the basis that AI systems can make decisions that are more accurate and objective than humans, and that these systems can help reduce medical errors and improve patient outcomes. (81)

The use of AI in critical care settings, such as in triage or treatment recommendations, may justifiably require a paternalistic approach due to the high stakes and life-or-death decisions that are involved. On the other hand, critics of AI paternalism argue that it can undermine patient autonomy, or that autonomy should never be overridden based on beneficence, except for when there is no autonomy present. (60) This reflects a widely held position regarding paternalism and is not of course specific to AI. However, there is also concern about the impact of paternalistic AI on trust in healthcare providers, and that it may lead to unequal access to care, as AI systems may prioritize certain groups of patients over others based on pre-existing biases or systemic factors, as discussed above. (60) Overall, the question of paternalism in the context of AI in healthcare is a complex and controversial issue that requires careful consideration of the ethical implications and potential consequences of AI decision-making in healthcare. With no clear guidelines and privacy law – If poorly regulated in other words – paternalistic AI decisions could pose significant threats to patient autonomy. (60) On the other hand, if these technologies are not regulated or no medical practitioner exercises oversight and raises queries for a given outcome, treatment could be hindered, and health inequity might be widened.

3.6 Justice

Justice as a principle in healthcare refers fundamentally to a concern for fairness and the fair treatment of patients. (58) Justice as an obligation can be conceptualised in three ways: legal justice (respect for morally acceptable laws), rights-based justice (respect for people’s rights), and distributive justice (fair distribution of limited resources). (82) Legal justice includes the use of laws to defend patients’ and doctors' rights. (67) When patients are harmed while receiving an AI and/or ML based treatment, a sense of injustice is created. Failure to resolve this unfair treatment may result in litigation. Knowledge of and respect for morally acceptable laws is therefore an ethical requirement in terms of upholding the principle of justice. (67) There is an urgent need to develop a legal framework with clear ethical guidelines, clarify liability and responsibility issues relevant for AI in healthcare. (83,84) It is therefore critical to understand the laws and their impact on healthcare.

Rights justice may be regarded as an entitlement that is considered valuable and therefore, a claim to what is regarded as a fundamental or absolute right does not require justification. (85) Examples of rights relevant to the healthcare context include the right to healthcare, the right to be seen on time as per appointment, among others. These rights are reflected in the patients' Rights Charter, (86) the South African Constitution, (87) and the South African Bill of Rights. (88) Aside from the POPIA, (59) which does not directly regulate AI, there are no laws or ACTs that guide or regulate the use of AI in healthcare settings in South Africa. Rights, entitlements, responsibilities, and obligations are connected. A patient is entitled to enjoy a specific right, the doctor is obligated to provide service, linked to the right, and both the patient and doctor have responsibilities. The rights of patients are accompanied by corresponding obligations. In the context of patient-doctor relationship, the patient has a right to competent treatment from a health professional and the latter has the obligation to treat the patient to the best of his/her ability. In the same vein, the patient has the responsibility to follow the doctors' orders, such as prescriptions and necessary precautions. With regards to AI, it is crucial to determine whose responsibility it is to ensure that all parties involved are accountable for the care and services provided.

According to Beauchamp and Childress, distributive justice is defined as "fair equitable and appropriate distribution of benefits and burdens determined by norms that structure the term of social co-operation." (58,67) The concept of distributive justice is focused on the distribution of scarce resources in terms of fairness (what one deserves, giving to each his or her due). The healthcare industry is fraught with challenges arising from limited resources. This was particularly evident in the peak of the pandemic, with both LMIC and HIC countries experiencing resource challenges. AI stands to address such challenges, but unregulated advancement may further exacerbate inequity in the healthcare industry. AI in Healthcare should strive to be accessible, inclusive, and equitable irrespective of race, age, and gender. (85) The right to healthcare is a fundamental human right. Thus, AI regulators are obligated to ensure that basic human rights are not violated, and that manufacturers maximumly appreciate these rights. Furthermore, manufacturers of these technologies should have employees from diverse backgrounds, cultures, and ethnicities to ensure that all cultural and ethnic perspectives are considered, as discussed above. When developing AI algorithms, contexts is critical in the different healthcare settings, therefore, representation of datasets in a specific group should be directly proportional to its population size. Accurate representation would ensure that AI achieves accurate and quality results in that population. (85)

Given the discussion of bias in previous sections, it is clear that bias is thus one of the major threats to the equitable inclusion of AI in Healthcare. AI in Healthcare should not inadvertently increase inequity in access to healthcare. As healthcare is a means of increasing well-being, welfare and advancing the public good, public healthcare should be accessible for all as a matter of justice. Disparities between AI-service providers, patients, and public and private healthcare should therefore be monitored and controlled. The rapid emergence of technologies emphasizes the need for artificial intelligence law, which would help ensure justice and address privacy concerns in terms of who should have control over the data collected. This law should protect the rights and welfare of vulnerable populations with severe penalties if bias emerges.

3.7 Trust in AI Systems

Trust, or trustworthiness, is a value, or virtue, that doesn't sit definitively under any one of the four principles, but is nevertheless, a fundamental component of the practice of medicine. A relationship of trust involves two or more individuals or parties, where one, or more, is in a position of vulnerability or dependency (the trustor) on the other (the trustee) for different reasons. (89) In a medical context, trust is critical because patients have no choice but to put themselves in the hands of doctors who have the specialist knowledge that patients need. The relationship between patients and physician is thus one of dependency or asymmetry, whereby patients must assume that the doctor has their best interests at heart. Previous abuses of trust in clinical and research contexts indicate the tenuous nature of this relationship. The case of Henrietta Lacks, whose cell line came to be known as HeLa is illustrative here. Lacks was diagnosed with aggressive cervical cancer, and during treatment, some of her tissues were given to a researcher without Lack's knowledge or consent. (90) These cell lines were immortalized and used for great commercial gain. The Tuskegee syphilis study[6] has come to symbolize the most egregious abuse of authority on the part of medical researchers. Tuskegee has also come to serve as a point of reference for African Americans distrustful of those with power, emblematic of the history of a people enslaved and then subject to social, legal, and political oppression after the end of formal servitude. (91)

As discussed in previous sections, trust is also critical for the deployment of AI technologies at a wider, societal level. In the case of trust concerns related to public health care data, there are numerous contexts where this is evidenced. At the peak of the COVID-19 pandemic in South Africa, there were

[6] The Tuskegee syphilis study was conducted from 1932 to 1972 with the intention of observing the natural progression of untreated syphilis. However, participants were not informed or given the option to consent, and were not offered treatment even when it became readily available.

low levels of public trust in the South African government to protect, and not misuse, health data, which arguably led to minimal use of the COVID-19 tracking and tracing alert app. (86) Trust in the commercial realm also seems lacking. For example, there are no guarantees that a company like Google would not sell or reuse their morbidity data for profit gain or to any third party for other purposes. (56,79) Trust plays a critical role in human-AI relationships due to the perceived risks associated with AI. The complexity and nondeterminism of AI behaviors, by way of machine learning, and its future role in the workplace is also concerning to many. Moreover, widespread concern that AI will eventually take over a wide arrange of human occupations and jobs as well as fundamentally transform the structure of organizations also fuels mistrust of these technologies. (56,79)

While trust is a major area of ethical concern, it is a complex issue in the relationship between AI and humans. Uncertainties exist regarding decision-making and the norms and values guiding it. The acceptability, and success of AI in healthcare will depend increasingly on trust in such systems, but questions remain about who to trust and how to establish a trust relationship. While human interaction with AI outside healthcare may be less ethically uncertain, healthcare poses higher stakes due to the potential life-threatening consequences of even simple mistakes. (79,92,93)

The major dilemma of AI in healthcare is that to harness its full potential, some machines or systems must be allowed to work autonomously and make decisions with less human input. This will, however, require well-placed trust in such systems. If we are to imagine a society without trust in healthcare practitioners, teachers, pilots, and drivers, this would be a context in which significant resources and time would be required to ensure duties are done properly. (94) The enormous cost of supervision would be burdensome and could come at the cost of efficiency, leading to a dysfunctional society. Therefore, for society to use AI in the most optimal way possible, there must be well-founded societal trust that machine learning algorithms will indicate the best decision to diagnose diseases and identify potential cures, and these algorithms must be free of bias. (79,92,93) The fostering of trust thus has instrumental value insofar as the fact that constantly supervising a machine learning algorithm in its decision making would require considerable time and resources, resulting in a lack of feasibility and defeating the purpose of using such systems in the first place. (94)

However, the other side of the coin, is that complete lack of supervision in healthcare settings, where clinical decision-making is not clear-cut and can be fraught with ethical dilemmas, at the best of times, is also problematic. Lack of oversight might lead to serious harm, breaches of privacy and security, and compromise of values that underpin our societies. It is essential to identify an effective way to build trust in digital technologies in healthcare to harness their value while protecting fundamental rights and fostering the development of open, tolerant, just information societies. (95,96) In a human and AI combined healthcare facility, trust is crucial when we consider the complex relationship between a patient-physician relationship. However, the question is: how do we find a balanced level of trust? To answer this question, in the short term, design could play a crucial role in addressing the problem e.g., pop-up messages alerting users to algorithmic search engine results that have considered the user's online profile or flagging that the outcome of an algorithm may not be objective. In the long term, there is a need for an infrastructure to reinforce norms and values such as fairness, transparency, and accountability across all healthcare sectors. (95,96)

3.8 Conclusion

In conclusion, this chapter emphasized that the application of AI in healthcare must be guided by a principled ethical framework that considers autonomy, non-maleficence, beneficence, and justice. This framework must ensure the protection of patients' rights and well-being. Beauchamp and Childress' principlist framework provides a solid foundation for identifying ethical concerns and informing ethical decision-making in the context of AI in healthcare. Beyond the principlist framework, there is a need to develop robust regulations, privacy laws, and oversight mechanisms to protect patients' rights and ensure the responsible and equitable use of AI technologies. By adhering to these principles and addressing the challenges and risks associated with AI in healthcare, we can harness the potential of AI to improve patient care and advance global health systems. In the following chapter, I will conduct an in-depth discussion of how the three famous theories: Kantian ethics (deontology), consequentialism (utilitarian theory) and virtue ethics can inform the question of moral robots in healthcare.

Chapter 4

4.1 Overview

In the previous chapter, I provided an overview of the ethical concerns associated with the use of AI in healthcare, using the Principlist framework. In this chapter, I build on the technical definitions provided in chapter 2, and some of the points raised in chapter 3, to consider some of issues associated with the use of robots in healthcare and the development of morally competent or 'ethical' robots**.** I also applied the three famous ethical theories namely: consequentialism, deontology, and virtue ethics to analyse the development of a morally competent robots. Finally, I explore the practical implementation of these theories in a robot and/or machine to empower it with the ability to make ethical decisions.

4.2 'Moral' Robots

For AI to be used effectively, it is arguably important for it to be programmed or trained to be sensitive to human moral norms, values, and ethics.[7] This could include ensuring that the decisions or 'behaviour' of such systems is not at odds with values and principles such as privacy, fairness, non-discrimination, and human dignity. AI systems that lack an 'understanding' of these values and ethical considerations could produce unintended consequences, such as biased decision-making or violations of privacy. AI can either be a system contained within a computer, or it can be housed within a machine or robot. As discussed in the chapter two a robot is a machine engineered to execute a sophisticated sequence of tasks. Robots can be designed for a wide range of purposes, including manufacturing, transportation, medical, and research applications. They can take various forms, including stationary machines, mobile machines, and even robots that have a human form. Robots can be controlled by a computer program or by a human operator, and can be equipped with sensors, cameras, and other devices that allow them to perceive their environment and respond to it in real-time. They can also be programmed to make decisions, carry out tasks, and perform actions based on input from their sensors and their programming.

The goal of designing a 'moral' robot is to create an AI system that can operate in a manner that is consistent with human ethical and moral norms, including being trusted to make decisions that are in keeping with society's interests. (97,98) While the nature of morality is a complex and multifaceted

[7] Some have argued that creating an ethical robot is a bad idea https://alumni.berkeley.edu/california-magazine/online/good-bad-and-robot-experts-are-trying-make-machines-be-moral/#

topic that has been debated by philosophers, theologians, and social scientists for centuries, there are certain components of, and intuitions about, morality that are widely recognized and considered uncontroversial. These include:

- Principles and values: Morality is often grounded in a set of principles and values that define what is considered right and wrong, good, and bad, just, and unjust. These principles and values can vary across cultures and societies, but they often include concepts such as fairness, justice, equality, and respect for others. (67)
- Moral emotions: Morality is often tied to the capacity for empathy and compassion, the ability to understand and feel the experiences and emotions of others. This component of morality is often seen as essential for promoting prosocial behaviour and reducing harm to others. (67)
- Consequences: A common position in moral theorizing is that the rightness or wrongness of an action depends on its outcomes. This approach to morality, known as consequentialism, argues that the consequences of an action should be evaluated in terms of their impact on the well-being of individuals and society. (67)
- The nature of the action itself: Another dominant approach to morality holds that certain actions are wrong, in an absolute sense, regardless of any contextual factors, including their consequences. This approach to morality is often associated with the idea of moral duties and obligations, and with the view that some actions should be avoided or prohibited because they violate the rights of others. (67)

While these components are considered uncontroversial, there is ongoing debate about which of them should receive more weight in our moral decision making, and about how these theories should be applied in practice. Nevertheless, they provide a useful starting point for thinking about the challenge of designing robots that are sensitive to ethical concerns. One of the main differences between human persons and non-human persons, including robots, is the formers' ability to identify an action or behaviour as either moral or immoral, to understand the reasoning behind such a judgment, and to respond appropriately. This is a result of the capacities for self-awareness, introspection, and empathy that are unique to human beings. While robots can be programmed to follow moral and ethical rules and principles, they lack the capacity to understand the underlying moral reasoning behind these rules, and they do not experience emotions or moral sentiments in the same way that humans do. (99) However, this does not mean that robots cannot be designed to make morally congruent decisions or to respond appropriately in ethical situations.

In fact, researchers and AI developers are working on developing AI systems that can incorporate such moral and ethical considerations into their decision-making processes, and which can be programmed to act in a way that aligns with human values and ethical norms. (99) By incorporating principles in the algorithms/data, such as fairness, transparency, privacy, and non-discrimination, these AI systems can be designed to operate in a manner that is likely to be more consistent with human ethical and moral standards. (57,99) Examples of human dispositional attitudes include admiring what is praiseworthy, defending the vulnerable, accepting responsibility and being held accountable, demanding justifications, or accepting apologies. (57,99) We reside in societies or communities where members observe numerous moral norms and values; these norms and values are taught through explicit instruction and passed down from generation to generation. A competent moral agent possesses an understanding of human norms, values, and virtues that are appropriate for their context. A moral agent could, however, engage in ordinary moral information processing to arrive at a moral decision, but still perform an act that is contrary or considered unethical. (97) For example, a pharmaceutical company that develops and produces a vaccine during an outbreak is acting in accordance with public well-being and engaged in an ethical action. However if they then hike the prices of purchase of this vaccine, to increase profit, this would be considered unethical insofar as their profits are being made at the expense of the lives or health of those who then cannot access these vaccines.

When it comes to developing *autonomous* moral robots, there are a few questions that warrant consideration. The first question is whether robots have the capacity to mimic the moral behaviour of human beings. (97,100) The ability to carry out these functions falls squarely within cognitive computing science, contextualising, integrating and practically theorising with machine learning models. The second question is: which ethical theories and moral values should be considered when building a morally competent robot, and which moral standards should robots be programmed to adhere to, in cases where there are clashes between two mutually exclusive approaches? The third question is, should we even attempt to develop morally competent robots? This last concern aside, and, in answer to the second question, some of the most important moral considerations include firstly: designing and programming robots to be responsive to respect for the dignity and worth of all human beings. This means ensuring that robots should not act in a way that causes harm or suffering, exacerbates inequality or is discriminatory. (100) Secondly, robots should be designed to protect the privacy of individuals, including avoiding collecting, using, or sharing personal information in ways that are unethical or violate privacy rights. (100) Thirdly, robots should be designed to make decisions that are fair and unbiased, including avoiding perpetuating or reinforcing discriminatory practices or stereotypes. (100) Fourthly, robots should be designed to operate in a manner that is responsible and be able to be accountable at the level of having the capacity to consider the potential consequences of their actions. (100) This includes being transparent about their decision-making processes and allowing for human oversight and intervention when necessary. This would, of course, require that robots are designed to be transparent in their operations and decision-making processes, and to allow for human inspection and understanding of how they work. Robots should be designed to consider the impact of their actions on society, and to promote the common good. (97,100) These ethical standards and moral values, which are informed by a range of ethical theories and philosophical perspectives, could be a start in helping to promote the ethical behaviour of robots.

4.3 "Moral' robots in healthcare

As mentioned in the previous chapter, in healthcare settings, the patient-doctor relationship depends fundamentally on trust. If robots are to be successfully utilised in healthcare settings, patient trust in such systems will be equally crucial. When dealing with robots in the healthcare setting, patients or clients would likely be concerned about moral competencies, moral expectations, and moral standing. (100) Any patients visiting a clinic or a hospital that uses autonomous robots would want to be certain of the technical competency of the robot in order to trust its ability to safely perform the assigned procedure. However, it is not only the presence of technical competence that informs a relationship of trust, moral competence also plays a role. As ascription of moral competence in the case of a robot might refer to a perception that a robots is able to behave responsibly due its perceived ability to make judgements or decisions that are congruent with those we regard as ethical and in a manner that implies awareness of the consequences of its actions. (101–103) There should also be a balance between both technical and moral competency for effective and ethical operation. (101–103) The former ensures efficiency, while the latter ensures alignment with ethical considerations and respect for human values. (101–103) Patients would inevitably also have moral expectations from their robotic doctors; these expectations stem from well-supported speculation that autonomous healthcare robots would be treated like their human counterparts, regardless of whether this is unintended or anticipated by the robot manufacturer. (104–106)

As briefly alluded to in the previous chapter, there is also the question of moral responsibility and blameworthiness, and the related issue of liability, in the case of errors or unintended harms. The more that robots operate autonomously, the greater the possibility of errors, particularly during the learning part of their programming. The awarding of autonomy creates a responsibility or obligation that must be fulfilled in some way, hence the need for robust ethical standards to guide computing abilities, and functional capacities, to minimise errors in the development of such robots. (97) In the case of errors, and insofar as these robots develop to the extent that they are able to act autonomously with minimal supervision, the issue of who should rightfully be responsible will require settling as this will have implications for potential cases of medical litigation. (107)

However, the question of who should be held responsible for the actions of autonomous robots is a complex one and there is ongoing debate about how to address the so called "responsibility gap" that might arise in the case of errors or harms. (45,76,108) There are several factors that can influence the answer to this question, including the specific design and capabilities of the robot, the context in which it is being used, and the legal and ethical norms that apply in that context. (107) In some cases, the manufacturer of the robot may be held responsible for any harm caused by the robot, particularly if it can be shown that there were defects in the design or manufacturing of the robot. In other cases, the user or operator of the robot may be held responsible, especially if they failed to properly supervise or monitor the robot's actions. (45,76,108) In the case of healthcare facilities or departments of health, they may also be held responsible if they failed to adequately oversee the use of robots in their care, or if they failed to properly implement policies and procedures for ensuring that robots are used in an ethical and safe manner.

It is unlikely that the robot itself will ever be a fitting target for our ascriptions of responsibility in the absence of consciousness, sentience, or a capacity to be better or worse off. If the latter capacities were attained by robots, and there was some way of definitively knowing that they had been attained, it might make sense to consider robots as having a moral standing of some sort. While there is currently no legal framework that recognizes robots as having a moral status of any kind, some philosophers and legal experts have argued that as far as robots become more autonomous and sophisticated, this may warrant further attention. (109–113) In the event of a medical dispute involving a moral robot, in the sense stipulated above, it could be treated as a medical malpractice case, and would be subject to the same legal and ethical standards that apply to other medical devices and technologies that malfunction. (105,111) The outcome of such a case would depend on a variety of factors, including the specific circumstances of the case and the laws and regulations that apply in that jurisdiction. (105,111)

In the section that follows, I discuss human interaction with robots, the issue of the moral competency of robots, and how ethical theories might assist in developing AI robots suitable for healthcare contexts.

4.4 Human interaction with autonomous robots

Human beings are complex creatures with human interaction through communication being a case in point. This is because in addition to language, we communicate through body language, tone of voice, and sometimes by what is implied or unstated. (114) While robots outperform human beings in various cognitive tasks, i.e., much faster processing speeds, they are unable to perform tasks considered basic from a human perspective. This is subject to continual and rapid change of course, with large language modules or 'chatbots,' like Open AIs Chat GPT now able to process certain cues or respond to questions or requests in ways that are deemed appropriate from a human perspective. (114)

As human interaction with robots becomes more common, conflict with machines, or between humans, due to machines, is likely to increase. The ability of AI to flood the internet with fake news and misinformation is an example of the latter risk which may have dire consequences. In terms of how we relate to robots, a study was conducted to evaluate whether parents prefer robot tutors that are polite to their children, to robots that use imperatives in their instructions. (115) The authors observed that initially, people did not have any preference when approaching the robots, but once they interacted with both robots, they preferred the polite robot. Another example of a social norm violation was that of "truth-telling" in which a robot announced itself as the winner of a "rock-paper-scissors" game played against a human counterpart even though it had been defeated by its opponent. It has also recently come to light that chatbots, in this case an earlier iteration of Microsoft Bing's language model, have used abusive language and threatened people in human-AI conversations when the system recognises disrespect or insult in their interactions. (116) The introduction of autonomous self-driving cars has also raised moral questions such as how such cars should be programmed in cases of inevitable collision with human beings; who should be avoided and thus saved – to do nothing is not an option. (117) Moreover, in Russia, during a tournament, a chess-playing robot was shown to break the finger of a 7-year-old. (118) Opinions were divided as to the reasons for the cause of the malfunction, however. While the tournament official claimed the child's violation of the rules of engagement by taking his turn too quickly caused the computer's reaction, others claimed that the robot mistook the child's finger for a chess piece. Either way, there was a child with a broken finger. Without staking a position in this debate, as interaction increases, there is likely to be more potential for errors and harms, hence a stronger case for policies and legal regulations for ethically competent robots.

4.5 Developing morally competent and ethical robots for healthcare contexts

Turning to the question of how designers of robots could ensure sensitivity to values considered crucial to human beings, several key steps and considerations are required, particularly in terms of their use in healthcare contexts. For instance, robots might be used for tasks such as monitoring vital signs, administering medications, or assisting physiotherapists. Once the intended use has been determined, ethical and moral principles that will guide the design and programming of the robot should be identified. (119) These principles could include respect for patient autonomy, confidentiality, and informed consent, just to begin with. (119) It is also crucial to consider the impact of the robot on healthcare providers and patients when developing morally competent robots in healthcare. For example, robots might automate certain tasks that were previously performed by healthcare providers, freeing up time for more complex tasks, however, this could also result in job loss for healthcare providers.

Once the appropriate ethical principles and values have been identified and justified, the robots must be trained or programmed to adhere to them. The developers must take responsibility for continuously evaluating and adjusting the programming of the robot, to ensure that it remains in line with ethical norms and moral principles, given the continued development and improvement of the underlying technologies. This might involve making updates to the robot's programming or incorporating new ethical norms or moral principles as challenges present themselves. In the sections that follow, I consider some of the dominant ethical theories that could be used for the task of designing a morally competent robot in this regard, as well as some of the potential problems that may arise.

4.6 Deontological considerations

The first ethical approach to consider is the view that our rightful moral focus should be on assessing the permissibility or impermissibility of actions themselves, regardless of contextual factors, such as the consequences they produce. On this account, ethical action is primarily based on adhering to rules, including not performing forbidden actions. These rules can be applied to AI machines. (120) Hutler et al. suggest that robots could be programmed to follow a set of basic rules to exhibit ethical behaviour, for instance, ‘don’t kill,” “do not deceive,” and “obey the rule” and so on. (121) While there are different forms of deontology the most influential deontological approach is that of the German Enlightenment philosopher Immanuel Kant. Kant’s two formulations of the categorical imperative stipulate to: “Act only on that maxim through which you can at the same time will that it should become a universal law and Act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means but always at the same time as an end.” (120) Importantly, the consequences, whether positive or negative, have no bearing on whether we should perform the action, the only criterion of relevance is the ability of the action to meet the requirements of the categorical imperative. The categorical imperative is famous for grounding the respect for the dignity and fundamental equal moral worth of all human beings, which has numerous implications for human rights and obligations. (122,123) It is also powerful in its stipulation that individuals should act in such a way that they would accept their actions to be universalised. These ideas are crucial for the development of AI in healthcare because they provide a framework for ethical decision-making. As paraphrased by Sandel, "Every person is worthy of respect, not because we own ourselves but because we are rational beings capable of reason; we are also autonomous beings capable of acting and choosing freely.” (Sandel 2009) Therefore to act freely for Kant means to exercise our capacity to legislate for ourselves as morally rational beings.

The direct translation of Kant’s philosophy to robots would mean that insofar as more robots become autonomous, there might be a desire for them to be programmed for the capacity to make moral judgements at some level and to be aware of areas where they fall short, so as to improve their ‘moral’ capacity. To ensure that their decisions are made in an ethical and responsible manner, the categorical imperative can be used as a guide. (58) if an AI system is being developed to make medical diagnoses, it is important to ensure that the algorithms used to make these diagnoses are transparent and can be willed as universal laws in the Kantian sense. For example, if a developer manufactures an algorithm to analyse medical images to diagnose lung cancer, the algorithm should be transparent and explainable to healthcare practitioners. (13) The explanation could be in the form of a heatmap with colour codes to show where the lung is severely affected. Therefore, to strengthen this explanation, back-end algorithms must be representative and free of bias. (56) If the AI system were to make a misdiagnosis, it should not be because of biased data or algorithms, but because of a lack of available data or limitations in the AI system's design.

The categorical imperative also requires that individuals treat others as ends in themselves, rather than merely as means to an end. This principle is particularly relevant to the development of AI in healthcare, where patients are often seen as sources of data rather than individuals with unique needs and desires. To ensure that AI is developed in an ethical manner, it is important that patient autonomy, privacy, and well-being are respected and protected. The ideas underpinning the categorical imperative are crucial for the development of AI in healthcare because they provides a framework for ethical decision-making and ensure that patients are treated with respect and dignity. By incorporating the principles of the categorical imperative into the development of AI in healthcare, healthcare organizations can help ensure that AI is used in a responsible and ethical manner. (58)

However, it must be noted that although this approach makes an excellent case for the rational basis of morality and moral status, deontology has serious shortcomings for AI in healthcare in that it does not translate from the level of abstract theory into practical action guidance in particular situations. Many of the shortcomings of the approach itself would be potentially carried through in its application in AI. In particular its disregarding of the consequences and context of our actions and moral decisions would be the major concern here. Moreover, absolute forms of morality like this do not recognize moral dilemmas. Many decisions in medical contexts take the form of real dilemmas where there is not one clear course of action and absolute moral rules might clash. The other limitation of programming moral robots with this approach is that it would not cultivate moral wisdom in robots but would lead to inflexibility, that, at best, might not be helpful, and at worst might have more serious implications. For example, In the case of the robot and the teenager in the chess game, the robot has been programmed to look for pawns, and if we consider the case of the robot that mistakenly announced itself as a winner of the chess game, it reveals the shortcomings of its programming. This shortcoming underscores the importance of developing all-encompassing algorithms for artificial intelligence systems. The latter case suggests a lack of truth-telling, while the former suggests rule-abiding robots that cannot differentiate between a finger and a chess pawn. The deontological approach struggles to provide guidance for AI to make prudent decisions. This is because an emphasis on rule following, at all costs, runs into concerns about potential cases in which rule-following directly endangers the life and safety of humans.

To further explore this approach, let us consider the following thought experiment which illustrates the moral complexity that is often encountered in practical contexts. (Adapted from Ethics of Artificial Intelligence). (124) Mr. X is the new digital officer in the city of Cape Town. He is approached by the new digital minister Mr. U and asked whether the city's health care organization should move from "reactive healthcare to preventive healthcare". Mr. X proposes a novel sophistical learning system that could assist healthcare authorities to predict the possible health risks of Capetonians. (124) This novel AI-powered machine learning method produces predictions that combine and analyze multiple sources of medical and healthcare systems. With the help of big data, many criteria datasets will be collected through medical aid companies and provincial and district hospitals. Therefore, high-risk individuals would be identified and prioritized for precise healthcare facilities. These high-risk citizens would be proactively invited for doctor's appointments and follow-ups with the adequate required treatment. (124)

On the one hand, when we consider this scenario, there are numerous advantages. For instance, this proposed intervention would provide better 'impact estimation' and more effective planning for basic healthcare services. It would also prevent future unmanageable illness with potential to improve the health quality of citizens. Most importantly, this method would facilitate preventative healthcare with the potential to significantly reduce social and health costs. On the other hand, there are a number of concerns and problems, including legal and ethical issues regarding privacy, security, consent, and the future use of data. There are also fundamental questions about the city's role in healthcare provision that need to be considered. For example, what is acceptable prevention and what is non-acceptable intrusion, given other goods that we value such as privacy and autonomy? In addition, does the city have the right to use sensitive medical data to identify high-risk patients? How would consent be obtained from people who cannot consent because they are ill? What would happen to those individuals who cannot provide consent? How would consent be given, and what would happen to the citizens that refuse to give consent? Beyond these questions, the digital and real-life scenarios differ; for instance, if someone is ill, people within the vicinity have a right to call an ambulance without having explicit permission or consent, but in the digital space, contacting an ambulance because one has access to their medical information could be considered as an intrusion of privacy and breach of POPIA. A complex example such as this shows the kinds of ethical challenges that arise in real life contexts. If a robot, programmed to follow moral rules, were to be tasked with decision making or involvement at any level in a complex scenario such as this, it is unlikely that it would be up to the task.

4.7 Consequentialism

The second approach, which could inform the design of ethically sensitive robots is consequentialism. Consequentialist approaches refer to moral theorising that considers the outcomes of an action as the only concern in assessing the moral rightness or wrongness of an action. The most well-known consequentialist theory, utilitarianism, argues that the right action is the one that maximizes overall happiness, pleasure or wellbeing and minimizes overall suffering or pain. Jeremy Bentham, considered the father of utilitarianism, characterized utility as "that property… (that), tends to produce benefit, advantage, pleasure, good, or happiness…(or) to prevent the happening of mischief, pain, evil, or unhappiness to the party whose interest is considered." (125) In the case of a surgical robot, for example, this implies that the robot evaluates all options, considering success, prognosis, opportunities, and benefits for all parties in context for the best decision. (126)

In the case of the thought experiment, mentioned in the previous section, a utilitarian would advise that Mr. X and the new minister should follow through with their plans considering it would maximize the most benefit over harm for everyone affected. There are various permutations of utilitarianism that have been developed to overcome some of the flaws of the original version, some of which reflect fundamental disagreements. For instance, should actions be chosen solely based on their results – the more traditional form of act utilitarianism – or should agents choose rules to conform to on the basis that such rules are established means of maximizing utility – a position referred to as rule utilitarianism. Moreover, risk/benefit analysis, another variant of consequentialism which is extensively used in healthcare and research contexts, can also assist in evaluating the use of AI in healthcare. As the name suggests, this approach involves weighing the potential benefits of using AI in healthcare against the potential risks or negative consequences. As discussed in chapter 2, in healthcare, the use of AI has various potential benefits, such as improving accuracy in diagnosis, providing personalized treatments, and reducing healthcare costs. However, there are also potential risks associated with the use of AI, such as errors in decision-making, privacy concerns, and bias. Benefits must always be balanced with risks, and both must be communicated to those who stand to be impacted.

Returning to our thought experiment, AI-powered healthcare would provide the appropriate ends or 'state of wellnesses' for all citizens involved. In the context of privacy, transparency, and security, utilitarianism would consider the potential benefits of using AI in healthcare against the potential risks to individuals' privacy and security. This would involve weighing the positive outcomes that AI in healthcare could bring about (such as improved diagnosis and treatment, better healthcare outcomes, and reduced healthcare costs) against the potential harms that could result from breaches of privacy and security (such as compromised medical records, identity theft, and other forms of personal harm). Therefore, the goal would be to maximize the overall well-being of all individuals affected by the technology. This would require considering not just the benefits and harm to patients, but also to healthcare providers, insurers, and society.

As highlighted above, the advantages of AI in healthcare are numerous; however, when we consider monitoring patients' health via AI, there is a need for transparency and security. Act Utilitarian would argue for actions that maximize net well-being or happiness, after weighing this against any disadvantages of the action which would detract from well-being. In the long term, an AI-centric monitoring system would save costs on employees' salaries and would also help healthcare workers avoid redundant and time-consuming tasks. In terms of the former, this would be a positive outcome in terms of financial considerations, but it would be negative outcome for those losing their livelihoods due to being replaced by AI systems. This important point aside, act utilitarianism might argue that for some patients, 24-hour monitoring of diabetes patients, for example, could be the difference between life and death; therefore, the benefits of surveillance outweigh security or privacy concerns. On the other side of the spectrum in LMICs, most healthcare centres have limited resources, including a lack of infrastructure such as Wi-Fi, security to ensure privacy, and adequate storage facilities. So, there is a possibility that these kinds of innovations would further exacerbate inequities in the healthcare sector meaning that AI might not be available in these settings.

One of the most concerning issues about AI that has been raised is the protection of privacy and the potential misuse of unregulated information. (84) Privacy and the misuse of information would not be a concern for utilitarianism unless it detracted from net happiness or well-being. While a utilitarian might justify the importance of privacy in terms of the good outcomes it ensures, it might easily trade this off if it is outweighed by an alternative. This is one of the problems with the utilitarian reduction to happiness; a right to privacy, as a moral good, only makes sense within a Kantian framework. (122) As emphasized in various sections thus far, the doctor-patient relationship is centred around trust. While doctor-patient trust may have eroded significantly, for several reasons, it is still regarded as fundamental. The limitation of utilitarianism is that it pits various goods against one another, many of which are incommensurable. In other words, arguments that are quantitative or problem-solving in nature are pitted against concerns that are qualitative in nature and are difficult to reconcile.

Once again, as was the case with our discussion of the use of deontological considerations in the design of ethical AI, the flaws of Utilitarianism are translated into challenges that would be faced if considering such an approach as the primary one for AI ethical decision making. Utilitarianism as a moral theory struggles as far as it is difficult to quantify well-being or happiness and to balance competing forms.

As mentioned above, rule utilitarianism regards a morally right action as one that complies with moral codes and rules that have been shown to lead to better outcomes and consequences. A rule utilitarian approach to AI in healthcare would focus on establishing rules or principles that, if followed consistently, would maximise net happiness or well-being. In the context of AI in healthcare, a rule utilitarian might argue that ethical principles should be established that guide the development, deployment, and use of AI in healthcare, on utilitarian grounds. These principles might include considerations such as transparency, accountability, and informed consent as far as they are associated with better outcomes for the wellbeing of patients and healthcare professionals. For example, a rule utilitarian might argue that the ethical use of AI in healthcare should be guided by rules such as the principle of transparency, which would require that AI algorithms and decision-making processes be open and understandable to patients and healthcare providers, not on the grounds of autonomy but on utilitarian grounds. A rule utilitarian might also advocate for the principle of accountability, which would require that those responsible for developing and deploying AI in healthcare be held responsible for any negative consequences that result from its use. By establishing these kinds of ethical rules or principles, a rule utilitarian would seek to promote the greatest overall happiness or well-being over the long-term, rather than simply focusing on short-term outcomes in specific cases. By consistently following these ethical rules or principles, a rule utilitarian would argue that we can create a more just and ethical healthcare system that benefits all members of society, and makes everyone better off than they would have been in the absence of such rules and principles.

There are, however, several further challenges, over and above those already mentioned above, that consequentialist AI in healthcare faces. First, to assess the consequences of a decision or action, a consequentialist AI must have a clear definition of what constitutes a desirable or undesirable outcome. In healthcare, this can be challenging, as different patients and stakeholders may have different goals and preferences. Second, consequentialist oriented AI must consider both the immediate consequences of a decision or action and the long-term consequences as well as both direct and indirect consequences. In healthcare, this can be especially challenging, as some interventions may provide short-term benefits but have negative long-term consequences, such as medications that relieve pain but lead to dependency or addiction. Third, consequentialist AI must consider the distribution of benefits and harms across separate groups, which includes ensuring that decisions are made fairly and without bias. In healthcare, this can be particularly challenging, as certain patient populations may be underrepresented in datasets used to train AI models, leading to biased predictions and recommendations. Fourth, consequentialist AI in healthcare must be transparent about the decision-making process and the data used to make decisions. This can be challenging, as AI algorithms can be complex and difficult to understand. In addition, accountability mechanisms must be in place to ensure that those responsible for AI decision-making are held accountable for any negative consequences that result. (98)

Given the above discussion, both deontological and consequentialist AI in healthcare clearly face significant challenges, but addressing these challenges will be crucial for developing ethical and effective AI systems in healthcare. Put differently, the challenge with using ethical theories to design ethical robots, is that all the theories are flawed in one way or another, hence the realisation that one approach cannot be endorsed unilaterally. Given that the theories sometimes provide opposing recommendations for how to behave in similar situations, a level of insight or moral knowledge is needed to discern the more appropriate approach to follow. It seems unlikely that a robot would ever be able to possess such discernment. Moreover, it also remains unclear whether a robot programmed with a particular moral theory would be acceptable to community members whose opinions are not in agreement with the theory in question. (98,100)

4.8 Virtue Ethics

The third approach I will discuss is virtue ethics, the oldest form of ethics to be found in the Western canon. This approach is an ethics of character insofar as it focuses not on actions or consequences but rather on character or moral traits. While this approach has its roots in the work of the classical Greek philosopher Aristotle, it was revived in the 20th century by prominent philosophers such as Elizabeth Anscombe, (127) and Alasdair MacIntyre. (128) Aristotle defined a virtue as a well-balanced middle ground determined by rational principles and prudence. It exists as a mean between two extremes of excess and deficiency, choosing the right measure in both feelings and actions. For instance, the virtue of courage is a balance between cowardice and foolhardiness, while generosity finds the mean between stinginess and extravagance. Similarly, diligence is the balance between recklessness and avoidance of danger, and pride lies between humility and arrogance. (129)

According to Anderson (2020), “implicit ethical agents have a kind of built-in virtue—not built-in by habit but by specific hardware or programming.” (130) In other words, robot virtues could be directly coded into their behaviour. For instance, an autonomous surgery device could initiate a defensive or protective manoeuvre, such as overriding a physician’s judgement, that could prevent a potentially life-threatening error. On the one hand, the potential benefits of having an autonomous system that can identify and intervene to correct errors quickly could be significant, potentially saving lives and improving healthcare outcomes. On the other hand, there are concerns about the impact that such a system could have on the relationship between patients and physicians, and the potential loss of trust in human decision-making that could result.

From a virtue ethics perspective, the ethical question around the use of autonomous surgery equipment would be framed in terms of the virtues and character traits that are necessary for promoting good outcomes in healthcare. Virtue ethics would focus on the moral character and intentions of the physician, and the role that technology can play in supporting or undermining those virtues. A virtue ethics approach would emphasize the importance of cultivating virtues such as wisdom, discernment, and practical judgement in healthcare professionals. From this perspective, the role of the autonomous surgery equipment would be to support these virtues. If the equipment can help the physician make better decisions, avoid errors, and promote the good of the patient, then it could be seen as a valuable tool in promoting the virtues of the physician. However, there are also potential risks associated with the use of an autonomous surgery equipment from a virtue ethics perspective. For example, if the equipment undermines the virtues of the physician by leading them to rely too heavily on technology or to abdicate their responsibility as decision-makers, then it could be seen as a threat to the integrity of the medical profession. (76) In keeping with an ethically comprehensive approach to the issue, the decision of whether to use an autonomous surgery equipment that can override the physician's judgement should be based on a careful assessment of the potential benefits and risks, as well as consideration of ethical principles such as respect for autonomy, beneficence, and non-maleficence.

Continuing our consideration of a virtue ethics informed robot; Aristotle described virtue by asking a fundamental question about the “good for man”, with the latter described as “an activity of the soul in accordance with virtue”. (131) Rachels and Rachels, paraphrasing Aristotle, describe a virtue is “a character trait that is manifested in habitual (i.e., sustained or maintained) action that it is good for a person to have.” (132) Truth telling can be considered as a habitual action that is good for a moral person as compared to withholding the truth. Therefore, it would seem that robots that cause harm (either deliberately or not) by modifying or withholding the truth cannot be considered or appreciated as moral robots. In terms of the discussion above, if Kantianism or utilitarianism are the ethical frameworks used to construct AI robots, they could have trouble solving complex dilemmas outside the ethical frameworks, when their pre-programmed rules are unable to do so. Therefore, the robots might endanger or cause harm in such circumstances. However, for these machines to become moral these systems would first need to be sensitive to both moral approaches and have the capacity to act in a way that is associated with virtue. (58)

The benefits of using virtue ethics insights for AI in healthcare are numerous. First, virtue ethics places a strong emphasis on ethical character and personal responsibility, which can be useful in promoting ethical behaviour in AI systems. (131) By focusing on the development of ethical character and intentions, AI developers and healthcare professionals can work to ensure that the technology is being used for the benefit of patients, thus aligning it with other ethical principles. Second, this framework is known for being flexible and context-sensitive, which can be useful in the complex and rapidly evolving field of AI in healthcare. (133) By considering the specific context and ethical challenges of AI in healthcare, developers and healthcare professionals can work to create AI systems that are context specific and tailored to the unique needs and values of patients and healthcare providers. (133) Third, there is a strong prominence on the good of the person, implied by virtue ethics, which can be useful in ensuring that AI systems in healthcare are being used for the benefit of patients. Fourth, virtue ethics as a framework encourages a broader perspective on ethical issues, considering the wider social and ethical context in which AI in healthcare is being developed and used. By considering the potential social, economic, and political impacts of AI in healthcare, developers and healthcare professionals can work to ensure that the technology is being used in a responsible and ethical manner. (58)

Virtue ethics as a framework is not without problems, however. The approach focuses on virtues, as such, and not their appropriate deployment, which will provide little guidance if the aim is to create relatively uniform ethical standards for AI in healthcare. Put differently, a virtue ethics approach would not help to resolve disagreements and inconsistencies in the ethical principles and practices that are applied to AI systems. This challenge is sourced in a criticism that is frequently directed at virtue ethics itself, namely, that it does not provide specific guidelines for action and can thus be vague in its application. In other words, although, it provides guidance for developing ethical character and intentions, it may not provide clear guidance for making specific decisions in complex situations. This can be challenging for healthcare professionals and AI developers who need to make decisions about the use of AI in healthcare. The absence of clear guidance on addressing bias and discrimination in AI systems, for example, poses a potential risk for biased outcomes and discriminatory practices. This can be a significant challenge in healthcare, where AI systems may be used to make decisions that have a significant impact on the health outcomes of patients. Virtue ethics places a strong emphasis on personal responsibility, however, it does not provide clear mechanisms for holding individuals or organizations accountable for ethical violations related to AI in healthcare. This can make it challenging to ensure that AI systems are being used in an ethical and responsible manner. While virtue ethics can be a useful framework for promoting ethical behaviour in AI systems, it may need to be complemented by other ethical frameworks and practices to address these challenges. (58)

4.9 Conclusion

In this chapter, I have discussed the complexity of developing morally competent or ethical robots, particularly in healthcare contexts. This is clearly a multifaceted task which requires careful consideration of various ethical theories and principles to ensure that robots are responsive to human values and ethical norms. To support this claim, I explored the three main ethical approaches namely: deontology, consequentialism, and virtue ethics. All these approaches offer valuable insights for designing ethical AI systems. Deontology emphasizes the importance of adhering to moral rules and principles, regardless of the consequences. Nevertheless, this approach is flawed because there is a need to provide practical guidance in complex moral dilemmas and a rule based approach is unlikely to assist in cultivating moral wisdom in robots. Consequentialism focuses on the outcomes and consequences of actions, aiming to maximize overall happiness or well-being; however, this approach falls short because it is quite unclear how to quantify well-being and balance competing forms of happiness. Also, this approach needs to consider the distribution of benefits and harms adequately. Virtue ethics emphasizes the development of ethical character and intentions, focusing on virtues and personal responsibility. Unfortunately, it lacks specific guidelines for action and does not provide precise mechanisms for accountability.

Practically, a combination of these ethical approaches may be necessary to address the challenges and limitations of AI – but all these approaches have major limitations themselves. By incorporating the strengths of these main theories, AI systems can be designed to closely mimic humans in a manner that is consistent with ethical and moral standards. To achieve this level of human-like behave we must also take the insights associated with principlism, discussed in Chapter 3, into consideration. Developing morally competent and ethical robots in healthcare requires a comprehensive and integrated approach that considers the values and principles of various ethical theories. By integrating these ethical considerations into the design and programming of AI systems, we can strive to create AI that aligns with human values and promotes the common good. In the next chapter, I shall further explore the ethics of responsibility and phronesis as an ethical framework to consider for AI in healthcare.

Chapter 5

5.1 Overview

In this chapter, I consider an ethical approach that has the potential to transcend some of the limitations of utilitarianism, deontology, and virtue ethics, as discussed in chapter 4, namely, the “ethics of responsibility”. I contend that this approach is well-equipped to address some of the dilemmas and ethical challenges that confront the use of AI in healthcare, in practice. The proposed framework of responsibility serves as a dual construct, encompassing both forward and backward responsibilities. It aims to support developers and healthcare professionals in taking accountability for the outcomes of their decisions. This includes assuming proactive responsibilities to shape their behaviour and prevent future harms, as well as accepting retrospective responsibilities for any harm that has already occurred. On the one hand, the framework emphasizes the understanding that the consequences of their decisions cannot be solely attributed to the guidelines dictated by ethical theories. Instead, it encourages a holistic view of responsibility and decision-making. On the other hand, this approach acknowledges the essential contributions of existing normative ethical approaches in the exercise of moral responsibility.

To consider this potential framework, I firstly discuss and incorporate the general ideas of various influential philosophers whose general ideas are helpful for a deeper understanding of the ethics of responsibility. These thinkers include: Max Weber, (134) Hans Jonas, (80) Emmanuel Levinas, (135,136) Zygmunt Bauman, (135,136) and Judith Butler. (137) I then draw insights from the general ideas discussed that are relevant to the ethical application of AI in healthcare and future recommendations. Thereafter, I consider how Aristotle's notion of phronesis can inform the framework. To deepen my consideration of the ethics of responsibility framework, I incorporate ideas regarding the possibility of failure and an ethics of futurity that have been articulated by various thinkers. (138,139,140) I conclude with some insights developed by Shannon Vallor. (133)

5.2 Historical Development of Ethics of Responsibility

Max Weber was the first to introduce the notion of an "ethics of responsibility" during his renowned speech, "Politics as a Vocation", in 1919. (134) However, the German philosopher Hans Jonas[8]

[8] In 1984, Hans Jonas authored a book titled "*The Imperative of Responsibility: In Search for an Ethics for the Technological Age*." Jonas believed that existing ethical frameworks were inadequate for tackling ethical dilemmas arising from emerging technologies. He pointed out that these frameworks focused too narrowly on the immediate consequences of human actions, making them ill-suited to address the long-term ethical

expanded upon Weber's concept and emphasized the "imperative of responsibility." (80) According to Jonas, this imperative emphasises the importance of considering the future consequences of present actions. (80) By doing so, individuals and societies can take responsibility for their actions and ensure that they do not cause harm or negative consequences to future generations thereby fostering a sustainable and ethically conscious approach to technology and its impact on humanity and the world. The idea of an ethics of responsibility was further developed in different ways by the French phenomenologist Emmanuel Levinas and Polish sociologist Zygmunt Bauman. (135,136)

Levinas' exploration of the ethics of responsibility, emphasizes the primacy of ethical relationships and our responsibility towards one another. (141) Levinas argues that the ethics of responsibility emerges through our interactions with others. According to his perspective, forming ethical relationships with plants and animals is challenging due to their inability to communicate through language or exhibit human-like qualities. He refers to these interactions as "face-to-face" encounters with nonhuman entities. (142) He further posits that the human face carries a unique ethical significance, evoking a call for responsibility and ethical engagement. This encounter disrupts our self-centeredness and demands a response that transcends self-interest. While Levinas's perspective holds value, it may conflict with contemporary thinkers who argue that current ethical priorities should be centred around the treatment of the natural world, including animals, plants, and the environment. (143) It is crucial to acknowledge this criticism and ensure that any framework built on Levinas's ideas sufficiently addresses these broader ethical considerations.

According to Levinas, ethical responsibility is not something that can be chosen or avoided; it is an ethical obligation that precedes any conscious decision. (136) He argues that our responsibility towards the Other is fundamental and unconditional, and it takes precedence over any other ethical considerations or systems. The ethics of responsibility provides a profound perspective on our ethical obligations towards others and the transformative power of ethical encounters. (140) Levinas also criticizes traditional ethical theories for their tendency to prioritize abstract principles or rules, which he believes can lead to a dehumanization of the Other. Instead, he emphasizes the need to focus on the concrete, immediate, and face-to-face encounters with others, where the ethical imperative is revealed. (136) On this view, the ethics of responsibility is not merely about doing good deeds or

implications of advancements like AI, genetic engineering, and the possibility of nuclear disasters. Jonas advocated for a more comprehensive ethical framework that could tackle the complex challenges presented by these technological developments. (Jonas 1984:5–6). Of course, Jonas was writing in the 1980s and these concerns he voices, are even more pertinent in the contemporary context which has seen significant advancements in the areas he identifies.

following rules, but about acknowledging the infinite responsibility we have towards the Other. It involves being receptive to the vulnerability and alterity of the Other and recognizing their absolute and irreducible value. Through this recognition, Levinas argues, we can cultivate a more just and compassionate society. (142)

Turning then to Bauman; the basic tenet of his investigation of an ethics of responsibility revolves around the notion that individuals in modern societies face more complex moral dilemmas and uncertainties than ever before. His work also focuses on the social and cultural dimensions of responsibility in the context of modernity. (135,136) He argues that consumer behaviour patterns impact every aspect of human life resulting in citizens consuming more, thereby becoming commodities themselves on the consumer and labour markets. In his writings he also explains how globalization, consumerism, and the fluid nature of social relations create ethical challenges that demand a responsible approach. (144) The ethics of responsibility according to Bauman involves recognizing the interdependencies and consequences of our actions in an interconnected world. It requires individuals to be mindful of the potential impacts of their decisions on others and the environment. He emphasizes the value of ethical reflection, critical thinking, and empathetic engagement as key components of assuming responsibility in our contemporary world. (145) In this regard, his work contributes to the broader discussions on ethics and responsibility, offering insights into the challenges and possibilities of ethical decision-making in the context of modernity.

Feminist philosopher Judith Butler has also made significant contributions to the ethical discourse. Her concept of "precariousness" or "vulnerability" may be useful for AI healthcare because it emphasizes the risks of errors, biases, and inadequate addressing of individual patient needs. (137) Recognizing and addressing these factors is crucial for ethical and responsible deployment of AI systems in healthcare. In her writing she critically examines power dynamics and social structures that shape individuals' identities and roles. She also highlights how power operates in society, influencing the distribution of responsibility and moral agency. (146) By questioning dominant power structures, Butler challenges traditional notions of responsibility and calls for an ethics that takes systemic inequalities and social justice concerns into account. Butler argues that our inherent vulnerability and interdependence are essential components of our humanity and that recognizing and acknowledging this vulnerability is crucial for ethics and responsibility. (137) She challenges the notion of autonomous and self-sufficient individuals and emphasizes the relational nature of responsibility. (137) In her writing, she advocates for an ethics of care, [9] (58) and solidarity that involves recognizing and responding to the vulnerability of others and challenging oppressive structures. (137)

5.3 Application of these perspectives on the ethics of responsibility

While Levinas, (136) Bauman, (135,136) Jonas, (80) and Butler (137) all address the ethics of responsibility, they do so from distinct perspectives. Levinas stresses the significance of engaging in a personal interaction and the value of establishing an ethical connection with others. He highlights the primacy of face-to-face encounters, the ethical demand they impose, and the transcendence of self-interest. (136) Bauman strongly emphasizes recognising and fulfilling our moral obligations in a rapidly changing and interconnected world. He unambiguously advocates for ethical behaviour that promotes social cohesion and sustainability, with a firm emphasis on acknowledging the consequential impact of our actions. (135,136) Jonas directs attention to environmental and technological considerations, (80) whereas Butler focuses on power, social structures, and the politics of identity. Each philosopher offers unique insights into the ethical dimensions of responsibility, but their approaches differ in terms of the emphasis placed on power dynamics, interpersonal encounters, and environmental concerns. (137)

Applying Levinas' perspectives to AI in healthcare entails recognizing and prioritizing the patient as the other. It involves designing AI systems that respect and respond to individual needs and the vulnerabilities of patients. The focus should be on developing AI technologies that enhance human connection and personalized care. Furthermore, the ethical responsibility lies in ensuring that AI supports healthcare practitioners rather than replacing them. (136) On the other hand, Bauman stresses the importance of acknowledging the correlation between actions and consequences, while also considering social and cultural considerations. By thoroughly examining ethical considerations and assessing the impact of AI systems on patients, stakeholders can prioritize human well-being and compassion. This conscientious approach ensures strict adherence to ethical standards and guarantees favourable outcomes for patients. However, Jonas's environmental ethics provides critical insights on our environmental responsibility, and how this responsibility can be applied to AI in healthcare by considering the ecological impact of AI systems and data collection. (137) This includes implementing AI technologies that prioritize energy efficiency, sustainability, and minimize their carbon footprint. Furthermore, responsible data governance practices should be in place to protect

[9] Ethics of care is an ethical framework that emphasizes the importance of relationships, empathy, and compassion in moral decision-making. It places value on interconnectedness, considering the needs and interests of individuals and recognizing the importance of caring for others. The ethics of care framework also emphasizes the moral significance of promoting well-being, and addressing social inequalities and injustices.

patient privacy and prevent unauthorized access or misuse of sensitive medical information. Butler's approach involves explicitly acknowledging and seeking to understand the power dynamics and social structures embedded in AI systems. This will ensure that the design and implementation of AI technologies promote equity, fairness, and inclusivity in healthcare delivery. Moreover, this will require proactively addressing biases in AI algorithms, ensuring transparency in decision-making processes, and fostering collaboration between healthcare professionals and AI systems as crucial steps towards promoting fairness. Additionally, Butler's emphasis on vulnerability calls for AI technologies that prioritize patient well-being, respect patient autonomy, and protect privacy rights. (137)

For our purposes, if we are to compare the views of these thinkers in terms of the points that are salient for the framework of responsibility, what we can take from the ideas of the abovementioned thinkers is when determining what is morally right and wrong. It is crucial to recognize that social phenomena can have positive and negative effects. Therefore, we cannot rely on a single perspective or paradigm when making decisions. It is incumbent upon individuals to take responsibility for their actions and carefully consider all potential outcomes before making choices. It is important to consider that individuals have varying values, which can lead to potential conflicts. Therefore, it is crucial to acknowledge and address these differences to move forward towards a solution that works for everyone involved. This level of responsibility is essential to ensure that our actions align with our values and contribute to the betterment of society as a whole. (139) Recognizing the strengths and weaknesses of different ethical approaches underscores the significance of adopting a nuanced, context-sensitive, and comprehensive approach to the concept of responsibility. This comprehensive approach requires everyone, not just healthcare workers and ethicists, to take responsibility for science and technology's impact on our world. In morally complex situations, simply following external rules and laws is not enough to justify our actions, according to current ethical standards. It is our responsibility to take ownership of everything we invent, design, and implement, as we construct, apply, and evaluate our personal values. By doing so, we can ensure that we create a world that is just, equitable, and in line with our collective values and aspirations. (108,138,139)

In terms of an ethics of responsibility and AI in healthcare, as discussed in Chapter 2, AI has the potential to transform healthcare, however, as with any innovative technology, and as discussed in Chapters 3 and 4, ethical considerations must be considered. This is particularly important given AI's potential and significant impact on patients' lives. Just as Hans Jonas emphasised the imperative of considering the future consequences of present actions, so too must developers and practitioners of AI in healthcare consider the potential long-term impacts of their decisions. (80)

The main difference between the ethics of responsibility and other ethical frameworks is that this framework specifically emphasises the moral obligation and accountability of individuals for their actions and their consequences.(134) Therefore, stakeholders have a duty to consider the potential effects and consequences of their actions on other and the broader society. The ethics of responsibility framework accentuates the role of developers' integrity, patient-centered care, accountability, duty, transparency, continuous improvements, and the recognition of one's responsibility towards others. (140)

5.4 Phronesis and Ethics of Responsibility

In this section, I discuss Aristotle's concept of phronesis to consider how it might further assist in developing a framework of responsibility. Phronesis, which is derived from the Greek word for practical wisdom, is a philosophical concept that refers to the ability to consider general moral insights, values, practical consequences, and the wider impacts of actions in terms of the broader context. (147) Put differently, phronesis, which is a kind of art or skill, implies relating and balancing practicality [judgement] with general insights [theory]. (147) This practical wisdom is acquired through experience. In particular its acquisition includes the practice of a form of cognitive reasoning that involves reflecting on one's actions and asking: "What am I doing?" – and … considering how [one's actions] impact the world and those around us … . Additionally, one should ask "Can I do this better?" while keeping in mind the well-being of both the world and others. (148) Practical wisdom embraces "peculiar interlacing of being and knowledge, determination through one's own becoming." (149)

Aristotle referred to phronesis in his Nicomachean Ethics and defined it as the moral ability to engage in critical self-reflection and reasoning, seek truth, make sound decisions, and be aware of the outcomes of our actions. (147) A wise individual can rely on their moral understanding to make decisions in different circumstances. This understanding is influenced by internal guidelines derived from religious beliefs, education, personal conscience, and societal customs. (108,138,139)

The integration of ethical considerations and context-specific judgments into the development, implementation, and use of AI systems in healthcare is a complex process that requires this kind of practical wisdom. Practical wisdom also implies recognising the complexity of healthcare contexts, including the unique needs, values, and preferences of patients and healthcare providers, as well as the socio-cultural, legal, and regulatory aspects that impact the ethical implications of AI applications. Engaging diverse stakeholders, such as healthcare professionals, patients, ethicists, and policymakers, fosters inclusivity and shared decision-making and supports a more comprehensive comprehension of the ethical implications of AI applications in healthcare. (7,56) As mentioned in chapter 4, transparency and explainability of AI algorithms and systems are critical for building trust, accountability, and addressing biases or errors. Applying the notion of phronesis to these ends will entail ensuring that as AI systems are continuously learning and adapting, they are continually evaluated for their impact and effectiveness in the context of wider regulations and ethical requirements. (7,56)

By taking a multidisciplinary and integrated approach, healthcare organizations can address ethical concerns while maximizing the benefits of AI in healthcare and prioritizing patient welfare. The application of phronesis in the development, implementation, and use of AI systems in healthcare serves to enhance patient outcomes, improve healthcare delivery, and foster responsible and ethical use of technology in healthcare. (147) This is essential for developing effective AI regulations that account for the complex ethical and social considerations of AI in healthcare.

In making decisions related to AI in healthcare, policymakers and regulators must consider the principles of Phronesis as part of an Ethics of Responsibility. This means carefully assessing the benefits and risks of AI technologies and ensuring they align with societal values. On the one hand, the Phronetic approach highlights the significance of patient-centered outcomes, emphasizing the importance of respecting and promoting the well-being and dignity of individuals. (139,150) On the other hand, the Ethics of Responsibility, emphasizes accountability and responsible decision-making. (45,140) This framework is particularly relevant when it comes to regulating AI in healthcare, as it ensures that developers, and users of AI technologies are held accountable for their actions.

This framework can therefore be used by regulators to create guidelines and standards that encourage ethical and responsible development, deployment, and use of AI technologies. In developing regulatory frameworks for AI in healthcare, it is essential to consider the practical realities and context of the healthcare sector. This includes the complexity of medical decisions, the significance of human expertise, and the potential impact on patient-provider relationships. The direct application of phronesis in the context of AI is to guide the development of guidelines and regulatory frameworks in way that is nuanced. The World Health Organisation guidelines and frameworks should be adaptable and responsive to the rapidly evolving landscape of AI in healthcare. (151) This requires ongoing monitoring and evaluation of AI systems and the development of mechanisms for addressing new ethical and social challenges as they emerge. By taking a Phronetic approach, regulators can develop nuanced policies that address these unique considerations and challenges, contributing to the safe and beneficial integration of AI in healthcare. The integration of AI technology in healthcare requires a careful and ethical approach. I content that a framework that includes considerations informed by the insights of Phronesis and the Ethics of Responsibility can provide useful guidance for policymakers and regulators in navigating the complex issues surrounding AI in healthcare. (139)

Another helpful aspect of Aristotelian thought is the notion of teleology. Aristotle posited that all objects, including human beings, have what he called a final cause or purpose. He claimed that "...for all things that have a function or activity, the good and the 'well' is thought to reside in the function" (147), which means that in achieving its purpose or goal, an object achieves its good. For example, the telos or ultimate purpose of a car is to transport, and a car is thus judged in terms of the extent to which it supports one's daily requirement of getting from one location to another without mechanical faults or breakdown. (147) In comparison to a car, the telos or ultimate purpose of AI in healthcare is a much more complicated and contested matter. Some argue that the justification for AI in healthcare is that it facilitates more accurate diagnoses and personalised treatment regimens and improves clinical workflows. (136) If the general purpose of healthcare is human health and well-being, a matter which is also subject to contestation, then the telos of AI in healthcare might be improved patient outcomes and well-being. (16) In particular, the justification could be improvements that might not be possible without the presence of AI in healthcare. On the other hand, some argue that the justification for incorporating AI in healthcare might be a matter of cost effectiveness or purely profit-focused which could result in it being inimical to patient autonomy, and thus, ultimately, to well-being. (58) The nature of practical reasoning employed to ethical ends is such that it will entail acute awareness of such deeper goals, and their implications. It will also include constantly prioritising regulations that promote the ethical use of AI in healthcare and to ensure that AI is developed and equitably deployed to align with the values and goals of patients and healthcare providers. (147)

Recognizing that some medical decisions may not alleviate suffering or improve quality of life, even if this is not the intention, highlights the need for phronesis in medical decision-making. While it is important to acknowledge that phronesis is fallible, it can protect and guide us when scientific knowledge falls short. (108,138,139) It is crucial to acknowledge the boundaries of scientific knowledge when it comes to AI and place greater emphasis on cultivating practical wisdom to navigate uncertain decision-making scenarios. (108,138,139) In the absence of certainty, an Ethics of

Responsibility that is enriched by the cultivation of phronesis reminds us that we are nevertheless still accountable for our actions. (108,138,139)

5.5 The possibility of failure

As discussed in previous sections, accountability requires us to carefully consider the reasons behind our decisions and actions and be prepared to explain the moral basis for them. It also means being open to constructive criticism when our reasoning is flawed or when we receive new information. In this context, the ethics of responsibility recognizes the potential for mistakes and failures. As such, a moral agent must be willing to take responsibility for their decisions, even if it means accepting blame or facing consequences. (139)

As mentioned above, decisions must frequently be made in the absence of certainty, and this is also the case for moral decisions. (108) Unlike certain more factual or concrete domains within the biomedical sciences, ethical questions cannot be answered with absolute certainty but instead rely on a "logic of validation" based on probability. (152) Therefore, a crucial component of the ethics of responsibility is the acknowledgement of fallibility; when making moral decisions we may make mistakes that have serious consequences. (108) However, inaction can also have equally harmful results. We must accept responsibility for our decisions, even though we cannot guarantee certainty. Instead, we should thoroughly explain our reasoning and the arguments behind our decision. (108) In this way, an ethics of responsibility does not guarantee perfect moral behaviour, but adherence to such a framework is more likely to ensure responsible, in the sense of considered, moral behaviour, which is a more realistic expectation for moral agents. (139) Recognizing the fallibility of human decision-making also entails recognising the value of learning from past mistakes that can contribute to the development of a more improved society. (108,139)

5.6 An Ethics of Responsibility as an ethics of futurity

As has been discussed throughout this thesis, the introduction of AI technology in healthcare has, and will continue to, broaden the range of ethical concerns to include more than just current human needs and daily interests. Responsibility requires that we also consider the impact of technologies developed, and in development, on "future generations”. (80) However, some argue that the uncertainty of the future makes it difficult to base moral decisions on predictions about what might happen. (108,138,139)

5.7 The future of AI in healthcare

In the face of uncertainty about the future, Shannon Vallor's updating of virtue ethics may be helpful. (133) Vallor expands upon traditional virtue ethics, reformulating phronesis as “technomoral wisdom” and the virtues as “technomoral virtues,” to address unique challenges and ethical considerations posed by emerging and disruptive technologies and innovations, such as AI. (133) Vallor’s reformulation is likely to assist healthcare practitioners and stakeholders to go beyond narrow utilitarian and Kantianism considerations and consider the broader human values and societal impact of technological advancements. Technomoral wisdom provides a means to integrate ethics into the design, deployment, and governance of these technologies, promoting a future that aligns with ethical principles and virtues. (133) It involves developing the intellectual, moral, and emotional qualities necessary to make rigorous ethical decisions and navigate complex technological landscapes. (133) Similarly, to the original Aristotelian notion, technomoral wisdom goes beyond mere technical expertise and encourages practitioners to consider the broader social, environmental, and ethical implications of their work.

Vallor emphasizes the importance of virtues such as honesty, humility, empathy, care, and justice in the development and use of technology. (133) Technomoral wisdom involves applying these virtues to anticipate and address the potential ethical challenges and unintended consequences of emerging technologies. It encourages stakeholders, developers, and practitioners to engage in ongoing ethical reflection, dialogue, and critique to ensure responsible and beneficial outcomes. (133) It also emphasizes the importance of interdisciplinary collaboration, seeking diverse perspectives, and engaging with stakeholders to ensure that technological innovations align with ethical principles and societal values. By fostering a culture of ethical inquiry, technomoral wisdom promotes continuous learning, improvement, and accountability in the realm of technology. (133) Technomoral wisdom also emphasizes a human-centred approach to patient care. Practitioners should prioritize the well-being and dignity of patients, understand their unique needs and values, and ensure that technological interventions enhance rather than harm the patient-provider relationship. Finally, it is essential to develop the necessary skills and knowledge to use healthcare technologies effectively and ethically. Technomoral wisdom encourages ongoing education and training to enable healthcare professionals to navigate and leverage technological advancements. (133)

5.8 Potential recommendations

Much of the discussion in this chapter has been at a more theoretical level with various suggestions and general ethical recommendations drawn from some of these theoretical observations. However, when considering the deployment of AI in healthcare, integrating both Vallor's notion of techomoral wisdom with the various insights associated with the ethics of responsibility can provide some practical guidance and more concrete recommendations. These recommendations aim to ensure that AI in healthcare is deployed in a manner that is ethically sound, respectful of patient rights, and advances equitable and high-quality healthcare outcomes. I propose the key recommendations derived from the discussion in this chapter:

1. When designing AI systems for healthcare, patient well-being and human interactions should be given top priority. This priority can only be achieved by strongly emphasising virtues such as empathy, compassion, and accountability. Furthermore, by leveraging technical expertise and virtuous design principles, we can create AI systems that enhance the quality of care and respect patient autonomy. In short, developing AI systems for healthcare requires an unwavering focus on patient well-being and a steadfast commitment to virtuous design principles.

2. Designers must prioritize transparency and explainability (as much as possible) in AI systems to promote trust among healthcare providers, patients, and stakeholders. This requires detailed documentation of algorithms, data sources, and decision-making processes to ensure understanding and accountability. By embracing the ethics of responsibility, designers can guarantee that all parties involved clearly comprehend how the AI system operates.

3. Effective and responsible AI deployment in healthcare necessitates ongoing ethical reflection and dialogue among healthcare professionals, ethicists, policymakers, patients, and AI developers. Through collaboration, these interdisciplinary stakeholders can efficiently tackle emerging issues and establish unambiguous guidelines for navigating ethical challenges. Technomoral wisdom must be leveraged to promote and foster this collaboration, ensuring responsible AI deployment is achieved.

4. Data Privacy and Security must be paramount when developing AI in healthcare. This requires robust data protection measures to safeguard patient information and ensure compliance with relevant regulations as well as ethical data governance practices that respect patient autonomy and informed consent.

5. In the South African context, there is a need to address the pitfalls of POPIA such as data minimisation and data security; as the act does not specifically address the challenges posed

by AI in healthcare. Due to the numerous disparities in terms of data protection legislation between South Africa and other countries, there is a possible need to review the existing laws or enact supplementary ones that take into consideration addressing these specific concerns as well developing a new regulatory framework for data protection in AI medicine.

6. Continuous monitoring and evaluation of the ethical implications of using AI in healthcare is crucial. The ethics of responsibility framework emphasises this point. Establishing mechanisms that assess the impact of AI systems on patient care, equity, and social justice is therefore necessary. In addition, regularly reviewing and updating ethical guidelines and policies is also crucial, to keep up with technological advancements.

7. Ethical education and training is key to fostering technomoral wisdom and an ethics of responsibility. This should include a specific focus on learning about the ethical implications of AI in healthcare to equip healthcare professionals and AI developers with the knowledge and skills to navigate these ethical challenges, make informed decisions, and apply appropriate ethical frameworks effectively.

5.9 Conclusion

In this chapter, I discussed the ethics of responsibility, and used Aristotle's concept of phronesis to navigate complex moral situations in order to ascertain the best ethical course of action for AI in Healthcare. In the context of AI in healthcare, I suggested that the ethics of responsibility can play a crucial role in guiding both manufacturers and individuals to make informed decisions. The advantage of this framework is its dual nature, in the sense that it is both forward and backward looking and open to the possibility of failure. The proposed framework, in tandem with the notion of phronesis, uses the insights gained from a deep understanding of the ethical implications of AI technologies to evaluate potential risks and benefits, to make an informed ethical decision. Phronesis also recognises the risk posed by biases, data security, privacy and promotes transparency and accountability. By leveraging phronesis, stakeholders can establish a healthcare system that serves everyone, and in particular, prioritizes patients' well-being.

5.10 Concluding remarks.

In conclusion, implementing AI technology in healthcare systems in Africa has immense potential due to the high burden of disease, shortages of healthcare professionals, and limited resources. Despite the consequences and risks posed by AI within the healthcare context, the importance and advantages cannot be underscored.

This thesis stresses the importance of deep ethical consideration in AI development to fully harness the potential of these technologies and to ultimately create a more efficient healthcare system. I approached the ethical consideration of AI in healthcare using moral norms, principles, ethical theories and certain moral frameworks to address unforeseen implications and comprehensively analyse AI in healthcare.

In keeping with the principlist framework my discussion centred around the fundamental ethical principles of respect for autonomy, non-maleficence, beneficence, and justice, with an emphasis on the importance of trust in AI systems for their effective deployment. I then critically discussed the development of morally competent or "ethical" robots and explored different ethical frameworks, such as deontological considerations, consequentialism, and virtue ethics. In my discussion, I highlighted the strengths and limitations of these frameworks and explored their fitness for purpose within the context of AI in healthcare.

Ethical risks are associated with various areas of AI, such as deep learning and machine learning. I provided a historical development of the ethics of responsibility and explored the concept of phronesis in relation to ethical responsibility. Thereafter, I suggested ethics of responsibility as a proposed framework because this framework provides us with both a forward and backward-looking approach to address contextually-associated risks whilst improving the acceptability of AI within the healthcare context.

This thesis aimed to provide a comprehensive understanding of the ethical implications of AI in healthcare and offered a framework for ethical decision-making in this context. By examining various ethical principles and frameworks, this thesis also aimed to contribute to the ongoing discussion on how to ensure the responsible and ethical use of AI in healthcare.

References

1. Mosch L, Fürstenau D, Brandt J, Wagnitz J, Klopfenstein SA, Poncette A-S, et al. The medical profession transformed by artificial intelligence: Qualitative study. Digit Heal [Internet]. 2022 Jan 13;8:205520762211439. Available from: http://journals.sagepub.com/doi/10.1177/20552076221143903
2. Shmatko A, Ghaffari Laleh N, Gerstung M, Kather JN. Artificial intelligence in histopathology: enhancing cancer research and clinical oncology. Nat Cancer [Internet]. 2022 Sep 22;3(9):1026–38. Available from: https://www.nature.com/articles/s43018-022-00436-4
3. Lee D, Yoon SN. Application of Artificial Intelligence-Based Technologies in the Healthcare Industry: Opportunities and Challenges. Int J Environ Res Public Health [Internet]. 2021 Jan 1;18(1):271. Available from: https://www.mdpi.com/1660-4601/18/1/271
4. World Health Organization. Ethics and Governance of Artificial Intelligence for Health: WHO guidance. World Health Organization. 2021. 1–148 p.
5. WHO. Ethics and Governance of Artificial Intelligence for Health: WHO guidance [Internet]. World Health Organization. 2021. Available from: http://apps.who.int/bookorders.
6. Rajaram, Pandey AK, Saproo S, Bansal S. Novel 4A Framework for Measuring Outcomes of Technological Implementation in The Healthcare Industry. Int J Life Sci Pharma Res [Internet]. 2023 Mar 1; Available from: https://www.ijlpr.com/index.php/journal/article/view/1615
7. Sallstrom L, And OM, Mehta H. Artificial Intelligence in Africa's Healthcare: Ethical Considerations. 2019;(312):12. Available from: https://www.orfonline.org/research/artificial-intelligence-in-africas-healthcare-ethical-considerations-55232/%0Ahttps://www.orfonline.org/wp-content/uploads/2019/09/ORF_Issue_Brief_312_AI-Health-Africa.pdf
8. Richardson JP, Smith C, Curtis S, Watson S, Zhu X, Barry B, et al. Patient apprehensions about the use of artificial intelligence in healthcare. npj Digit Med [Internet]. 2021 Sep 21;4(1):140. Available from: https://www.nature.com/articles/s41746-021-00509-1
9. Obasa AE, Palk AC. Responsible application of artificial intelligence in health care. S Afr J Sci [Internet]. 2023 Jun 5;119(5/6). Available from: https://sajs.co.za/article/view/14889
10. Jha S, Topol EJ. Adapting to Artificial Intelligence. JAMA [Internet]. 2016 Dec 13;316(22):2353. Available from: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2016.17438
11. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med [Internet]. 2019 Jan 7;25(1):44–56. Available from:

http://www.nature.com/articles/s41591-018-0300-7

12. Korteling JE (Hans)., van de Boer-Visschedijk GC, Blankendaal RAM, Boonekamp RC, Eikelboom AR. Human- versus Artificial Intelligence. Front Artif Intell [Internet]. 2021 Mar 25;4. Available from: https://www.frontiersin.org/articles/10.3389/frai.2021.622364/full
13. Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. npj Digit Med [Internet]. 2020 Sep 11;3(1):118. Available from: https://www.nature.com/articles/s41746-020-00324-0
14. Haenlein M, Kaplan A. A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. Calif Manage Rev [Internet]. 2019 Aug 17;61(4):5–14. Available from: http://journals.sagepub.com/doi/10.1177/0008125619864925
15. Mackworth DLPAK. Artificial Intelligence [Internet]. Vol. 7, Syria Studies. 2010. 37–72 p. Available from: https://www.cambridge.org/highereducation/books/artificial-intelligence/F90A3CDD1D34B6FF3B6235B1B3D9F0C1#overview
16. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Futur Healthc J [Internet]. 2019 Jun;6(2):94–8. Available from: https://www.rcpjournals.org/lookup/doi/10.7861/futurehosp.6-2-94
17. Svoboda E. Artificial intelligence is improving the detection of lung cancer. Nature [Internet]. 2020 Nov 19;587(7834):S20–2. Available from: http://www.nature.com/articles/d41586-020-03157-9
18. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, et al. Artificial intelligence in cancer imaging: Clinical challenges and applications. CA Cancer J Clin [Internet]. 2019 Feb 5;caac.21552. Available from: https://onlinelibrary.wiley.com/doi/abs/10.3322/caac.21552
19. Xiong Y, Ba X, Hou A, Zhang K, Chen L, Li T. Automatic detection of mycobacterium tuberculosis using artificial intelligence. J Thorac Dis [Internet]. 2018 Mar;10(3):1936–40. Available from: http://jtd.amegroups.com/article/view/19696/15545
20. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. Intensive Care Med. 1996;22(7):707–10.
21. Shickel B, Loftus TJ, Adhikari L, Ozrazgat-Baslanti T, Bihorac A, Rashidi P. DeepSOFA: A Continuous Acuity Score for Critically Ill Patients using Clinically Interpretable Deep Learning. Sci Rep [Internet]. 2019 Dec 12;9(1):1879. Available from: http://www.nature.com/articles/s41598-019-38491-0
22. Loh E. Medicine and the rise of the robots: a qualitative review of recent advances of artificial intelligence in health. BMJ Lead [Internet]. 2018 Jun;2(2):59–63. Available from:

https://bmjleader.bmj.com/lookup/doi/10.1136/leader-2018-000071

23. Gamble A. Artificial intelligence and mobile apps for mental healthcare: a social informatics perspective. Aslib J Inf Manag [Internet]. 2020 Jun 2;72(4):509–23. Available from: https://www.emerald.com/insight/content/doi/10.1108/AJIM-11-2019-0316/full/html

24. Report of the Secretary-General on SDG progress. Special edition. New York City (NY): United Nations [Internet]. 2019. Available from: https://sustainabledevelopment.un.org/content/documents/24978Report_of_the_SG_on_SDG_ Progress_2019.pdf

25. Agrebi S, Larbi A. Use of artificial intelligence in infectious diseases. In: Artificial Intelligence in Precision Health [Internet]. Elsevier; 2020. p. 415–38. Available from: https://linkinghub.elsevier.com/retrieve/pii/B9780128171332000185

26. Amisha, Malik P, Pathania M, Rathaur V. Overview of artificial intelligence in medicine. J Fam Med Prim Care [Internet]. 2019;8(7):2328. Available from: https://journals.lww.com/10.4103/jfmpc.jfmpc_440_19

27. Hoodbhoy Z, Hasan B, Siddiqui K. Does artificial intelligence have any role in healthcare in low resource settings? J Med Artif Intell [Internet]. 2019 Jul;2:13–13. Available from: http://jmai.amegroups.com/article/view/5049/html

28. Wahl B, Cossy-Gantner A, Germann S, Schwalbe NR. Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? BMJ Glob Heal [Internet]. 2018 Aug 29;3(4):e000798. Available from: https://gh.bmj.com/lookup/doi/10.1136/bmjgh-2018-000798

29. Pillai S V., Kumar RS. The role of data-driven artificial intelligence on COVID-19 disease management in public sphere: a review. DECISION [Internet]. 2021 Dec 30;48(4):375–89. Available from: https://link.springer.com/10.1007/s40622-021-00289-3

30. Gerke S, Minssen T, Cohen G. Ethical and legal challenges of artificial intelligence-driven healthcare. Artif Intell Healthc [Internet]. 2020;295–336. Available from: https://linkinghub.elsevier.com/retrieve/pii/B9780128184387000125

31. Panesar S, Cagle Y, Chander D, Morey J, Fernandez-Miranda J, Kliot M. Artificial Intelligence and the Future of Surgical Robotics. Ann Surg [Internet]. 2019 Aug;270(2):223–6. Available from: https://journals.lww.com/00000658-201908000-00007

32. Aruni G, Amit G, Dasgupta P. New surgical robots on the horizon and the potential role of artificial intelligence. Investig Clin Urol [Internet]. 2018;59(4):221. Available from: https://icurology.org/DOIx.php?id=10.4111/icu.2018.59.4.221

33. St Mart J-P, Goh EL, Shah Z. Robotics in total hip arthroplasty: a review of the evolution,

application and evidence base. EFORT Open Rev [Internet]. 2020 Dec;5(12):866–73. Available from: https://eor.bioscientifica.com/view/journals/eor/5/12/2058-5241.5.200037.xml

34. Kayani B, Konan S, Ayuob A, Ayyad S, Haddad FS. The current role of robotics in total hip arthroplasty. EFORT Open Rev [Internet]. 2019 Nov;4(11):618–25. Available from: https://eor.bioscientifica.com/view/journals/eor/4/11/2058-5241.4.180088.xml

35. Crew B. Worth the cost? A closer look at the da Vinci robot's impact on prostate cancer surgery. Nature [Internet]. 2020 Apr 23;580(7804):S5–7. Available from: http://www.nature.com/articles/d41586-020-01037-w

36. Kirchner EA, Bütefür J. Towards Bidirectional and Coadaptive Robotic Exoskeletons for Neuromotor Rehabilitation and Assisted Daily Living: a Review. Curr Robot Reports [Internet]. 2022 Jun 19;3(2):21–32. Available from: https://link.springer.com/10.1007/s43154-022-00076-7

37. Vélez-Guerrero MA, Callejas-Cuervo M, Mazzoleni S. Artificial Intelligence-Based Wearable Robotic Exoskeletons for Upper Limb Rehabilitation: A Review. Sensors [Internet]. 2021 Mar 18;21(6):2146. Available from: https://www.mdpi.com/1424-8220/21/6/2146

38. Ciuti G, Caliò R, Camboni D, Neri L, Bianchi F, Arezzo A, et al. Frontiers of robotic endoscopic capsules: a review. J Micro-Bio Robot [Internet]. 2016 Jun 2;11(1–4):1–18. Available from: https://link.springer.com/10.1007/s12213-016-0087-x

39. Looi M-K. Sixty seconds on . . . ChatGPT. BMJ [Internet]. 2023 Jan 26;p205. Available from: https://www.bmj.com/lookup/doi/10.1136/bmj.p205

40. Schönberger D. Artificial intelligence in healthcare: a critical analysis of the legal and ethical implications. Int J Law Inf Technol [Internet]. 2019 Jun 1;27(2):171–203. Available from: https://academic.oup.com/ijlit/article/27/2/171/5485669

41. Schwartz RC. Racial disparities in psychotic disorder diagnosis: A review of empirical literature. World J Psychiatry [Internet]. 2014;4(4):133. Available from: http://www.wjgnet.com/2220-3206/full/v4/i4/133.htm

42. Braveman P. HEALTH DISPARITIES AND HEALTH EQUITY: Concepts and Measurement. Annu Rev Public Health [Internet]. 2006 Apr 1;27(1):167–94. Available from: https://www.annualreviews.org/doi/10.1146/annurev.publhealth.27.021405.102103

43. Neighbors HW, Jackson JS, Campbell L, Williams D. The influence of racial factors on psychiatric diagnosis: A review and suggestions for research. Community Ment Health J [Internet]. 1989;25(4):301–11. Available from: http://link.springer.com/10.1007/BF00755677

44. Hruska E, Liu F. Machine learning: An overview. In: Quantum Chemistry in the Age of Machine Learning [Internet]. Elsevier; 2023. p. 135–51. Available from:

https://linkinghub.elsevier.com/retrieve/pii/B978032390049200024X

45. Matthias A. The responsibility gap: Ascribing responsibility for the actions of learning automata. Ethics Inf Technol [Internet]. 2004;6(3):175–83. Available from: http://link.springer.com/10.1007/s10676-004-3422-1

46. Sarker IH. Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. SN Comput Sci [Internet]. 2021 Nov 18;2(6):420. Available from: https://link.springer.com/10.1007/s42979-021-00815-1

47. P. S. DV. How can we manage biases in artificial intelligence systems – A systematic literature review. Int J Inf Manag Data Insights [Internet]. 2023 Apr;3(1):100165. Available from: https://linkinghub.elsevier.com/retrieve/pii/S2667096823000125

48. Lauriola I, Lavelli A, Aiolli F. An introduction to Deep Learning in Natural Language Processing: Models, techniques, and tools. Neurocomputing [Internet]. 2022 Jan;470:443–56. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0925231221010997

49. Tanya de Villiers-Botha OPINION | Tanya de Villiers-Botha: Risks and limitations of ChatGPT and Bing Chat accreditation.

50. Drage E, Mackereth K. Does AI Debias Recruitment? Race, Gender, and AI's "Eradication of Difference." Philos Technol [Internet]. 2022 Dec 10;35(4):89. Available from: https://link.springer.com/10.1007/s13347-022-00543-1

51. Bohr A, Memarzadeh K. The rise of artificial intelligence in healthcare applications. In: Artificial Intelligence in Healthcare [Internet]. Elsevier; 2020. p. 25–60. Available from: https://linkinghub.elsevier.com/retrieve/pii/B9780128184387000022

52. Lalmi F, Adala L. Big Data for Healthcare: Opportunities and Challenges. In 2021. p. 217–29. Available from: http://link.springer.com/10.1007/978-3-030-62796-6_12

53. Xafis V, Schaefer GO, Labude MK, Brassington I, Ballantyne A, Lim HY, et al. An Ethics Framework for Big Data in Health and Research. Asian Bioeth Rev. 2019;11(3):227–54.

54. Li F, Ruijs N, Lu Y. Ethics &amp; AI: A Systematic Review on Ethical Concerns and Related Strategies for Designing with AI in Healthcare. AI [Internet]. 2022 Dec 31;4(1):28–53. Available from: https://www.mdpi.com/2673-2688/4/1/3

55. Ryan M, Stahl BC. Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications. J Information, Commun Ethics Soc [Internet]. 2021 Mar 3;19(1):61–86. Available from: https://www.emerald.com/insight/content/doi/10.1108/JICES-12-2019-0138/full/html

56. Nuffield Council on Bioethics. Artificial intelligence (AI) in healthcare and research. Bioeth Brief Note. 2018;1–8.

57. Pflanzer M, Traylor Z, Lyons JB, Dubljević V, Nam CS. Ethics in human–AI teaming: principles and perspectives. AI Ethics [Internet]. 2022 Sep 20; Available from: https://link.springer.com/10.1007/s43681-022-00214-z

58. Beauchamp TL, Childress JF. The capacity for autonomous choice. Principles of Biomedical Ethics. 1994. 120–142 p.

59. Protection of Personal Information Act (POPI Act). Relationships among Multiple Task Asymmetries. 2020.

60. Laitinen A, Sahlgren O. AI Systems and Respect for Human Autonomy. Front Artif Intell [Internet]. 2021 Oct 26;4. Available from: https://www.frontiersin.org/articles/10.3389/frai.2021.705164/full

61. Kim TW, Hooker J, Donaldson T. Taking Principles Seriously: A Hybrid Approach to Value Alignment in Artificial Intelligence. J Artif Intell Res [Internet]. 2021 Feb 28;70:871–90. Available from: https://www.jair.org/index.php/jair/article/view/12481

62. A.M. McLean S. Autonomy, Consent and the Law [Internet]. Routledge; 2009. Available from: https://www.taylorfrancis.com/books/9781135219055

63. McCradden MD, Joshi S, Anderson JA, Mazwi M, Goldenberg A, Zlotnik Shaul R. Patient safety and quality improvement: Ethical principles for a regulatory approach to bias in healthcare machine learning. J Am Med Informatics Assoc [Internet]. 2020 Dec 9;27(12):2024–7. Available from: https://academic.oup.com/jamia/article/27/12/2024/5862600

64. Grote T, Berens P. On the ethics of algorithmic decision-making in healthcare. J Med Ethics [Internet]. 2020 Mar;46(3):205–11. Available from: https://jme.bmj.com/lookup/doi/10.1136/medethics-2019-105586

65. Price WN, Cohen IG. Privacy in the age of medical big data. Nat Med [Internet]. 2019 Jan 7;25(1):37–43. Available from: http://www.nature.com/articles/s41591-018-0272-7

66. Santoni de Sio F, van den Hoven J. Meaningful Human Control over Autonomous Systems: A Philosophical Account. Front Robot AI [Internet]. 2018 Feb 28;5. Available from: http://journal.frontiersin.org/article/10.3389/frobt.2018.00015/full

67. Beauchamp T. The Principle of Beneficence in Applied Ethics. Stanford Encycl Philos (Spring 2019 Ed Edward N Zalta (ed),. 2019;

68. Aquino YSJ, Carter SM, Houssami N, Braunack-Mayer A, Win KT, Degeling C, et al. Practical, epistemic and normative implications of algorithmic bias in healthcare artificial intelligence: a qualitative study of multidisciplinary expert perspectives. J Med Ethics [Internet]. 2023 Feb 23;jme-2022-108850. Available from: https://jme.bmj.com/lookup/doi/10.1136/jme-2022-108850

69. Desouza KC, Dawson GS, Chenok D. Designing, developing, and deploying artificial intelligence systems: Lessons from and for the public sector. Bus Horiz [Internet]. 2020 Mar;63(2):205–13. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0007681319301582

70. Mhlambi S. Q&A: Sabelo Mhlambi on what AI can learn from Ubuntu ethics [Internet]. People + AI Research. Available from: https://medium.com/people-ai-research/q-a-sabelo-mhlambi-on-what-ai-can-learn-from-ubuntu-ethics-4012a53ec2a6

71. Hofstede, Geert. 2001. Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations. Sage Publications.

72. Goffi, Emmanuel R., Colin, Louis, and Belouali, Saida. 2021. Ethical Assessment of AI Cannot Ignore Cultural Pluralism: A Call for Broader Perspective on AI Ethics. Human Rights in Africa & the Mediterranean International Journal 48–71.

73. Fisher S, Rosella LC. Priorities for successful use of artificial intelligence by public health organizations: a literature review. BMC Public Health [Internet]. 2022 Nov 22;22(1):2146. Available from: https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-022-14422-z

74. Brand D, Singh JA, McKay AGN, Cengiz N, Moodley K. Data sharing governance in sub-Saharan Africa during public health emergencies: Gaps and guidance. S Afr J Sci [Internet]. 2022 Oct 26;118(11/12). Available from: https://sajs.co.za/article/view/13892

75. Tobin MJ, Jubran A. Pulse oximetry, racial bias and statistical bias. Ann Intensive Care. 2022 Dec;12(1):2.

76. Santoni de Sio F, Mecacci G. Four Responsibility Gaps with Artificial Intelligence: Why they Matter and How to Address them. Philos Technol [Internet]. 2021 Dec 14;34(4):1057–84. Available from: https://link.springer.com/10.1007/s13347-021-00450-x

77. Königs P. Artificial intelligence and responsibility gaps: what is the problem? Ethics Inf Technol [Internet]. 2022 Sep 24;24(3):36. Available from: https://link.springer.com/10.1007/s10676-022-09643-0

78. Nebeker C, Torous J, Bartlett Ellis RJ. Building the case for actionable ethics in digital health research supported by artificial intelligence. BMC Med. 2019;17(1):1–7.

79. Omrani N, Rivieccio G, Fiore U, Schiavone F, Agreda SG. To trust or not to trust? An assessment of trust in AI-based systems: Concerns, ethics and contexts. Technol Forecast Soc Change [Internet]. 2022 Aug;181:121763. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0040162522002888

80. Jonas H. The Impreative of Respoonsibility: In search of an Ethics for Technological Age.

University of Chicago Press.; 1984.

81. Diaz Milian R, Bhattacharyya A. Artificial intelligence paternalism. J Med Ethics [Internet]. 2023 Jan 20;jme-2022-108768. Available from: https://jme.bmj.com/lookup/doi/10.1136/jme-2022-108768

82. Varkey B. Principles of Clinical Ethics and Their Application to Practice. Med Princ Pract [Internet]. 2021;30(1):17–28. Available from: https://www.karger.com/Article/FullText/509119

83. Leonelli S. Locating ethics in data science: Responsibility and accountability in global and distributed knowledge production systems. Philos Trans R Soc A Math Phys Eng Sci. 2016;374(2083).

84. Dhirani LL, Mukhtiar N, Chowdhry BS, Newe T. Ethical Dilemmas and Privacy Issues in Emerging Technologies: A Review. Sensors [Internet]. 2023 Jan 19;23(3):1151. Available from: https://www.mdpi.com/1424-8220/23/3/1151

85. Ulman YI. Social Ethics. In: Encyclopedia of Global Bioethics [Internet]. Cham: Springer International Publishing; 2016. p. 2632–41. Available from: https://link.springer.com/10.1007/978-3-319-09483-0_395

86. Charter PR. The Patients' Rights Charter [Internet]. 1996 p. 1–3. Available from: http://www.doh.gov.za/docs/legislation/patientsright/chartere.html

87. Constitution of the Republic of South Africa, 1996. Available from: https://www.gov.za/documents/constitution-republic-south-africa-1996

88. Rights BOF. CHAPTER 2. (1):5–20. Available from: https://www.gov.za/documents/constitution/chapter-2-bill-rights#

89. Kaplan AD, Kessler TT, Hancock PA. How Trust is Defined and its use in Human-Human and Human-Machine Interaction. Proc Hum Factors Ergon Soc Annu Meet [Internet]. 2020 Dec 9;64(1):1150–4. Available from: http://journals.sagepub.com/doi/10.1177/1071181320641275

90. Henrietta Lacks: science must right a historical wrong. Nature [Internet]. 2020 Sep 3;585(7823):7–7. Available from: https://www.nature.com/articles/d41586-020-02494-z

91. Fairchild AL, Bayer R. Uses and Abuses of Tuskegee. Science (80-) [Internet]. 1999 May 7;284(5416):919–21. Available from: https://www.science.org/doi/10.1126/science.284.5416.919

92. Jacovi A, Marasović A, Miller T, Goldberg Y. Formalizing Trust in Artificial Intelligence. In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency [Internet]. New York, NY, USA: ACM; 2021. p. 624–35. Available from:

https://dl.acm.org/doi/10.1145/3442188.3445923

93. Winfield AFT, Jirotka M. Ethical governance is essential to building trust in robotics and artificial intelligence systems. Philos Trans R Soc A Math Phys Eng Sci [Internet]. 2018 Nov 28;376(2133):20180085. Available from: https://royalsocietypublishing.org/doi/10.1098/rsta.2018.0085

94. Taddeo M. Trusting Digital Technologies Correctly. Minds Mach [Internet]. 2017 Dec 15;27(4):565–8. Available from: http://link.springer.com/10.1007/s11023-017-9450-5

95. Floridi L. Tolerant Paternalism: Pro-ethical Design as a Resolution of the Dilemma of Toleration. Sci Eng Ethics [Internet]. 2016 Dec 9;22(6):1669–88. Available from: http://link.springer.com/10.1007/s11948-015-9733-2

96. Taddeo M, Floridi L. How AI can be a force for good. Science (80-) [Internet]. 2018 Aug 24;361(6404):751–2. Available from: https://www.science.org/doi/10.1126/science.aat5991

97. Scheutz M, Malle BF. Moral Robots. In: Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction [Internet]. New York, NY, USA: ACM; 2015. p. 117–24. Available from: https://dl.acm.org/doi/10.1145/2696454.2696458

98. Scheutz M. The Need for Moral Competency in Autonomous Agent Architectures. 2016;(2009):517–27.

99. Bankins S, Formosa P. The Ethical Implications of Artificial Intelligence (AI) For Meaningful Work. J Bus Ethics [Internet]. 2023 Feb 11; Available from: https://link.springer.com/10.1007/s10551-023-05339-7

100. Scheutz M, Malle BF. Moral robots. Routledge Handb Neuroethics. 2017;(July 2017):363–77.

101. Morley J, Machado CCV, Burr C, Cowls J, Joshi I, Taddeo M, et al. The ethics of AI in health care: A mapping review. Soc Sci Med [Internet]. 2020 Sep;260:113172. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0277953620303919

102. Russell RG, Lovett Novak L, Patel M, Garvey K V., Craig KJT, Jackson GP, et al. Competencies for the Use of Artificial Intelligence–Based Tools by Health Care Professionals. Acad Med [Internet]. 2023 Mar 6;98(3):348–56. Available from: https://journals.lww.com/10.1097/ACM.0000000000004963

103. Owe A, Baum SD. Moral consideration of nonhumans in the ethics of artificial intelligence. AI Ethics [Internet]. 2021 Nov 6;1(4):517–28. Available from: https://link.springer.com/10.1007/s43681-021-00065-0

104. Asaro PM. A body to kick, but still no soul to damn: Legal perspectives on robotics. 2012;

105. Coeckelbergh M. Robot rights? Towards a social-relational justification of moral consideration. Ethics Inf Technol [Internet]. 2010 Sep 27;12(3):209–21. Available from:

http://link.springer.com/10.1007/s10676-010-9235-5

106. Pagallo U. Robots of Just War: A Legal Perspective. Philos Technol [Internet]. 2011 Sep 3;24(3):307–23. Available from: http://link.springer.com/10.1007/s13347-011-0024-9

107. Heine K, Quintavalla A. Bridging the accountability gap of artificial intelligence – what can be learned from Roman law? Leg Stud [Internet]. 2023 Jan 18;1–16. Available from: https://www.cambridge.org/core/product/identifier/S0261387522000514/type/journal_article

108. Van Niekerk A. The ethics of responsibility: Fallibilism, futurity and phronesis. STJ | Stellenbosch Theol J [Internet]. 2020 Aug 28;6(1):207–27. Available from: https://ojs.reformedjournals.co.za/stj/article/view/2062

109. Darling K. Extending Legal Rights to Social Robots. SSRN Electron J [Internet]. 2012; Available from: http://www.ssrn.com/abstract=2044797

110. Gunkel DJ. The other question: can and should robots have rights? Ethics Inf Technol [Internet]. 2018 Jun 17;20(2):87–99. Available from: http://link.springer.com/10.1007/s10676-017-9442-4

111. Coeckelbergh M. David J. Gunkel: The machine question: critical perspectives on AI, robots, and ethics. Ethics Inf Technol [Internet]. 2013 Sep 3;15(3):235–8. Available from: http://link.springer.com/10.1007/s10676-012-9305-y

112. Gunkel DJ. A Vindication of the Rights of Machines. Philos Technol [Internet]. 2014 Mar 30;27(1):113–32. Available from: http://link.springer.com/10.1007/s13347-013-0121-z

113. Tavani H. Can Social Robots Qualify for Moral Consideration? Reframing the Question about Robot Rights. Information [Internet]. 2018 Mar 29;9(4):73. Available from: http://www.mdpi.com/2078-2489/9/4/73

114. Ray PP. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet Things Cyber-Physical Syst [Internet]. 2023 Apr; Available from: https://linkinghub.elsevier.com/retrieve/pii/S266734522300024X

115. Louie B, Björling EA, Kuo AC, Alves-Oliveira P. Designing for culturally responsive social robots: An application of a participatory framework. Front Robot AI [Internet]. 2022 Oct 20;9. Available from: https://www.frontiersin.org/articles/10.3389/frobt.2022.983408/full

116. Park N, Jang K, Cho S, Choi J. Use of offensive language in human-artificial intelligence chatbot interaction: The effects of ethical ideology, social competence, and perceived humanlikeness. Comput Human Behav [Internet]. 2021 Aug;121:106795. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0747563221001187

117. Bonnefon J-F, Shariff A, Rahwan I. The social dilemma of autonomous vehicles. Science (80-)

[Internet]. 2016 Jun 24;352(6293):1573–6. Available from: https://www.science.org/doi/10.1126/science.aaf2654

118. The Guardian. Chess robot grabs and breaks finger of seven-year-old opponent. Available from : https://www.theguardian.com/sport/2022/jul/24/chess-robot-grabs-and-breaks-finger-of-seven-year-old-opponent-moscow#:~:text=Last%20week%2C%20according%20to%20Russian,match%20at%20the%20Moscow%20Open.

119. Gerke S, Minssen T, Cohen G. Ethical and legal challenges of artificial intelligence-driven healthcare. In: Artificial Intelligence in Healthcare [Internet]. Elsevier; 2020. p. 295–336. Available from: https://linkinghub.elsevier.com/retrieve/pii/B9780128184387000125

120. Kasher N. Deontology and Kant. Rev Int Philos [Internet]. 1978;32(126):551–8. Available from: https://www.jstor.org/stable/23944141

121. Hutler B, Rieder TN, Mathews DJH, Handelman DA, Greenberg AM. Designing robots that do no harm: understanding the challenges of Ethics for Robots. AI Ethics [Internet]. 2023 Apr 17; Available from: https://link.springer.com/10.1007/s43681-023-00283-8

122. Immanuel K. Groundwork of the Metaphysic of Morals. M GM and KC, editor. Philosophy: The Classics. Cambridge University Press; 1785. 151–158 p.

123. Immanuel K. Grounding for the metaphysics of morals: With on a supposed right to lie because of philanthropic concerns. Hackett Publishing Company. 1993;

124. CourseHero. Tutorial SDL AI Ethics.docx - Foundation of AI Tutorial [Internet]. Available from: https://www.coursehero.com/file/125194868/TutorialSDL-AI-Ethicsdocx/

125. Bentham J. An Introduction to the Principles of Morals and Legislation. Problemos. 2013;83:188–90.

126. Ingrand F, Ghallab M. Deliberation for autonomous robots: A survey. Artif Intell [Internet]. 2017 Jun;247:10–44. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0004370214001350

127. Teichmann R, editor. The Oxford Handbook of Elizabeth Anscombe [Internet]. Oxford University Press; 2022. Available from: https://academic.oup.com/edited-volume/43987

128. Wax ML, MacIntyre A. After Virtue: A Study in Moral Theory. Contemp Sociol [Internet]. 1982 May;11(3):346. Available from: http://www.jstor.org/stable/2067167?origin=crossref

129. Aristotle. The Ethics of Aristotle – The Nicomachean Ethics. Harmondsworth: Penguin Books, 1953.

130. Anderson M, Anderson SL. Machine Ethics: Creating an Ethical Intelligent Agent. In: Machine Ethics and Robot Ethics [Internet]. Routledge; 2020. p. 237–48. Available from:

https://www.taylorfrancis.com/books/9781000108934/chapters/10.4324/9781003074991-22

131. Aristotle, “Nicomachean Ethics, Book 2,” in Aristotle in 23 Volumes, trans. H Rackham, vol. 19, 23 vols. (London: Harvard University Press, 1934), http://data.perseus.org/citations/urn:cts:greekLit:tlg0086.tlg010.perseus-eng1:2.

132. Rachels J and Rachels S. The element of moral philosophy. Boston: MCGraw Hill; 2010.

133. Vallor S. Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting. Oxford: Oxford University Press. 2016.

134. Weber M. Politics as a Vocation. http://anthropos-lab.net/wp/wp-content/uploads/2011/12/Weber-Politics-as-a-Vocation.pdf (accessed 3 May 2023).

135. Bauman Z. Postmodern Ethics. Oxford: Blackwell, 1993.

136. Levinas E. Ethics and Infinity. Pittsburgh: Duquesne University Press, 1985.

137. Ruti M. THE ETHICS OF PRECARITY: JUDITH BUTLER’S RELUCTANT UNIVERSALISM. In: Remains of the Social [Internet]. Wits University Press; p. 92–116. Available from: http://www.jstor.org/stable/10.18772/22017030305.8

138. van Niekerk AA. Ethics for Medicine and Medicine for Ethics. South African J Philos [Internet]. 2002 Jan 28;21(1):35–43. Available from: http://www.tandfonline.com/doi/full/10.4314/sajpem.v21i1.31334

139. Van Niekerk AA, Nortjé N. Phronesis and an ethics of responsibility. South African J Bioeth Law [Internet]. 2013 Jun 24;6(1):26. Available from: http://hmpg.co.za/index.php/sajbl/article/view/7528

140. Huber W. Ethics of responsibility in a theological perspective. STJ | Stellenbosch Theol J [Internet]. 2020 Aug 28;6(1):185–206. Available from: https://ojs.reformedjournals.co.za/stj/article/view/2083

141. Thomas EL. Emmanuel Levinas: Ethics, Justice and the Human beyond Being, London: Routledge. 2004.

142. Davy BJ. An Other Face of Ethics in Levinas. Ethics and the Environment. 2007;12(1):39-65.

143. De Villiers J-H. Thinking-of-the-Animal-Other with Emmanuel Levinas. Potchefstroom Electron Law J [Internet]. 2020 Nov 3;23:1–18. Available from: https://journals.assaf.org.za/index.php/per/article/view/8974

144. Dalgliesh B. Zygmunt Bauman and the Consumption of Ethics by the Ethics of Consumerism. Theory, Cult Soc [Internet]. 2014 Jul 17;31(4):97–118. Available from: http://journals.sagepub.com/doi/10.1177/0263276413508447

145. Zygmunt Bauman and LD. Moral Blindness: The Loss of Sensitivity in Liquid Modernity. 2013.

146. Mills C. Undoing Ethics: Butler on Precarity, Opacity and Responsibility. In: Butler and Ethics [Internet]. Edinburgh University Press; 2015. p. 41–64. Available from: https://www.degruyter.com/document/doi/10.1515/9780748678860-004/html

147. Aristotle. The Ethics of Aristotle – The Nicomachean Ethics. Harmondsworth: Penguin Books, 1953.

148. Hacking, I. 2002. Inaugural lecture: chair of philosophy and history of scientific concepts at the College de France, 16 January 2001. Econ Soc 31(1):1–14.

149. Bernstein, R. 1986. Philosophical profiles. Cambridge: Polity Press.

150. Thompson, M. 2017. "How managers understand wisdom in decision making: a phronetic research approach", Ch. 9 in Kupers W, Gunnlaugson O (eds) Wisdom learning: perspectives on wising up business and management education, Routledge, London and New York, Tay.

151. World Health Organisation. Guidance for managing ethical issues in infectious disease outbreaks. World Heal Organ [Internet]. 2016;62. Available from: file:///G:/AM/Research/AMR/WHO .pdf

152. Ricoeur P. Hermeneutics and the Human Sciences. Cambridge: Cambridge University Press, 1981.