UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Robots and Dignity from an Afro-Communitarian Perspective

by

Karabo Samuel Maiyane

A dissertation submitted in fulfilment of the requirements for the degree

PhD in Philosophy

in the Department of Philosophy at the

UNIVERSITY OF PRETORIA

FACULTY OF HUMANITIES

SUPERVISORS: Prof. Emma Ruttkamp-Bloem and Prof. Thaddeus Metz

April 2024

ACKNOWLEDGEMENTS

What to say?

After writing so many words for sooo long I'm all worded out. All that I am left with is humility and gratitude. This was by far the most difficult thing I ever had to do. I cannot tell you how much and how many times I wanted to quit. But were here now 😊

To my supervisors Prof Emma Ruttkamp Bloem and Prof, Thaddeus Metz, thank you for putting up with me. I imagine this was not an easy feat. But through your patience and guidance we have made it here. I can say with confidence that through your guidance I am a better researcher now.

To my mentor, Dr Mpho Tshivhase, thank you for always being in my corner. You know all too well to cost of this journey for me. Thank you for always looking out for me.

To my sponsors, I would also like to extend my gratitude to CSIR-CAIR for beginning this journey with me from my honour's year until my early PhD just. It was unfortunate we had to part ways, but I am grateful for your support throughout the most of my postgraduate journey.

To the NIHSS, the financial assistance of the National Institute for the Humanities and Social Sciences, in collaboration with the South African Humanities Deans Association towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at are those of the author and are not necessarily to be attributed to the NIHSS and SAHUDA.

To my friends and family. You know my heart, you know what it cost, thank you for sticking with me.

I love you all.

PLAGIARISM DECLARATION

PLAGIARISM, ORIGINALITY, and INTEGRITY STATEMENT

I Karabo Maiyane (29505852……..) confirm that I have read and understand the following integrity statement and description of declaration of originality:

The University of Pretoria commits itself to produce academic work of integrity. The Department of PHILOSOPHY places great emphasis upon integrity and ethical conduct in the preparation of all written work submitted for academic evaluation.

While academic staff teaches you about referencing techniques and how to avoid plagiarism and engaging in unoriginal work practices, you too have a responsibility in this regard. If you are at any stage uncertain as to what is required, you should speak to your lecturer before any written work is submitted.

Plagiarism is a serious form of academic misconduct. It involves both appropriating someone else's work and passing it off as one's own work afterwards. Thus, you commit plagiarism when you present someone else's or AI generated written or creative work (words, images, ideas, opinions, discoveries, artwork, music, recordings, computer-generated work, etc.) as your own. Only hand in your own original work. Indicate precisely and accurately when you have used information provided by someone else. Note however that you are not allowed to use any AI generated text directly in your work even if you acknowledge it. Referencing of all other sources, including information downloaded from the Internet, must be done in accordance with a recognised system.

Students who commit plagiarism or submit AI generated text will not be given any credit for this work. The matter may also be referred to the Disciplinary Committee (Students) for a ruling. Plagiarism and unoriginal work are regarded as serious contraventions of the University's rules and can lead to expulsion from the University. Ensure you read the University of Pretoria Plagiarism Policy which is available on your ClickUP page for this module.

The declaration that follows must accompany all written work submitted while you are a student of the Department of PHILOSOPHY. No written work will be accepted unless the declaration has been completed and attached.

DECLARATION OF ORGININALITY

I affirm that I am aware of and have read the Rules and Policies of the University, more specifically the Disciplinary Procedure, the Plagiarism Policy, and the Tests and Examinations Rules, which prohibit any unethical, dishonest, or improper conduct and any form of plagiarism (including passing off AI generated work as your own) during tests, completion of assignments or essays, examinations, and/or any other forms of assessment.
I am aware that no student or any other person may assist or attempt to assist another student, or obtain help, or attempt to obtain help from another student or any other person or AI system during the writing of tests or examinations and/or any other forms of assessment such as class tests and tutorial assignments.

Full names of student: ..Karabo Samuel Maiyane .....

Student number: .....29505852 .....

Declaration

1. I understand what plagiarism and unoriginal work are and am aware of the University's policy in this regard.
2. I declare that this .....Thesis ..... (e.g. essay, report, project, assignment, dissertation, thesis, etc.) is my own original work. Where other people's work has been used (either from a printed source or Internet source), this has been properly acknowledged and referenced in accordance with university requirements.
3. I have not used work previously produced by another student or any other person or AI system to hand in as my own.
4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her or their own work.

DATE **……30/04/2024……**

SIGNATURE .....  .....

ETHICS STATEMENT

Fir

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Humanities
Fakulteit Geesteswetenskappe
Lefapha la Bomotho

Humanities 100. Since 1919

6 April 2021

Dear Mr KS Maiyane

Project Title:	Understanding human dignity in relation to artificial intelligence within the human-robot interaction domain
Researcher:	Mr KS Maiyane
Supervisor(s):	Prof EB Ruttkamp Prof TH Metz
Department:	Philosophy
Reference number:	29505853 (HUM028/0221)
Degree:	Doctoral

Thank you for the application that was submitted for ethical consideration.

The Research Ethics Committee notes that this is a literature-based study and no human subjects are involved.

The application has been **approved** on 25 March 2021 with the assumption that the document(s) are in the public domain. Data collection may therefore commence, along these guidelines.

Please note that this approval is based on the assumption that the research will be carried out along the lines laid out in the proposal. However, should the actual research depart significantly from the proposed research, a new research proposal and application for ethical clearance will have to be submitted for approval.

We wish you success with the project.

Sincerely,

Prof Innocent Pikirayi
Deputy Dean: Postgraduate Studies and Research Ethics
Faculty of Humanities
UNIVERSITY OF PRETORIA
e-mail: PGHumanities@up.ac.za

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Humanities

RECOMMENDATION REGARDING TITLES OF THESES/DISSERTATIONS/MINI-DISSERTATIONS,

APPOINTMENT OF SUPERVISORS/CO-SUPERVISORS

THIS FORM MUST PLEASE BE TYPED AND SUBMITTED TO THE RESEARCH ETHICS / POSTGRADUATE COMMITTEE.

Student number: 11067561 Title: Mr

Surname: Maiyane Initials: K

Course e.g. MA(Anthropology): PhD (Philosophy)

Department: Philosophy

Date of ethical clearance: 25 March 2021

Ethics protocol number: 29505853 (HUM028/0221)

Please ensure that the title is grammatically correct.
Please do not type the full title in capital letters.

For notification of the faculty board, please mention:

THESIS: __X__
DISSERTATION: _____
MINI-DISSERTATION: ____

Current research title:
Understanding human dignity in relation to artificial intelligence within the human robot interaction domain

New research title (to be approved):

Robots and Dignity from an Afro-Communitarian Perspective

a) SUPERVISOR:
Please state staff number
(If a person outside the University is nominated, a CV and motivation should be attached to the proposal)

Prof EB Ruttkamp-Bloem

02598531

b) CO-SUPERVISOR:
Please state staff number
(If a person outside the University is nominated, a CV and motivation should be attached to the proposal)

APPROVAL:

- ☐ Ethical clearance to be considered by the Ethics Committee
- ☐ Appointment of external supervisor/co-supervisor (motivation and cv attached)
- ☐ Change of supervisor/co-supervisor
- ☐ Research proposal and title to be considered by the Postgraduate Committee (where applicable)
- ☒ X Editorial change of title
- ☐ Other (mention): _____

ROUTE

1. SUPERVISOR:

Signature

Date 25 June 2023

2. DEPARTMENTAL RESEARCH/POSTGRADUATE COMMITTEE:

Signature

Date 22 June 2023

3. HEAD OF DEPARTMENT:

Signature

Date 22 June 2023

4. FACULTY RESEARCH ETHICS/ POSTGRADUATE COMMITTEE:
Where applicable (to approve research proposal/ethics; change of title; appointment or change of internal/external supervisor/co-supervisor;

Signature..... obo Prof Pikirayi

Date 26/06/2023

5. STUDENT ADMINISTRATION:

Signature.....

Date.....

ABSTRACT

One of the criticisms raised against using technologies powered with artificial intelligence (AI) is that some of their uses would undermine human dignity. This argument is prevalent in healthcare and military robotics, where it is argued that the use of these AI technologies in various domains undermines or violates the dignity of those human beings who are in contact with them. In this thesis, I look to investigate *Under which conditions various AI technologies would undermine human dignity*. To answer this question, I first develop an African communitarian conception of dignity, which I will then use to evaluate the impact of AI technologies on dignity in the domains of healthcare and warfare. What is different about communitarian conceptions proposed here is that they prize different normative values, such as community over individual interest and duties over rights. In my conception, I argue that one has dignity when they have the capacity for communal relationships as subjects or objects. To honour this dignity, this dignity is respecting other agents capacity. Based on this evaluation, I argue that there are clear cases where AI technologies would undermine dignity, just as there are also clear cases where their use would enhance it. This thesis contributes to the philosophical discourse in two main ways. First, it advances an African Communitarian conception of dignity as a plausible conception to evaluate the impact of AI technologies on human dignity. Secondly, it evaluates the impact of dignity using the proposed conception formulated, bringing out relational considerations that have not been salient in debates up to now.

KEYWORDS

Dignity, human dignity, Human-Robot interaction, Afro-communitarianism, social robots, care robots, autonomous weapon systems

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

1.1. Introduction

The development in the fields of robotics and artificial intelligence (AI) has seen massive growth. Lately, such noticeable growth has been seen at the intersection of the two fields where the robots are now powered with artificial intelligence capacities, making them more responsive and intuitive. While this is a welcomed development, it does not come with challenges. Prevalent among these issues and what is often cited as the reason why some of these technologies must be welcomed is that some of their uses would undermine dignity.

It is this claim that I seek to test in my thesis. Two questions emerge: 1) what is meant by dignity, and 2) under which conditions would the use of AI technologies undermine it?

In this chapter, as an introduction to my thesis, I will provide the background of the development of artificial intelligence and robotics. Following this, I will outline the ethical issues that emerge in general and as they pertain to dignity. I will briefly explain what I mean by dignity and under which conditions it can be undermined. After this, I will provide the outline of the following chapters.

1.2.Artificial Intelligence and Robotics

Artificial intelligence (AI) has been a focused academic interest since the 1950s. Although the term itself was coined in 1956 by John McCarthy and others such as Simon Newell and Marvin Minsky at a conference in Dartmouth, UK (Bringsjord and Govindarajulu 2018), there is much evidence of work related to the field dating from before 1956. One such piece of evidence is Alan Turing's (1950) paper titled 'Computing Machinery and Intelligence', where he asks whether a machine could think. In this paper, Turing argues that if a machine can deceive a human being into believing it is a person, it can be considered intelligent. He developed a test for machine intelligence known as the 'Turing Test'. Also, the work of John

von Neumann, one of the fathers of computing, must be mentioned here (see Von Neumann and Burks 1966; O., von Neumann, and Taub 1964). However, of course, the very first original roots of AI lie in the work of George Boole, captured in his 1854 *Laws of Thought* with Aristotle as the father of classical syllogistic logic.

The academic scope of Artificial Intelligence was interdisciplinary from the start of the conception of the domain in the 1950s, combining work by computer scientists, mathematicians, philosophers, linguists, and psychologists. A shared interest in how *intelligent* agents – biological or artificial – acquire, process, represent, and convey knowledge forced researchers to work together when conducting research focused on communication, perception, action, reasoning, learning and representation. The unifying notion of this collaborative and interdisciplinary research is that of intelligent agents. So, what is intelligence? In what follows, I will refer to both human and artificial intelligence.

Human intelligence refers to the capacity to *learn* from experience, *adapt* to new situations, *understand* and handle abstract concepts, use knowledge to manipulate one's environment, and/or deal flexibly and effectively with practical, theoretical problems (Sternberg 2023). Human beings are considered intelligent because they have mental capacities that associate them with one or more of these capacities. Artificial Intelligence refers to "any kind of artificial computational system that shows intelligent behaviour, i.e., complex behaviour that is conducive to reaching goals" (Müller 2021). In this regard, intelligence does not refer to the strict humancentric understanding explained above. In the context of AI, it means to "*incorporate* a range of machines, including those in 'technical AI' that show only limited abilities in learning or reasoning but excel at the automation of tasks, as well as machines in work on 'general AI' that aims at creating a generally intelligent agent" (ibid. my emphasis). These tasks involve, "sensing, modelling, planning, action, perception, text analysis, natural language processing, logical reasoning, game-playing,

decision support systems, data analytics, predictive analytics, as well as autonomous vehicles and other forms of robotics" (ibid.).

Additionally, Selmer Bringsjord and Naveen Govindarajulu (2018) argue that we can consider AI systems to be reason-based or behaviour-based. Reason-based AI refers to systems that mimic human rationality - for example, decision systems in the financial sector. Behaviour-based systems act in a way that could be construed as rational. For example, the humanoid robot named Sophia interacts with human beings like other human beings. Turing and others focused on the latter, while researchers in knowledge representation and reasoning focused on the former.

In its first 50 years or so, the focus on AI was mainly knowledge or logic-based. The focus was on how an 'intelligent' agent would represent knowledge, to be able to act on it or reason about it. This focus is why logic-based systems use rules to encode human knowledge and reasoning. During the early 2000s, it became clear that the explosion of big data was heralding a movement away from logic-based AI to data-driven AI. This movement is underpinned by methods different from those used in logic-based AI, such as neural networks, Bayesian networks, and evolutionary algorithms, to name some (Bringsjord and Govindarajulu 2018).

It must be noted again that much earlier work was done in terms of data-driven AI as well, for instance, the work of the neuropsychologist Donald Hebb on how learning affects neurons in the human brain. However, research on artificial neural networks was computationally very expensive, and it was only with the speeding up of computer micro-processing that this kind of research on neural networks received a new boost in the late 1980s (Anderson and Rosenfeld 1998). A notable breakthrough in machine learning was the work of Geoffrey Hinton (2012), who applied deep learning techniques to computer vision. Another explosive area of research in modern AI is work done with probabilistic methods

(Bringsjord and Govindarajulu, 2018). The advent of Bayesian networks, a new formalism which combines graph probabilism and graph theory (ibid.), heralded new methods to model conditional dependence and, thus, causation. The work of Judea Pearl (1988), *Probabilistic Reasoning in Intelligent Systems,* was a pivotal text in this field.

Until the advent of data-driven AI, philosophical interest in the field of AI was mostly limited to debates in the philosophy of mind, on the nature of consciousness and cognition, and in epistemology, on the nature of reasoning about knowledge and belief. More recently, philosophical interest has been aroused by the fact that "new [data-driven] technologies challenge current norms and conceptual systems" (Müller 2021) concerning intelligent behaviour and reasoning. Thus, the current philosophical project of AI aims to investigate aspects of artificial entities that have features central to how we humans perceive ourselves, such as feeling, thinking and intelligent action. Specifically, the interest is in the moral implications of creating and ascribing such features to AI in its various applications.

Research fields that have gained prominence as a result of this interest include data ethics, information ethics, robot ethics, machine ethics and neuro-ethics (Ruttkamp-Bloem 2020). Data ethics refers to a "branch of ethics that studies and evaluates moral problems related to algorithms and corresponding practices. [It] focuses on ethical problems posed by the collection and analysis of large datasets and on issues ranging from the use of big data in biomedical research and social sciences to profiling, advertising and data philanthropy as well as open data. It aims to formulate and support morally good solutions (Floridi and Taddeo 2016, pp. 1–4).

Data ethics is the most significant field when we want to understand AI because when we refer to an AI, we refer to an algorithm. It is this algorithm that is manipulated using different tools to make it perform certain tasks. It can operate as software, e.g., algorithms in banks that determine who is eligible for a loan, and it can also be programmed into robots.

When we evaluate the moral acceptability or permissibility of the robot's behaviour, we call that reflection 'robot ethics'. In the domain of machine ethics, we study the possibility and moral permissibility of this algorithm embodied in a robot as an agent. Thus, when we consider the ethical issues of an AI, be it in the form of software or hardware, at the core of our concern is the algorithm that drives its behaviour. Luciano Floridi and Mariarosaria Taddeo (2016, p. 3) argue, "It is not the hardware that causes ethical problems, it is what the hardware does with the software and the data that represents the source of our new difficulties".

Robot ethics is a field of study that focuses on a combination of the professional ethics of roboticists, ethical issues that arise as a result of human-robot interaction, the moral code programmed in robots, and the self-conscious ability to do ethical reasoning by robots (Abney 2014; Asaro 2006). When the debate is about the machine having its moral code and self-conscious abilities, we have moved to machine ethics. This is a research field concerned with "ethics *for* machines, for 'ethical machines', for machines as subjects, rather than for the human use of machines as objects" (Müller 2021; Siau and Wang 2020). It deals with the design and behaviours of machines as artificial moral agents (AMAs) (see also M. Anderson & Anderson, 2007; Moor, 2006; Wallach & Allen, 2009).

All these fields are important and hold a wide array of literature. There are also many intersection points that exist among them. I have deliberately withheld explaining the ethics of AI because it is the main area of interest in this thesis, with an overlap in robot and machine ethics. As such, I have dedicated a section to the domain of the ethics of AI. In this thesis, I am particularly interested in the intersection between artificial intelligence as it is embodied in robotics in the domains of healthcare and warfare because my aim is to investigate (from an African perspective) the use of the concept of human dignity to call for limitations or prohibitions in the use of AI in these domains.

1.3. The Ethics of Artificial Intelligence

The ethics of AI refers to a field that "studies the ethical principles, rules, guidelines, policies, and regulations that are related to AI"(Siau and Wang 2020, p. 74). It ultimately aims to investigate whether various applications of AI are morally (and legally) *permissible* under certain circumstances and whether automatic decision-making happens fairly and transparently. For example, is it morally permissible to relegate the care of children or elders to robots? Is it justifiable that the *NYPD stop-and-frisk* program mostly indicated Black and Hispanic persons as targets? Is it permissible for a human to have a love relationship with an AI, either as a bot or a robot? Is it acceptable that the decision-making processes of AI systems are not transparent? Is it permissible for autonomous weapons to target people or autonomously choose whom to harm? Who is responsible if a self-driving car kills a pedestrian? Is it permissible to replace humans with robots in cases of life and death? These questions and many others are investigated in a field called the ethics of AI. According to Muller (2021), the main debates in this domain include privacy and surveillance, manipulation, opacity, bias, human-robot interaction (HRI), automation and employment, autonomous systems, machine ethics, artificial moral agents and singularity (Gordon and Nyholm n.d.). These debates are held in different domains and sectors where certain AI technologies are applied. Debates in the Human-Robot Interaction (HRI) domain on the design, development, and deployment of autonomous systems are of interest to my thesis. I will briefly outline what both domains entail.

Human-Robot Interaction (HRI) is an academic field that studies interactions between humans and robots. It pays significant attention to ethical matters, the dynamics of perception from both sides (the human and the robot side), and the different interests present in them (Müller 2021). Essentially, it investigates the ethical issues that emerge as a results of the development and deployment of robots in direct contact with human beings. The main

domains and technologies that are considered in HRI are care robotics and sex robotics. Some of the ethical issues that emerge as a result of this are discrimination, dehumanisation and deception (Wullenkord and Eyssel 2020). On the issue of discrimination, the argument is that robots are programmed by humans and may be biased. That is, they might discriminate against specific groups because of how available data portrays them. An example of this is the NYPD stop and frisk systems, which targeted African Americans and Hispanics as potential committers of crime (Asaro 2016; Müller 2021; Sparrow 2020). Another issue that is often raised is that the successful use of robots, especially in healthcare, would lead to reduced human contact. This could lead to the nature of care being dehumanised(N. Sharkey and Sharkey 2012).

Regarding deception, the argument is that certain AI applications, especially in robotics, are developed so that they 'deceive' humans. In robotics, some developers prey on the human tendency to anthropomorphise objects. They feed into this tendency by creating robots that give the impression that they can reason and understand (e.g., Sophia), care (care robots), or be suitable companions (sex robots and chatbots). The ethical issue with deception is that it "threatens human dignity or violates the Kantian requirement for respect for humanity" (Müller 2021).

This phenomenon is evidenced mostly in healthcare robots and sex robots. There has been a proliferation of the use of robots in the healthcare sector to assist vulnerable patients, such as older persons, people with disabilities, and children, with daily tasks and social and emotional support (A. Sharkey 2014; A. Sharkey and Sharkey 2011, 2012). Since their introduction, these robots have found the most prominence in older persons' care for several reasons, such as their potential to mitigate an increasing shortage of caregivers for ageing populations (Girling 2021) and reported cases of ill-treatment in certain care facilities by caregivers (Sharkey 2014). The ethical issues (potential and actual) in this regard include

robots infantilising, objectifying, deceiving, and denying privacy to patients needing care (A. Sharkey and Sharkey 2011, 2012; Sparrow and Sparrow 2006).

There is also a growing development and distribution of robots that can be utilised for sexual pleasure. Some argue that the advantages of using robots in this sector such as enabling sexuality without the risk of sexually transmitted diseases and unwanted pregnancies and potentially reducing sex-work-related problems, such as sex trafficking (Döring and Pöschl 2018; Wullenkord and Eyssel 2020) are worth considering. Others argue for the possibility of developing love relationships with such robots (Frank and Nyholm 2017; Sullins 2012). A prevalent issue with sex robots is this issue of deception. Currently, robots do not have sentience, so at best, they provide impressions of human emotional states and do not necessarily experience these states themselves. However, they can use manipulation tools to fool one into believing they do. Part of the deception is also because of the tendency of human beings to anthropomorphise objects. Whether this is an issue in itself is a subject of debate (Muller 2021), which I will touch on below. Other issues associated with sex robots are "consent, aesthetic concerns, and the worry that humans may be 'corrupted' by certain experiences" (ibid.).

In addition to the HRI domain, another domain that has garnered attention in terms of ethical has been autonomous systems. as AI technologies develop and become more sophisticated, there are fears that they could arrive at a point where they are considered to be considered *autonomous*. The argument for those in favour of this claim is that if robots can attain a level of sophistication whereby they can operate solely based on their programming, then they can be considered autonomous (Floridi and Sanders 2004; Moor 2006; Sullins 2006; Wallach and Allen 2009). The major philosophical implication of this claim is that "autonomy is [traditionally known as] the basis for responsibility and personhood" (Müller

2020), meaning that we associate autonomous agents with the capacity for moral *responsibility* and personhood.

Autonomous agents are typically rational persons who, because of their rationality, can make their own choices and act based on free will; thus, they can be held responsible for their actions. In the context of autonomous robots, some arguments suggest that robots can be autonomous but not responsible. These arguments raise questions of whether they are autonomous and, if so, who bears responsibility for their actions. The field of machine ethics aims to study and resolve such ethical issues. Examples of technologies that have (or possibly could have) such capacities are self-driving cars and autonomous weapons systems. Although they are outside of the scope of HRI, in this thesis (chapter 5), I will evaluate autonomous weapons. This is because they are one of the major technologies under consideration when discussing AI technologies' impacts on human dignity.

These domains (healthcare, specifically care-giving and the use of AI in the military) are important for my thesis because AI systems are applied in ways that were traditionally relegated to human beings. As stated, I will specifically focus on care robots and autonomous weapons, and it is noteworthy for the aim of my thesis that both these technologies are applied in domains where the patients or recipients of them are vulnerable. Care robots are mostly used for childcare or caring for persons with cognitive disabilities and older persons (commonly known as marginal or vulnerable groups). Autonomous weapons are used in warfare where they risk the lives of people and other related kinds of harm. Another salient issue prevalent among the two technologies and the core focus of this thesis is human dignity. Many authors argue that the use of these technologies would undermine human dignity. In what follows, I provide a brief literature review regarding these AI technologies and human dignity to set the scene.

1.4. AI and Human Dignity – A Brief Overview

Healthcare robotics and autonomous weapons are some of the key culprits when it comes to the arguments about robotics and dignity. In healthcare robotics, it is argued that the use of certain care robots undermines the dignity of the patients and caregivers in that it infantilises, objectifies, deceives and denies privacy to the very people it is meant to care for (Coeckelbergh 2010, 2012, 2016; A. Sharkey 2014; A. Sharkey and Sharkey 2011, 2012; Sparrow 2016; Sparrow and Sparrow 2006). The main concern here is that introducing AI systems in care spaces traditionally relegated to human beings would affect the quality of care in that care by robotic technologies would be impersonal, humiliating and deceitful.

Additionally, In the healthcare sector, there are concerns regarding the impact on human dignity and responsibility, which may result from over-reliance on AI technologies to make decisions. An example of this is medical surgeons relying on decision systems to make decisions for them. The underlying concern is that the development of such technologies, be it in software or hardware form, would tend to be used as default decision tools instead of support tools. Based on their level of effectiveness, their use as decision-support technologies could replace their users' decision-making capacities. Over time, the users might merely accept the decisions computed instead of critically making them themselves.

For example, surgeons could uncritically implement recommended procedures in intensive care without any thought. Such an action would defer the surgeon's responsibility whilst arguably also undermining the patient's dignity. In this regard, the patient would be treated as an arbitrary entity unworthy of moral consideration (see Friedman and Kahn 1992; Wallach and Allen 2009, p. 40). Such involvement of algorithms in decision-making also affects the autonomy of the patients in making informed decisions since some of the decision procedures are opaque.

Opacity is a common issue in AI decision systems, whereby it is often impossible to learn how machines explain how machine learning decision systems came to a diagnosis of one sort or the other. According to Muller (2021), such decisions will "typically be opaque even to the expert, who will not know how a particular pattern was identified, or even what the pattern is". Thomas Grote and Philipp Berens (2020) argue that such a system undermines the epistemic authority of the clinician. They also undermine the dignity of the patient in that they take away their autonomy to make informed choices about their recommended procedures (see also McDougall 2019).

In terms of autonomous systems, specifically in the context of autonomous weapons, it is argued that using autonomous weapons to target people undermines the dignity of those who are being targeted. Their use might undermine laws or cause armed conflict (Asaro 2012; A. Sharkey 2019). Also, it is argued that autonomous weapons are inanimate objects, thus, incapable of understanding the consequences of killing. For that reason, being targeted by them undermines the value of the agent targeted, treating them as mere means to an end, not ends in themselves. (see Asaro 2012; Goose and Wareham 2016; Heyns 2017; ICRC 2018; Johnson and Axinn 2013; Rosert and Sauer 2019; Ulgen 2018).

These issues will be discussed at length in the following chapters (4 & 5). However, from this brief introduction, it is evident that there is a case to be made regarding how the introduction and use of certain AI technologies would impact human dignity. From this motivation, I consider the following question in this thesis*: under which conditions do various AI technologies undermine human dignity? I look to investigate how – and if so, under which conditions – various AI technologies would have an impact on human dignity, i.e. if they would respect or undermine it.* This question is essential in the context of the ethics of AI, machine ethics and robot ethics because dignity is often cited as a reason these

technologies should not be developed or applied in general but also specifically in the healthcare and military domains.

The investigation of the relationship between AI applications and human dignity is not a new one. Authors like Amanda Sharkey (2014), Arto Laitinen, Marketta Niemelä & Jari Pirhonen (2019), Ozlem Ulgen (2018) and Christoph Heyns (2017) have conducted similar investigations. Amanda Sharkey (2014) argued that to evaluate whether it is the case that various AI applications do undermine dignity, there must first be an understanding or at least a consensus of what we mean by human dignity. She argues that thus far, even the authors who argue that various AI applications undermine human dignity do not seem to agree on what they mean by dignity in this context. Some consider dignity as a status (Ulgen 2018), some as a value/foundational principle (Kant (1785) 1998, United Nations 1945, 1948), and a right (South African Constitution; Heyns 2017). Some authors even go as far as arguing that dignity is vague, multifaceted, and heavily contested (see A. Sharkey 2014; 2019; Werner 2014; Macklin 2003; Ikuenobe 2016; Heyns 2017).

In the literature on dignity and care robots, Sharkey (2014) highlights two different approaches to human dignity - one that looks at dignity as a multifaceted concept option bearing different meanings and Martha Nussbaum's capabilities approach to human dignity – and suggests Nussbaum's capabilities approach (CA) as a more comprehensive and unified perspective on human dignity for different domains of AI applications. Laitinen and colleagues (2019) take on a multifaceted approach to human dignity and argue that the severity of the threat to dignity in the hands of AI technologies, specifically in eldercare robotics, depends on whether the care is robot-based, robot-assisted or teleoperated. They conclude that the threat to dignity is if the care is robot-based.

In the case of military robots, specifically autonomous weapons systems, Ozlem Ulgen (2018) applies Immanuel Kant's conception of dignity to argue that, firstly, the use of

autonomous weapons denies the right of equality of persons since combatants are removed from physical risk by using autonomous weapons. Secondly, using autonomous weapons diminishes the duty not to harm others. In his turn, Christof Heyns (2017) asks whether it is not an affront to human dignity if robots have the power of life and death over humans. He (ibid., p. 49) argues that there are reasons, based on a right to life and dignity, why full machine autonomy over critical functions, such as using force against those who may otherwise be targeted, should not be permitted. Even if someone may be liable to be killed, as is the case with combatants, it does not mean fully autonomous machines may kill that person. He argues that using autonomous weapons to target humans makes the lives of those targeted arbitrary. He continues to argue that there should be 'deliberative human decision-making or – what may amount to the same thing – 'meaningful human control'. Failure of this undermines the right to life and the protection of a dignified life. He (ibid., p. 62) continues to argue, "[e]ven if there is a legitimate limitation on someone's right to life, as is the case where the person is not protected under IHL[1], that person retains his or her other rights, such as the right to dignity".

What is similar about the conceptions of dignity that have been applied to evaluate AI technologies' impact on human dignity is that they are based on individualistic normative concepts prevalent in Western societies. These normative frameworks prize the individual person as a priority and argue that societies must be created in such a way that caters for the individuals and enables their capabilities. Western conception of individualism is not a problem in itself, but it is not a universal viewpoint. Other societies, such as those in Africa

[1] The term 'IHL' refers to International Humanitarian Law. These are laws that are put in place to try and limit the effects of armed conflict (see https://www.icrc.org/en/document/what-international-humanitarian-law).

and some Asian parts, understand the individual as they relate to the community. Essentially that our existence and thus thriving is centred on our relations with others.

So, what is apparently missing in the literature is an evaluation of these AI technologies based on other normative ideas, such as communitarian or communal ideas. It is this gap that I aim to contribute to in this thesis. I aim to develop a communitarian conception of dignity that would be suitable for evaluating the impact AI technologies could have on human dignity.

In African normative thought, it is argued that there are four prevalent conceptions of dignity: the vitalist, communitarian, based on spiritual nature and based on personhood (see Metz 2012, 2014, 2021, Molefe 2022, Ikuenobe 2016). Of the four conceptions, I discuss the communal conceptions in detail, specifically Thaddeus Metz's (2012a, 2021) and Polycarp Ikuenobe's (2018; 2016) conceptions. Metz argues that dignity is the capacity for communal relations ascribed to agents who are subjects in a communal relationship. In contrast, Ikuenobe argues that dignity is the active use of said capacity to ensure harmony in a communal sense. Of these concepts, Metz's conception is the more plausible, with the exception that it is not as inclusive. Thus, I modify Metz's conception of dignity to agents with this capacity but only as objects. I argue for the inclusion of human beings with a partial moral status but who have the capacity for communal relationships as objects. I argue that they have dignity for these reasons: they have a higher moral status than any other entity with the capacity for communal relations as objects, and we recognise and relate with them more meaningfully. My conception provides a plausible African conception of dignity that is suitable to evaluate the impact AI technologies could have on human dignity to formulate a conception that would be plausible to evaluate the impact AI technologies could have on human dignity in communitarian concepts.

Ultimately, I want to determine 1) whether the arguments that AI applications undermine dignity are valid, and, if so, which applications of AI do undermine dignity and which do not, and 2) which conception of human dignity is suitable to evaluate AI applications in an African context and under which conditions this is so. I argue that there are clear cases where AI technologies would enhance human dignity and where their use would undermine human dignity. Care robots[2], for example, would enhance if their use enhanced one's capacity for communal relationships. Care robots could also ensure respect for inherent dignity if their use enables patients to connect better with their caregivers and families. For example, an assistive robot that aids a patient with mobility ensures that patients can reach their loved ones with greater ease than it would otherwise be the case if such contact were only reliant on their loved ones.

They would undermine dignity if, in their use, they discourage social connection between the patients and their caregivers, evoke feelings of humiliation in the patient or are deceitful. For example, it is argued that the success of these robots in the care domain would create less incentive for human caregivers to be present. This, in effect, will evoke feelings of loneliness on the part of the patient. In this regard, I argue that, although less prevalent in communal contexts, technologies that would reduce human social contact would undermine dignity because, in a communal setting, dignity is honoured through acts of cooperation and solidarity.

In terms of humiliation, the argument is that using robots that can be viewed as humiliating undermines the patient's dignity. Regarding privacy, the argument is that surveilling patients without their consent would undermine dignity. In my case, I argue that although privacy is a right that must be respected, such respect cannot trump the duties the

[2]Some arguments on the impact of care robotics to human dignity have already been published (Maiyane 2023a)

patient has towards others. So, the consent in question would not only be of the individual patient but of a network of all affected members. This means the patient cannot decide alone, but in the same breath, it cannot be decided for, even if it is 'in their best interest'. The principle that holds greater value is that of transparency among all the actors. This also applies to the notion of deception. The general intuition in African moral theory is that any act that is deceitful would undermine dignity because deception constitutes unfriendliness. The idea is to appraise all parties involved in the decisions made fully. So, technologies that would rely on deception would be unwelcome. There have to be serious mitigating circumstances where deception would be necessary. However, in more circumstances, transparency is vital.

There are clear cases where there is agreement between the Western and the African conceptions. There is agreement that technologies that would minimise human interaction or cause humiliation would be unwelcome. The difference is in the motivations. While in the Western conception, the motivation is that such technologies would undermine the patient's autonomy, in African conceptions, the motivation is that they would undermine the patient's capacity to coexist with others.

On the other hand, autonomous weapons operate in conflict areas where certain rights and privileges of agents are limited. By this, I mean that during the time of armed conflict, human rights as we know them do not apply when IHL comes into effect. This means certain rights that we hold sacrosanct, such as 'the right to life', are no longer applicable holistically but contextually. For example, not everyone has a right to life in this context. In this case, I argue that the use of autonomous weapons would undermine the dignity of combatants if they were used to target combatants who have surrendered or who are held captive or by targeting combatants in a way that causes unnecessary pain or suffering(torture). Their use would also undermine dignity if they target non-combatants. It is important to note here that it is not the

technologies per se that undermine dignity but how they are applied. So, the issue is more about the agents that apply them than the technologies[3].

My findings in this sense are different to those existing in the literature in that where the literature places the undermining of dignity on the technologies. I argue that the disrespect would occur among agents - autonomous weapons function as instruments to initiate decisions made by human beings. Whereas my colleagues assume that autonomous weapons would have agency as such, we can place responsibility on them; I argue they are merely sophisticated instruments that must be used with care and in compliance with existing war conventions.

This thesis contributes to the philosophical discourse in several ways. Firstly, I will illustrate that there is reason to explore a non-Western and specifically African analysis of dignity. Secondly, regarding the conception of dignity in relation to AI technologies, I introduce an Afro-communitarian conception of dignity as a plausible conception to evaluate the impact of AI applications on human dignity. Thirdly, I will evaluate this proposed conception in the context of healthcare robots and autonomous weapons to stipulate the conditions under which AI applications in the HRI domain would undermine human dignity. The evaluation will provide relevant information on AI technologies' challenges to the dignity of communal environments, thus providing a different lens for policymakers to look through when developing policies for communal societies, such as in Africa and some parts of Asia.

As I have briefly explained above, the conception I will use considers dignity as the capacity for communal relationships. This dignity is ascribed to agents with a partial and full moral status for the reasons provided above. I will evaluate the impact of AI technologies in

[3] some arguments I make here have been published (see Maiyane 2023b)

the case of care robots and autonomous weapons. I show that there are cases where the use of care robots would undermine human dignity. In the case of autonomous weapons, dignity will primarily be undermined if the agents involved in warfare fail to respect the different moral statuses of the agents involved or affected by hostilities. The weapons are merely instruments that must be used in ways that comply with the war convention. In what follows, I outline the chapters of the thesis.

1.5. Chapter Outline

Below, I briefly outline what each chapter entails to sketch the flow of my argument. As an introduction to the thesis, this chapter provided background on the philosophical study of artificial intelligence, specifically ethical issues related to using certain technologies. This was to lay a foundation for the research question: under which conditions would AI technologies undermine dignity? I argued that based on the available literature, there is a clear case to be made that the use of these technologies in certain contexts could undermine dignity. It was also evident that there are cases where their use would enhance respect for human dignity. I highlighted that what is missing in the literature is that the conceptions of dignity used to arrive at these conclusions are mainly, at present, based on Western individualistic normative understandings of dignity, and what I seek to introduce to the discourse is a plausible conception of dignity from an African communitarian perspective that would be capable of conducting the same evaluation. I briefly outlined cases where the use of AI technologies would have an impact on human dignity. The chapters that follow explain this claim in more detail.

Chapter two discusses the concept of dignity in detail. This is done to resolve the claim that there is a lack of consensus regarding the concept of ‘dignity’. It argues that the claim that there is a lack of consensus regarding the concept of dignity is based on neglecting to distinguish between a ‘concept’ and a ‘conception’ of dignity. Clarifying this provides a

working definition of the concept from which all other conceptions emerge. The working definition that will be applied in the thesis is that dignity is an intrinsic value that we have based on our capacities. There are many such capacities, such as rationality, autonomy, sentience, relationality, community, or recognition, to name a few. Different conceptions of dignity emerge based on what capacities the author prize.

Following this, it will then delve into the different conceptions of dignity that are salient in the HRI domain, i.e. Kant's ([1758]1998) and Nussbaum's (2008, 2011). Although these conceptions provide interesting outcomes about how AI technologies would impact human dignity, they both have shortfalls. Kant's basis of dignity on rationality tends to exclude all human beings that might lack it but have other capacities, such as the capacity to relate, sentience, or the capacity to be recognised. Basing dignity only on rationality excludes infants and people with cognitive disabilities. Nusbaum does well in resolving the shortfall levelled against Kant. By basing her conception of dignity on what she terms 'basic capabilities', her conception is much more inclusive. The challenge is that basic capabilities such as sentience extend beyond the scope of human beings. Many animals, such as mice and cows, also have sentience, meaning they would also have dignity in her formulation. So, her conception goes far beyond the scope of a conception that must be suitable to evaluate *human* dignity.

Additionally, these conceptions are all based on Western individualistic conceptions of dignity. Such views prize individual values such as rights, rationality and autonomy. Given that not all societies subscribe to these individualist normative ideals, I propose a communitarian conception of dignity as another plausible conception to evaluate the impact of AI technologies. Communitarian conceptions place greater emphasis on different ideals, such as duty towards a collective, the capacity to relate or the common good. Such values

would be beneficial to societies that prize communal normative ideals, such as those in Africa.

Chapter three proposes the exploration of African conceptions of dignity to find one that would be suitable for evaluating the impact AI technologies could have on human dignity. I begin by briefly outlining conceptions of dignity that are salient in African normative thought, which are vitalist and communitarian, based on spiritual nature and personhood. Following this, I discuss two Afro-communitarian conceptions of dignity that are salient in African philosophical discourse: one by Metz (2012, 2021) and one by Ikuenobe (2016). This aim is to find a conception that would be plausible to evaluate the impact AI technologies could have in communitarian contexts. Metz's conception of dignity is stronger among those discussed because it is more inclusive than Ikuenobe's conception. By basing his conception on the *capacity* for communal relationships as opposed to *acting* on such capacities, his conception is able to include even those people who cannot act or do not want to, such as a prisoner in solitary confinement or a hermit. However, it has some limitations that would make it challenging to evaluate the impact AI technologies could have on human dignity, and as such, I suggest some additions.

The main addition I propose to Metz's conception is to include even human individuals with a partial moral status but with the capacity for communal relationships as objects only. This is because, although such human individuals might lack full moral status, they still hold a higher moral status than any other entity with a partial moral status. As such, we recognise and relate with them better because we are kin to them. For that reason, they ought to have the same privileges in a communal setting. I argue that these reasons provide a sufficient justification for why we should also have dignity. This amended view provides a robust and plausible, more inclusive conception than any other discussed here. In chapters four and five, I use this conception to evaluate the impact AI technologies could have on

dignity in terms of human interaction with care robots and autonomous weapons, respectively.

Chapter four evaluates the impact care robots could have on human dignity based on my plausible conception of human dignity. It starts with explaining what care robots are and then reflects on the ensuing ethical challenges, including the (potential) impact of these robots on human dignity. Using my Afro-communitarian conception of dignity, I evaluate the possible impact of three types of care robots on human dignity: assistive robots, monitoring robots and companion robots. I show that there are clear benefits to using assistive technologies as these robots will be additional tools to aid in an already existing intricate network of care. These robots would undermine dignity if their use would humiliate or be perceived as humiliating, deny social contact, and are deceitful. That is because humiliation can be perceived as disrespectful to a person's capacity for communal relationships.

Also, any robots that create or encourage human distancing dishonour dignity in a communal sense because dignity in this sense is honoured through interdependence, cooperation, or solidarity. In a communitarian sense, value is placed more on coming together than being separated. On the point of deception, the argue that deception constitutes unfriendliness and thus should be discouraged. The more welcomed move would be transparency; only in extreme circumstances where there are clear benefits to the patients and others would such technologies be permissible. The evaluation arrives at similar conclusions in terms of the circumstances whereby dignity would be undermined but for different reasons. For example, the benefits of care robots for Sharkey are the promotion of independence and self-reliance. In contrast, in my case, such assistive robots promote interdependence and the capacity to relate. Another example is the need for robots to be deceptive to make better companions. My conception frowns upon technologies that would require deception to be of assistance because being deceptive constitutes unfriendliness.

Chapter five evaluates the impact the development, deployment, and use of autonomous weapons systems (AWS) could have on human dignity. It first outlines the principles related to the convention of warfare. Following this, it outlines some ethical issues that are often levelled against AWS, especially those pertaining to dignity. Regarding the debates against the deployment and use of AWS, I argue that most arguments against AWS are based on an unjustified assumption that their status in warfare would be that of combatants. Autonomous weapons are not 'autonomous' in a way we understand autonomy to mean for human beings. At most they are just sophisticated weapons.

If we consider AWS as instruments, like any other weapon currently in use, most of the issues labelled are either irrelevant or resolved. I also show that the arguments for dignity in warfare are more complicated than what might be acknowledged in current literature because of the different moral statuses agents hold and their associated entitlements. Essentially, the treatment of agents varies depending on their status.

Using my Afro-communitarian conception of dignity, I first discuss what dignity looks like in warfare and then evaluate whether the introduction of autonomous weapons undermines it. I argue that all human agents with the capacity for communal relationships have dignity in warfare. In this sense, dignity can be honoured by respecting agents' moral statuses and their associated benefits as combatants or non-combatants. I show that autonomous weapons in themselves cannot undermine dignity. As instruments, its users can commit these transgressions, not the weapon. It is the actors who use these weapons who would be undermining dignity if, in using them, they do not distinguish the different roles actors have in warfare and treat them accordingly. This differs from the existing arguments on dignity and autonomous weapons, where the charge of undermining dignity is laid on autonomous weapons.

Chapter six concludes the study and provides a summary of key findings. It also highlights the novel contribution this study provided and some reflections on what we could consider in the future. The findings can be distinguished into two: the conception of dignity and the conditions under which the use of AI technologies would undermine it.

What is key to the findings is that dignity is an important conception that we can use as a yardstick to evaluate the challenges of AI technologies. A plausible African conception of dignity suitable for evaluating the impact of AI technologies on human dignity is one that is inclusive, based on capacity, protects individuals' rights, promotes communal values, and ensures respect for all with capacities. While most conceptions discussed here are good in their own right, they are found wanting when trying to consolidate the features proposed here. My conception ticks all these boxes. Being able to tick off all these features, which are made up of the important qualities from other conceptions discussed in the thesis, means it can be used to evaluate the impact of diversity in both individualist and communal societies.

Now that we have conceptions, the next step is to test it. In the second part, I will reveal what ethical issues emerge when evaluating care robots and autonomous weapons using my conceptions. What is noteworthy about conception is that it is versatile. It can be used in individualist societies just as it can be used in communal societies. Also, unlike other conceptions that focus on one domain of application, I have applied conception in the HRI and the AMA domains. This means adopting it as a concept one can use in any domain as long as the affected parties are human beings. Such a conception will be beneficial to researchers, students and policymakers who can use it to conduct evaluations of any domain of choice

CHAPTER 2: WESTERN CONCEPTIONS OF DIGNITY

2.1. Introduction

The topic of human dignity is one of the essential topics in philosophy. Lately, it has become one of the important debates in the human-robot interaction (HRI) domain. In this domain, it is argued that certain applications of AI technologies would threaten human dignity. Some have argued that the challenge with responding to this argument is that there seems to be no consensus regarding the meaning of the term 'human dignity' (see A. Sharkey 2019; 2014; Werner 2014; Lin 2015). Part of the reasons stated for this supposed lack of consensus is the pervasive use of the concept in various fields (law, ethics, religion, media, literature and politics) or that, in use, it tends to hold different meanings (Riley and Bos n.d.; Schroeder 2010). This claim has even led to others claiming that the term is useless in certain contexts (Macklin 2003). To evaluate AI technologies' implications on human dignity, we must first understand what we mean by the term 'dignity'.

In this chapter, I discuss the concept of dignity to provide a working definition for this thesis. This discussion will also assist in disambiguating the claim that there is a lack of consensus regarding the term. I argue that the problem raised about dignity lacking consensus has more to do with authors not distinguishing between the '*concept*' and the resulting '*conceptions*' of dignity. As such, I will begin by distinguishing between the two notions. Making this distinction provides the necessary clarifications regarding whether it is a term in question or the resulting theory(s) from it.

Secondly, in this chapter, I critically discuss the salient conceptions of dignity prevalent in the HRI domain, i.e., Kant's and Nussbaum's. Although these conceptions provide interesting outcomes about how AI technologies would impact human dignity, they both have shortfalls. Kant's basis of dignity on rationality tends to exclude all human beings

who lack this capacity, such as infants and people with cognitive disabilities. Nusbaum's conceptions, on the other hand, are based on what she terms ‘basic capabilities’, these are “innate faculties of the person that make later development and training possible” (Nussbaum 2011, p. 24) such as sentience, imagination, emotions, choice and reason, to name a few. By basing her conception of dignity on basic capabilities, her conception is much more inclusive. The challenge is that basic capabilities such as sentience extend beyond the scope of human beings. Many animals, such as mice and cows, also have sentience, meaning they would also have dignity in her formulation. So, her conception goes far beyond the scope of a conception that must be suitable to evaluate *human* dignity.

Additionally, these conceptions and others in the domain are based on Western individualistic conceptions of dignity. Such views prize individual values such as rights, rationality and autonomy. Given that not all societies subscribe to these individualist normative ideals, I provide reasons why we should consider the dignity and its evaluation of AI technologies from a communitarian lens. I argue that the notions of individualist normative values and their associated values, such as autonomy and rights, are not representative of all societies. Other societies, mostly in Asia and Africa, are communal and thus prize different values like duty, obligation and family. So, it would be prudent to come up with a plausible conception of dignity that prioritises these values.

In chapter three, I will formulate a communitarian conception of dignity as another plausible conception to evaluate the impact of AI technologies. Unlike individualist conceptions, communitarian conceptions place greater emphasis on different ideals, such as duty towards a collective, the capacity to relate or the common good. Such values would be beneficial to societies that prize communal normative ideals, such as those in Africa.

The introduction of the African conception of dignity as a plausible conception to evaluate the impact of AI technologies on human dignity is a novel contribution. Currently,

debates from this domain usually occur in Europe and America and thus are based on Euro-American ideas of human dignity. There is currently no contribution where African conceptions are used to evaluate this impact. So, this introduction would provide a fresh perspective to the discourse. This chapter does not detail which African conceptions will be considered; it only offers reasons why they should be considered. African conceptions of human dignity will be discussed in detail in Chapter three.

2.2. The Concept of Dignity and Its Interpretations

In this section, I aim to disambiguate the claim that human dignity is a vague concept and, thus, possibly useless. I argue that the problem does not lie with the lack of consensus on the term's meaning. It lies in differentiating between a concept and a conception of dignity. I also outline which working definition of dignity I will be applying in my quest to find a suitable conception of dignity that is applicable within the HRI domain.

What is dignity? Some scholars argue that one of the challenges when discussing the concept of dignity is that there seems to be no consensus regarding what it means (Lin 2015; A. Sharkey 2014, 2019; Werner 2014). Amanda Sharkey (2019, p. 180) argues that "there is a lack of clear consensus about what dignity is". Micha Werner (2014, p. 343) argues that dignity is a "heavily contested, multifaceted and ambiguous concept" (see also Lin 2015). Mattson and Clark (2011, p. 305) also point out that "dignity is variously considered by diverse people to be an antecedent, a consequence, a value, a principle, an experience, and both a contingent and non-contingent exhibition". Some argue that one of the reasons for this lack of consensus could be that the concept is used in various fields in various ways (Riley and Bos, n.d.; A. Sharkey, 2014). Such ubiquitous application of the concept could be why others argue that the idea of dignity is useless in specific fields such as healthcare.

One of the proponents of dignity being a useless concept is Ruth Macklin (2003), who argues that the concept can be eliminated without losing any content in medical ethics. She argues that in its various applications, appeals for dignity are vague restatements of other concepts such as autonomy, respect, responsibility, or slogans that contribute nothing to understanding a topic in conversation. For example, in its use in many international human rights instruments, such as the United Nations Universal Declaration of Human Rights (1948)[4] for Macklin (ibid.), "these conventions do not address medical treatment or research". Even in those that do, such as the Convention for the Protection of Human Rights and Dignity of the Human Being regarding the Application of Biology and Medicine[5], "dignity seems to have no meaning beyond what is implied by the principle of medical ethics" (Macklin 2003, p. 1419).

Responding to Macklin's (2003) argument that dignity is a useless term, Doris Schroeder (2008) argues that this claim is based on the Kantian[6] concept of dignity and does not distinguish this concept from others. In addition to the Kantian concept of dignity, which considers dignity as an intrinsic value human beings have because of their capacity for rationality, she argues that there are three other ways we can consider dignity: aristocratic, meritorious and comportment. Aristocratic dignity refers to dignity as "an outwardly displayed quality of a human being who acts in accordance with her superior rank and position"; comportment dignity refers to "the outwardly displayed quality of a human being who acts in accordance with society's expectations of well-mannered demeanour and

[4] The preamble of the Universal Declaration of Human Rights states that "Whereas recognition of the inherent dignity and of the equal and inalienable rights of all members of the human family is the foundation of freedom, justice and peace in the world" (see https://www.un.org/en/about-us/universal-declaration-of-human-rights)

[5] For this report by the Council of Europe, see https://rm.coe.int/168007cf98

[6] I discuss the Kantian conception of dignity in detail in the next section.

bearing" and meritorious dignity refers to dignity as "a virtue, which subsumes the four cardinal virtues and one's sense of self-worth" (Schroeder 2008, pp. 233–235). Schroeder (2008, p. 236) argues that considering dignity in this regard shows that "[d]ignity is more than respect for autonomy, and an analysis of dignity concepts has the potential to illuminate debates as long as one does not expect dignity to have only one, clearly delineated meaning". Disambiguating dignity in this way, Schroeder (2008) also resolves other contradictory ways dignity is applied, such as how two opposing movements, for example, death with dignity and anti-euthanasia, can both appeal to the concept of dignity to justify arguments in favour of and against assisted suicide. In this example, those who are opposed to euthanasia invoke the Kantian concept of dignity, which is considered an inviolable value that all have to an equal extent. According to this view, "the purposeful acceleration of death of an intrinsically valuable being is contrary to this being's intrinsic dignity and must therefore be prevented" (Schroeder 2008, p. 236). However, those who advocate for death with dignity invoke meritorious and comportment dignity. They argue that:

> What one might have tried all one's life, to fit into society's standards of decent behaviour, one might not be able to achieve in death. And to have this witnessed by others on whom one is dependent leads to one's perception of lost comportment dignity. At the same time, unbearable pain and extreme anxiety may undermine the cardinal virtue of wisdom and, thereby, the associated cardinal virtues of courage, justice, and temperance.

Another example of Considering dignity this way asks why people of different statures would seem to have equal dignity and different dignity at the same time. For example, how can we appeal to dignity in the same way for a humanitarian like Nelson Mandela or a serial killer like Ted Bundy? For this dilemma, Schroeder points to the idea that international instruments, such as the Universal Declaration of Human Rights, provide

dignity for *all* human beings[7]. In this context, we would be considering rapists or murderers to have the same kind or amount of dignity as someone considered a hero like Nelson Mandela or Mother Teresa. Here, the response is that all humans with rational capacity have dignity in the Kantian sense. This form of dignity is equal and inviolable, meaning that a criminal would have it in the same sense a saint would, whether the difference between the saint and the criminal would lie in the other forms of dignity where their granting is contingent on the behaviour of the stature of the person for whom it is granted.

In summary, although it is the case that all humans have inherent value, some, such as Nelson Mandela, their actions in society warrant other forms of dignity. For example, Nelson Mandela gained a level of veneration from others in his struggles for human rights and justice. His actions have set him apart from murderers and rapists, who have intrinsic value because of their capacities, but lack virtues or act in ways that are contrary to their 'society's values. So, the way that Mandela is perceived will be different from that of someone who is a convicted criminal. Thus, Schroeder (2008) concludes that considering dignity as having different conceptions resolves the contradictions that have been raised above and places dignity not only as useful but also as a compelling concept in moral and political debates on human nature.

Building on Schroeder's argument that dignity has four concepts, Suzy Killmister (2010, p. 160) argues that although distinguishing dignity into four concepts is a useful tool, "the usefulness of dignity as a guiding principle in medical ethics can be much improved by

[7] This idea is invoked in the Universal Declaration of Human Rights (as quoted above); the United Nations Charter (1945) which states that the peoples of the united nations "reaffirm faith in fundamental human rights, in the dignity and worth of the human person, in the equal rights of men and women and of nations large and small", and the International Covenant on Civil and Political Rights (1966) which states that, "recognition of the inherent dignity and of the equal and inalienable rights of all members of the human family is the foundation of freedom, justice and peace in the world".

identifying the single conceptual link that ties together the various values flying under its banner". Finding this link would be a way to reconnect the conceptual strands. She argues that this can be done by defining dignity as a *capacity*. She (ibid., p. 162) argues that "dignity is the inherent capacity for upholding one's principles". This means that because everyone has the capacity to uphold their principles, although not everyone might be able to act on them, they are guaranteed dignity.

Considering Schroeder's four concepts of dignity, Suzy Killmister (2010) first categorises them into inviolable and aspirational dignity. Under inviolable dignity, she places Schroeder's (2008) reformulation of the Kantian conception of dignity. Killmister argues that this concept of dignity applies to all human persons equally and cannot be lost. Under aspirational dignity, she places the other three concepts of dignity (aristocratic, meritorious, and comportment). As a capacity, dignity would account for both inviolable and aspirational dignity. Killmister, however, finds an issue with having dignity in different forms because while inalienable dignity can inform ethical principles, it excludes others and aspirational dignity, as it varies from person to person, also excludes others. She (ibid. p. 162) argues that to resolve this issue and to understand dignity as both aspirational and inviolable, we must consider it as a capacity, i.e. "a latent potential" as opposed to an ability, i.e. an "immediate possibility of action". Such a link to the concept of dignity provides a harmonious way of ensuring upholding dignity in the healthcare sector.

I agree with these authors (Killmister 2010; Schroeder 2008) that the concept of dignity is not useless. In fact, it is a pervasive subject of conversation in almost all fields of enquiry in the humanities and related disciplines and a guiding principle (see Riley and Bos, n.d. and Mattson and Clark). This pervasiveness is evidence of how useful it is. It is the case, though, as evidenced even in this discussion, that the notion of dignity tends to hold different meanings for different people in different contexts. So, part of ensuring clarity is by making

an important distinction between a 'concept' and a 'conception'. I rely on Metz's (2012) distinction. He(ibid) writes:

> A conception of dignity is a *philosophical theory* of it, i.e., a comprehensive and basic principle that purports to entail that, and explain in virtue of what, things either have dignity or lack it. A conception of dignity aims to account for the 'underlying structure' of the myriad things with dignity by invoking as few properties as possible. The claims that beings have a dignity solely in virtue of, say, having a soul or being autonomous are different conceptions of dignity. In contrast, the concept of dignity is what these rival theories are about. The concept of dignity is that which makes a given theory one of dignity as opposed to something else, such as virtue or welfare (Metz 2012, p. 20).

The distinction between a concept and a conception of dignity is important because it makes significant strides to resolve the claim of 'vagueness' mentioned above. So, when Schroeder refers to a Kantian, aristocratic, meritorious or comportment, she speaks of different conceptions of dignity. Similarly, when Killmister (2010) refers to dignity as a capacity for upholding one's principle, she refers to Schroeder's conceptions of dignity. According to this distinction, a concept of dignity could be either a value, a capacity, a status, or an entitlement (Killmister 2010; Mattson and Clark 2011; Riley and Bos n.d.). Different conceptions of dignity would then emerge when different authors theorise about the concept of dignity. In that theorisation, they would ascribe different characteristics to their conceptions.

For example, Metz (2012, p. 19) defines dignity (the concept) as a "superlative non-instrumental value that deserves respectful treatment". Based on this concept of dignity, he

formulates conceptions of dignity based on African communitarian values[8]. Similarly, Ikuenobe comes up with his conception of dignity but accepts Metz's definition of the concept of dignity as a 'standard definition'. The point I am making here is that it is from the different characteristics or basis that different authors emphasise that multiple conceptions of dignity are formed. These characteristics would vary from lifeforce to rationality, communal disposition, sentience and relationality, to name a few.

Depending on the field in question, these conceptions would prioritise different characteristics. For example, in law, the conception of dignity that would be prioritised would emphasise human rights, whereas, in warfare, the priority would be civilian or combatant protection. In bioethics, the priority might be patient care. Sharkey (2019, p. 81) points out that "in health care, … concerns about dignity are often related to bodily functions, access to toileting facilities and being addressed respectfully. In a war zone, the interpretation of dignity is likely to be quite different, and overriding concerns about death, suffering, and stress will leave little room for worries about personal hygiene". This difference in priorities indicates how the attributions of dignity vary based on the context of use.

Thus, different concepts and related conceptions of dignity will exist, especially considering the ubiquitous application of the term. That is no reason for the concept to be rendered useless or recommended not to be used. What is essential is that the concept and the emanating conceptions are clarified in use. Also, in applying the concept, we must appreciate that varied conceptions exist, viewed from different lenses and, as such, different qualities are ascribed to them. In what follows, I seek to highlight how the concept of dignity is applied in the HRI domain and to clarify which concept of dignity I will be considering.

[8] I discuss Metz's conceptions of dignity in detail in chapter 3.

Within the HRI domain, there are claims that certain uses of AI technologies undermine human dignity. However, the concept of dignity seems to mean different things in different domains of application. For example, in care (eldercare), Sharkey (2014) argues for a conception of dignity as a *quality* that can be gained or lost. Using Nussbaum's Capabilities Approach to human dignity, she argues that certain uses of robots in elder care could enhance the dignity of the elders, while others could impede it. Ozlem Ulgen (2018) argues for dignity as both a status and value in the field of autonomous moral agency, specifically regarding autonomous weapons. Using the Kantian conception, she argues that using autonomous weapons would deny the right of equality of persons and would violate the duty not to harm others. Also, she argues that using autonomous weapons instead of human combatants debases human life and does not provide respectful treatment. In the same context, Christof Heyns (2017) argues for dignity as an *entitlement* that would be denied to those targeted by autonomous weapons instead of human combatants. He argues that using autonomous weapons denies those targeted the right to a dignified life.

Evidently, even within the HRI domain, there is more than one concept of dignity in use. I will focus on the concept that dignity is an *intrinsic human valu*e based on some property, typically a capacity[9]. This definition is the most common in philosophical discourse (see Kant 1998; Metz 2012; Molefe 2019). Humans could have this value for several reasons, such as the capacity for reasoning or being autonomous, forming meaningful relationships, or caring for others. These reasons would vary depending on whichever conception is applied. What is consistent is that most human beings have it, and having it ensures they have certain

[9] Another conception of dignity that has recently become popular is one made by Jeremy Waldron (2012), who argues that dignity is a rank or status people have because of their rights. Although popular in theory, this conception has not yet enjoyed popularity in the HRI domain.

entitlements and are treated with respect. This definition of dignity seems to be the most salient and widely used, at least in the domain of AI ethics.

The question emanating from this definition of human dignity is, what about humans that grants them this superlative value? Kant argues that humans have this value because of their capacity for rationality. Some challenge this notion, claiming that ascribing dignity only to those with rationality discriminates against many who lack it[10] (Schroeder 2008). Schroeder (2008) argues that given that the notion of human dignity is invoked in international instruments that account for every human being, dignity is an inviolable property applicable to human beings. Others argue that although rationality is an important feature that guarantees one dignity, it cannot be the only feature that does. Other features can include the ability to relate to others, contribute to harmonious living, or be made in the image of God. I believe that human beings have this value because of their dynamic qualities. They can have rationality, form bonds, and relate to others. All these features are why they have dignity.

In this section, I clarified that the claim that dignity is a vague concept emanates from the failure to differentiate between a concept and a conception of dignity. Doing so clarifies whether it is a definition, the theory, or the theories in question. I also posit that the definition of dignity I will be applying in my thesis views dignity as an intrinsic value that humans have because of their various characteristics. In the next section, I will explore the two most salient conceptions of dignity prevalent in the HRI domain - the Kantian conception and Martha Nussbaum's Capabilities approach. I will critically discuss these two approaches to ascertain whether one of them could be the most appropriate to evaluate AI technologies' impact on human dignity. I first turn to the Kantian conception of dignity.

[10] A detailed discussion on the Kantian conception and the argents is in the next section.

2.3. Kant's Conception of Human Dignity

Immanuel Kant's (1724-1804) conception of dignity is one of Western philosophy's most popular and used conceptions of dignity to date. In this section, I critically discuss Kant's conception of human dignity because it is still expressed in the HRI domain. The discussion will feature a summary of Kant's conception and arguments for and against it, generally and specifically within the domain. The discussion is essential because it will help the reader understand what the concept of dignity means for Kant and what conditions guarantee it. As it pertains to the HRI, I argue that Kant's conception in its original formulation would not be suitable for evaluating how certain AI technologies could undermine human dignity because it discriminates on the grounds of rationality and race, i.e., only those who have rationality have dignity, and only those who are white have rationality. However, certain features of his conception can be applicable in formulating the conception of dignity suitable for evaluating AI technologies.

In the *Groundwork of the Metaphysics of Morals* (1785), Immanuel Kant (1724-1804) argues that "[e]verything in nature works in accordance with laws" (4:412). For example, if one throws a ball in the air, the ball will return to the ground. The ball's behaviour is based on the law of nature (gravity) - that what goes up must come back down. This claim applies to every entity except for rational beings. For Kant, "[o]nly a rational being has the capacity to act in accordance with the representation of laws, that is, in accordance with principles, or has a will". Rational beings have "will", that is, practical reason "required for the derivation of actions from laws" (ibid.). This implies that, although laws of nature may hold, rational beings can consider, announce, or even denounce the principles according to which they act. Given that human beings have the capacity for rationality, they can act according to will. Thus, they do not act merely according to laws (of nature or any other kind) but according to their own conceptions of laws. In Kant's words, the human being is "subject only to laws

given by himself but still universal, and he is bound only to act in conformity to his own will" (4:432). Thus, rational beings are autonomous – they make laws and legislate them. Kant also separates rational beings from other beings. He argues that:

> Beings the existence of which rests not on our will but on nature, if they are beings without reason, still have only a relative worth, as means, and are therefore called things. Whereas rational beings are called persons because their nature already marks them out as an end in itself, that is, as something that may not be used merely as a means, and hence so far limits all choice (and is an object of respect) (4:432).

Essentially, any being without reason is a 'thing' and thus can be used solely as a means, i.e., that which contains the possibility of action towards an end of a person. On the other hand, rational beings cannot be used as mere means because they are ends in themselves. As rational beings, they have *absolute worth*. Tying this all together, Kant writes:

> In the kingdom of ends[11], everything has either a price or a dignity. What has a price can be replaced by something else as its equivalent, what on the other hand is raised above all price and therefore admits of no equivalent has a dignity. What is related to general human inclinations and needs has a market price, that which, even without presupposing a need, conforms with a certain taste, that is, with a delight in the mere purposeless play of our mental powers, has a fancy price but that which constitutes the condition under which alone something can be an end in itself has not merely a relative worth, that is, a price, but an inner worth, that is, dignity. Now, morality is the condition under which alone a rational being can be an end in itself since only through this is it

[11] Kant defines the kingdom of ends as "a systematic union of rational beings through common laws" (4:434).

> possible to be a law giving member in the kingdom of ends. Hence morality, and humanity insofar as it is capable of morality, is that which alone has dignity (4:435).

The passage above is important because it captures the essence of Kant's conception of dignity. It mainly captures the point that humanity has intrinsic value insofar as humans are capable of moral reasoning. Kant defines dignity as "an unconditional, incomparable worth" (4:436) afforded to rational beings because of their capacity for morality[12]. He later adds that "[a]utonomy is, therefore, the ground of the dignity of human nature" (4:436). Thus, human beings as morally rational and autonomous beings are ends in themselves and, for this reason, have an intrinsic worth known as dignity. For Kant, human beings as rational and autonomous beings are ends in themselves that possess an intrinsic worth known as dignity. Because of this intrinsic value, the only way they are to be treated is with respect.

Although Kant's conception of dignity is over 200 years old, it is still a relevant and prominent subject of conversation among theorists of dignity (Kerstein 2014; Matthias 2020; Schroeder 2008). It is prominent and relevant to the point that some Kantian scholars maintain that the notion of autonomy and dignity should be considered the central claim of Kant's ethics. Sunday Fasoro (2019, p. 89) argues that scholars such as Barbara Herman, Christine Korsgaard, Allen Wood and Thomas Hill Jr. hold this view. Schronecker & Schmidt (2018, p. 81) also points out that "for Kant, dignity and value are crucial, indispensable concepts".

[12] In the literature about Kant's conception of dignity in the HRI domain, many attribute the basis of Kant's conception of dignity being rationality and not the capacity for morality. This is understandable, given that the capacity for morality is made possible by one's capacity for reason. So, for the sake of consistency with others, I will refer to rationality as the basis of Kant's conception.

Although this is the case, like any other philosophical concept, Kant's conception of dignity is not immune to criticism. One such critic is Doris Schroeder (2008, 2012). She (2008, p. 233) argues that if Kant defines dignity as "a property of all *rational* beings", this, in essence, means that dignity is only for "those human beings who are morally self-legislative, who can distinguish what is morally right from what is morally wrong". Samuel Kerstein (2014) also raises the issue of how the Kantian notion of dignity seems to exclude those beings who seem not to have a rational nature, such as older people, people with mental conditions, children, and animals (see also Nussbaum (2008)). He (ibid. p.224) argues that "if these beings indeed fail to have rational nature, then they have a mere price, according to the (Kant's) account". The claim is motivated by Kant's insistence that "everything has either a price or a dignity"; only autonomous agents have dignity due to their rational nature. Thus, when one does not meet the necessary criteria for dignity (having a rational nature), they merely have a price. If this is the case, then it would mean that we are under no obligation, and our treatment of them must have moral constraints.

From the arguments above, it is evident that Kant's conception of dignity discriminates based on one's capacity for rationality. His conception discriminates against children, elders with mental disorders, and those born with cognitive disabilities due to their lack of rational capacities. It discriminates against children because they are yet to develop rational faculties, against elders in that, as they age, some tend to lose the capacity for rationality, and those with cognitive disabilities because they can never develop it. This kind of discrimination is problematic because it takes away a large chunk of the human population based on the lack of this capacity. Excluding them then begs the question, what moral dimensions are they relegated to if they are not persons? Bearing this in mind, it might not seem ideal to apply this conception to evaluate the impact on dignity caused by AI technologies, particularly in healthcare. Many robot technologies within the care paradigm

are meant as assistive technologies for older adults, children, and people with disabilities. Thus, one cannot apply a concept of dignity that does not consider them as entities under consideration. One would need a conception of dignity that considers these vulnerable groups worthy of this value, and Kant's conception does not.

Thomas Hill Jr. (2014, p. 218) tries to defend Kant by arguing that "the notion that ordinary competent adults' moral agents have dignity by virtue of their rational autonomy does not imply that *only* they have moral standing". The vital issue is not who has rational autonomy but whether those who do would be directed by the principles of morality in their treatment of those who do not. He (ibid.) argues that "these moral principles would require us to extend the kind of dignity to human beings who have lost or not yet developed moral capacities and also animals (despite what Kant says) to regard animals as more than mere means to our ends".

Although it is the case that our moral obligations would require us to treat everyone with dignity, this would not necessarily mean that, by treating them that way, they would have dignity under Kant's definition. If the necessary condition for one to have dignity is rational autonomy, then those who lack it do not have dignity, and by Kant's words, those who lack dignity have a price. By lacking rationality, they would be relegated to the realm of objects. As objects, they would not deserve any treatment except for the one that other agents deemed fit. Should they be treated well like Hill proposes, that would be of the agent's own accord, not because it would be required. Similarly, any ill-treatment of objects would not fall foul within the moral realm.

In her turn, Schroeder (2008, p. 233) tries to save Kant's notion by arguing that many international laws and state constitutions have "stressed the inviolability of human dignity and expanded its scope to include *all* human beings"; therefore, Kant's conception can be rephrased to represent this. She argues that the modern Kantian-inspired definition of human

dignity would be: "Dignity is an inviolable property of *all human beings*, which gives the possessor the right never to be treated simply as a means, but always at the same time as an end" (ibid.).

This could be another reason to consider Kant's conception. If we considered it this way, it would definitely be more inclusive. The challenge is that Kant stresses rationality as a basis for dignity. Extending his definition to cater to all human beings, one would have to account for the basis of this conception. If it is not rationality that grounds human dignity, then what is? Even in the international instruments from which the reformulation emanates, the basis of human dignity is not articulated.

One other issue often raised with Kant's conception of dignity is that it also discriminates on the grounds of race. In his earlier works, Kant argues that the notion of rational nature is one relegated to a specific race of people. Immanuel Eze (2002, p. 438) raises this point, arguing that Kant, in his several texts, "outlines the geographical and psychological (moral) classification of humans". The texts in question are "A*nthropology from a pragmatic point of view. Physische Geographie, Conjectural beginnings of human history (1875),* Bestimmung des Begriffs einer Menschenrassen, On the varieties of the different races of man (1775), and *Observations on the feeling of the beautiful and sublime (1764)*" (Eze 2002, 430).

From a geographic point of view, Kant argues that humans can be distinguished according to race based on their skin colour and can be distributed geographically: white (European), yellow (Asian), black (African) and red (American Indians). Accordingly, different customs also emanate because of geographic differences. For example, "it is customary to permit theft in Africa, or desert children in China or to bury them alive in Brazil or for Eskimos to strangle them" (ibid., p.439). From a moral perspective, such actions are based on "unreflective mores and customs, natural impulses (or the inclinations to evil), or

commands of authority, lack of ethical principles and therefore not properly (i.e., essentially) human" (ibid.). Essentially, it seems Kant argues that the reason why other races act in such derogatory ways is that they lack a rational nature and morality. Eze (2002, p. 450) concludes that "it is clear that what Kant settled upon as the 'essence' of humanity, that which one ought to become in order to deserve human dignity, sounds very much like Kant himself: 'white', European and male". According to Kant, only the white race has such reflective attributes because they are from Europe (see also Tshivhase 2021).

Now, one might ask why I would bring up Kant's theory of race while the subject of this paper is dignity. This is because Kant's writings have a logical order in which his arguments flow and follow. So, what he has done in his earlier works, especially on the relationship between race and rationality, feeds his later concepts such as dignity, autonomy, duty, and freedom. In his earlier anthropologic work, Kant argues about what constitutes a rational being. For him, according to Eze, what constitutes that is one's skin colour and geographic location. So, a rational being, according to Kant, is meant to be white and from Europe. Kant's many theories, such as autonomy, dignity, and respect, rely primarily on the notion of rationality. We cannot just assume that when he speaks of rational beings, he means everyone when he explicitly states that he does not. Thus, this discrimination of who is rational does not only exclude older people and children. It also excludes any other race that is not white and from Europe.

On the issue of racial discrimination, I agree with Tshivhase (2021) and Eze (2002) that Kant's conception of race is offensive in that it excludes Black people from moral consideration and, thus, the value of dignity purely based on the colour of their skins and their geographic locations. A conception that discriminates on racial grounds is problematic because it dehumanises a large body of people and excludes them from the moral realm. Thus, Kant's conception of dignity would not be suitable for evaluating the relationship

between robots and dignity because it is selective regarding who is worthy of moral consideration and who is not. It is not like the challenges that care robots cater for only exist for white people. These challenges exist across all humans, and robots developed in this regard are meant for everyone. Thus, the concept of dignity that is used to evaluate them should not be exclusionary.

Many other arguments exist in favour of (Hill 1992, 2014; Korsgaard 1996; Schönecker and Schmidt 2018) and against (Kerstein 2013, 2014; Schroeder 2008, 2010; Sensen 2009) Kant's notion of dignity. This thesis will not endeavour to elucidate and respond to all of them. It aims only to illustrate the appropriateness of considering Kant's conception of dignity in evaluating the relationship between AI applications and human dignity within the HRI domain. Given the arguments stated above, it is evident that the Kantian conception of dignity might not be the best to consider within the healthcare domain, especially within the care paradigm.

Although that is the case, it appears that the concept finds expression in other domains, such as autonomous moral agency, specifically in the case of autonomous weapons by, for instance, Ulgen (2018). Using Kant's two conceptual strands of human dignity: "human dignity as a status entailing rights and duties; and human dignity as respectful treatment" (Ulgen 2018, p. 1), Ulgen (ibid.) makes two arguments: Firstly, that the "use of autonomous weapons denies the right of equality of persons and diminishes the duty not to harm others, [and secondly that] replacing human combatants with autonomous weapons debases human life and does not provide respectful treatment". The first argument implies that by replacing combatants with autonomous weapons, the value of the combatants in hostilities is debased since others are removed from physical risk. Ulgen bases the second argument on Kant's principle of humanity and argues that using "[a]utonomous weapons would devalue humanity by treating humans as disposable inanimate objects rather than ends

with intrinsic value and rational thinking capacity" (ibid., p. 10). By such treatment, the respect of those being targeted would be undermined. Similarly, Christof Heyns (2017) argues that:

> A central thrust of the notion of human dignity is the idea that humans should not be treated as something similar to an object that simply has an instrumental value (as is the case, e.g., with slavery or rape) or no value at all (as with many massacres). The person against whom the force is directed by autonomous weapons is reduced to being an object that has to be destroyed, and that is even more clearly the case where incidental casualties are at stake. They have no avenue, futile or not, of appealing to the humanity of the enemy, or hoping their humanity will play a role because it is a machine on the other side.

By subjecting humans to autonomous weapons, they are reduced to objects that could be used to fulfil a particular goal. Their lives are not respected, and their value is viewed as instrumental instead of intrinsic. Other authors also pick up on this idea of respect for persons. Goose and Wareham (2016, p. 31) argue that "fully autonomous weapons would also undermine human dignity because as inanimate machines they could not understand or respect the value of life, yet they would have the power to determine when to take it away". I discuss these arguments in greater detail in Chapter five.

Despite my reservations about the Kantian conception of dignity, I cannot deny its relevance and ubiquitous use inside and outside the HRI domain. Also, some features such as dignity as an intrinsic value, autonomy of persons, persons with dignity deserving respectful treatment, and subject to certain entitlements/rights are most cited when referring to it. In the same vein, the challenges levelled against the conception cannot be denied. A plausible conception suitable for evaluating the impact AI technologies have on dignity must be

inclusive. Kant's conception excludes more people than usual. Schroeder's reformulation seems to resolve this issue, except that it does not provide the basis for its ascription. It just takes for granted that such a definition is already in use. For these reasons, neither conception in their original formulation seems suitable to apply in my thesis.

A few main takeaways worth considering are, firstly, Kant's conception of dignity is discriminatory to rationality and race. Placing primacy on rationality discriminates those who lack the capacity, have lost the capacity or are still to have it, such as children, elders, and those with cognitive disabilities, from consideration of human dignity. On race, his earlier anthropological work claims that only European whites have rationality, thus removing other humans of different races from moral consideration and dignity. I argued that his conception is thus unsuitable for applying within the HRI domain, given that many AI technologies within this domain fall under the ambit of care and aim to cater to these vulnerable groups. I also note that, although his entire conception might not be plausible, certain features can be extrapolated and applied when conceptualising a suitable conception of dignity within the HRI domain.

Another conception of human dignity that finds resonance in the HRI domain is Martha Nussbaum's conception, specifically as it pertains to her Capabilities Approach (CA). The following section will discuss Nussbaum's conception of dignity and how it relates to her capabilities approach.

2.4. Nussbaum's Conception of Dignity

In this section, I will critically discuss Nussbaum's conception of dignity and its relation to the capabilities approach. The motivation behind this is that in the HRI, the CA is invoked as a framework from which AI technologies can be evaluated in general (see Borenstein and Pearson 2010; Coeckelbergh 2010; Sharkey 2014; Vallor 2011), and as it pertains to dignity

in particular (see Coeckelbergh 2016; Sharkey 2014). I argue that although the CA resolves some major issues levelled against the Kantian conception, it is still not suitable for evaluating the impact AI technologies could have on human dignity in an African context for two reasons. First, the scope of dignity is too wide; it allows for granting dignity even to animals, something that would not be acceptable in African normative frameworks. Secondly, its overemphasis on individuality makes it an unattractive concept for evaluation in the African context, which prizes community and relationality over individuality.

Martha Nussbaum explains her conception of dignity in a text titled 'Human Dignity and Political Entitlements' (2008). In this paper, Nussbaum (2008, p. 357) argues that "human beings have a worth that is indeed inalienable because of their capacities for various forms of activity and striving". She argues that dignity goes beyond the capacity of rationality. Other capacities include sentience, imagination, emotion, or the capacity for reason, what She (2011, p. 24) *basic capabilities* - "innate faculties of the person that make later development and training possible". These faculties differ from person to person. They are also dependent on the world for their full development and for their conversion into actual functioning.

Regarding capabilities, Nussbaum (2011) developed a theory of justice based on the CA known as the partial theory of justice. In her theory, Nussbaum (2011, p. 18) defines the CA as "an approach to comparative quality-of-life assessment and to theorising about basic social justice. It holds that the question to ask when comparing societies and assessing them for their basic decency or justice is what each person can do or be?". These questions do not necessarily mean that everyone can strictly do or be what they want; they must exist in an environment that enables them to have opportunities and act on them should they wish to do so. Nussbaum highlights the essential features that are associated with her approach. She writes that:

> The approach takes each person as an end, asking not just about the total average wellbeing but about the opportunities available to each person. It is focused on choice or freedom, holding that the crucial good societies should be promoting for their people is a set of opportunities, or substantial freedoms, which people then may or may not exercise in action: the choice is theirs. It thus commits itself to respect for people's powers of self-definition. The approach is resolutely pluralist about value: it holds that the capability achievements that are central for people are different in quality, not just in quantity; that they cannot, without distortion, be reduced to a single numerical scale; and that a fundamental part of understanding and producing them is understanding the specific nature of each. Finally, the approach is concerned with entrenched social justice and equality, especially capability failures that are the result of discrimination or marginalisation. It ascribes an urgent task to government and public policy – namely, to improve the quality of life for all people as defined by their capabilities (Nussbaum, 2011, pp. 18–19).

In addition to the central features of her approach, Nussbaum (2011) also employs other related concepts to her theory of justice. These are the notions of human dignity, threshold, political liberalism, and a list of central capabilities. Although all of these are important, her notion of dignity is particularly interesting in my context. Nussbaum argues that some living conditions can assure people a life worthy of dignity, while some do not ensure the chance to attain such a life.

Nussbaum (2008, 2011) argues that all persons must be accorded the means to ensure they reach a minimum threshold level required for a life worthy of dignity. For people with cognitive disabilities, for example, "the goal should be for them to have the same capabilities as "normal" people, even though some of those opportunities might have to be exercised

through a surrogate. The surrogate may, in some cases, supply the person with such a disability with part of the *internal capabilities,* that is, the "trained or developed traits and abilities, developed in most cases, in interaction with the social, economic, familial, and political environment" (Nussbaum 2011, p.21). If the person cannot develop sufficient choice capability on her own, for example, voting on that person's behalf even if the person cannot make a choice" (ibid.). These exceptions, however, only apply to those who have active striving. They do not apply to people in permanent vegetative states or anencephalic children because there is no way that their condition can ensure active striving.

Her conception of dignity is also closely related to the notion of respect and equality – beings with dignity deserve equal respect from others. She (2008, p. 359) argues that Kant's conception was wrong in assuming that respect only requires reverence for one's inherent worth. Respect "requires creating the conditions in which capacities can develop and unfold themselves" (Nussbaum 2008, p. 359). Thus, her conception of human dignity combines the Kantian argument that all humans are ends in themselves with Aristotle's notion of *eudaimonia,* that human beings' basic capabilities must be married with active striving to ensure that they can reach their functioning. She argues that "a life that does not contain opportunities for the development and exercise of the major human capacities is not a life worthy of dignity" (ibid.). She formulates a definition of dignity, which encompasses all the features. She (ibid. 363) argues that "full and equal dignity is possessed by any child of human parents who has any of an open-ended disjunction of basic capabilities for major life activities". This formulation includes a wide range of children and adults with severe mental disabilities, some of whom are capable of love and care but not reading or writing and are severely challenged in the area of social interaction.

Following this idea of dignity as an equal and intrinsic faculty of humans, Nussbaum (2011, p. 32) asks, "What does a life worthy of dignity require?". As a response to this

question, she composes a list of ten central capabilities that should serve as a minimum threshold for a life worthy of dignity, and these are: (quoting from Nussbaum 2011):

> 1)Life. Being able to live to the end of a life of normal length; not dying prematurely, or before one's life is so reduced as to be not worth living.
>
> 2) Bodily health. Being able to have good health, including reproductive health; to be adequately nourished; to have adequate shelter.
>
> 3) Bodily integrity. Being able to move freely from place to place; to be secure against violent assault, including sexual assault and domestic violence; having opportunities for sexual satisfaction and for choice in matters of reproduction.
>
> 4) Senses, imagination, and thought. Being able to use the senses, to imagine, think and reason—and to do these things in a 'truly human' way, a way informed and cultivated by an adequate education, including, but by no means limited to, literacy and basic mathematical and scientific training. Being able to use imagination and thought in connection with experiencing and producing works and events of one's own choice, religious, literary, musical, and so forth. Being able to use one's mind in ways protected by guarantees of freedom of expression with respect to both political and artistic speech, and freedom of religious exercise. Being able to have pleasurable experiences and to avoid non beneficial pain.
>
> 5)Emotions. Being able to have attachments to things and people outside ourselves; to love those who love and care for us, to grieve at their absence; in general, to love, to grieve, to experience longing, gratitude, and justified anger. Not having one's emotional development blighted by fear and anxiety.

(Supporting this capability means supporting forms of human association that can be shown to be crucial in their development).

6) Practical reason. Being able to form a conception of the good and to engage in critical reflection about the planning of one's life.

7) Affiliation. (A) Being able to live with and toward others, to recognise and show concern for other human beings, to engage in various forms of social interaction; to be able to imagine the situation of another. (B) Having the social bases of self-respect and nonhumiliating; being able to be treated as a dignified being whose worth is equal to that of others. This entails provisions of non-discrimination on the basis of race, sex, sexual orientation, ethnicity, caste, religion, national origin.

8) Other species. Being able to live with concern for and in relation to animals, plants and the world of nature.

9) Play. Being able to laugh, to play, to enjoy recreational activities.

10) Control over one's environment. (A) Political. Being able to participate effectively in political choices that govern one's life, having the right of political participation, protections of free speech and association. (B) Material. Being able to hold property (both land and movable goods) and having property rights on an equal basis with others; having the right to seek employment on an equal basis with others; having the freedom from unwarranted search and seizure. In work, being able to work as a human being, exercising practical reason and entering into meaningful relationships of mutual recognition with other workers. (Nussbaum 2011, pp 33-34)

According to Nusbaum (2008, p.351), these ten capabilities are the necessary conditions for a life worthy of dignity and, thus, the bare minimum that every individual should be entitled to for them to have and live a life worthy of dignity. One of the central characteristics of these capabilities is that they "belong first and foremost to individual persons and only derivatively to groups" (Nussbaum 2011, p. 35). She insists that this focus cannot be taken away by pointing out apparent facts that people belong to communities, be it as a family, a member of an ethnic group, or a state. Priority must be placed on individuals to ensure these capabilities are met. Another characteristic of these capabilities is that they are "heterogeneous and irreducible" (ibid.). They are irreducible in that no capability can be derived from any other. They are heterogeneous in that in providing for these capabilities, a state cannot implement a one-size-fits-all approach to provide for such capabilities. The level at which any capabilities should be realised must vary based on individual needs. It is important that all central capabilities are realised to ensure that people are guaranteed a life worthy of dignity.

Some find issues with Nussbaum's conception of human dignity, arguing that her conception lacks sufficient theoretical elaboration (Claassen 2014; Formosa and Mackenzie 2014). Rutger Claassen (2014) asks what role the notion of dignity plays in her capabilities list, considering that they were formulated way before the concept was introduced in her work. According to this claim, it appears that one way to accept it is to argue that the notion of dignity was implied earlier and only stated later. He (ibid., p. 245) also argues that Nussbaum's theory is different (compared to other conceptions of dignity like rationality) "only because it happens to defend a broader set of capabilities, going beyond our rational capacities".

Similarly, Paul Formosa and Catriona Mackenzie point out that, although human dignity plays a key role in Nussbaum's conception, some of the main comparisons Nussbaum

makes between her conception and the Kantian conception are overstated and that there seems to be no significant difference between the two conceptions (Formosa and Mackenzie 2014, p. 891). Essentially, on this view, whatever Nussbaum deems fundamentally different between her conception and Kant's is not that different.

I agree with the view that Nussbaum's conception is not fundamentally different to Kant's conception. However, I do not think this is a view that she refutes. She (2008), in fact, states that she agrees with Kant's view except for its insistence on rationality being the *only* human faculty that guarantees dignity. Extending the scope of Kant's conception to look beyond rationality as the only basis for dignity does a great job of including a larger population than Kant's original conception. I am in full agreement with this move.

Also, arguing that we must not only have dignity but live a life worthy of dignity ensures prioritising active striving. This ensures that people's inherent value can translate into more than a mere capacity, to being an opportunity that a person can fulfil should they wish to. Ensuring that an environment promotes the development of basic capabilities empowers people from just having dignity to living a life worthy of dignity. I thus argue that the CA does a great job in resolving the discriminations that Kant's conception has, thus extending the scope of who is worthy of dignity.

The challenge I find with Nussbaum's conception is that its scope is so broad that it would not only accommodate human beings, but given that rats also have sentience and can be construed as behaving in a goal-driven in Nussbaum's articulation, they would by this formulation also have dignity. Although this is the case with her conception, it does not necessarily mean that they have dignity in the same way as human beings do. For her, each animal's nature has its own dignity. She (2006) writes that:

> Each form of life is worthy of respect, and it is a problem of justice when a creature does not have the opportunity to unfold its (valuable) power, to flourish in its own way, and to lead a life with dignity. The fact that so many animals never get to move around, enjoy the air, exchange affection with other members of their kind — all that is a waste and a tragedy, and it is not a life in keeping with the dignity of such creatures.(Nussbaum 2006, p. 4).

I agree with the idea that each animal is worthy of respect because they have characteristics that are worthy of moral consideration, albeit to a varying degree. These characteristics grant them certain protections. For example, one should not kill wild animals for trophy purposes or forcefully domesticate animals that are in the wild. The harm that befalls animals must be morally justifiable. This, however, should not grant them dignity. It can, at the most, account for the fact that they have some moral status that warrants them some humane treatment, but this cannot be considered dignity.

The other issue with Nussbaum's conception is that it prioritises individuals over communities. In her conception, communities are only considered relative to the individual. This is a challenge I hold against most Western normative conceptions (including Kant's and Laitinen et al.). Although these normative values are prevalent in many societies, they do not represent all societies. Some societies, such as many in Africa and Asia, value the community over individuals. In this regard, I agree with Mattson and Clarks (2011, p.306) that "although the Western conception of human dignity and its relation to human rights is embraced by individualist cultures of Europe, North America, and elsewhere, it is not widely accepted by people in communitarian cultures…Communitarian cultures tend to prize values such as duties and obligations as opposed to rights". Not all societies subscribe to Western individualist cultures, so conceptions that prize these values would be problematic in communitarian societies, such as most African societies. When evaluating the impact of AI

technologies on human dignity, the conceptions and values of those who would be affected must be taken into consideration.

Be that as it may, Nussbaum's CA still enjoys popularity and expression in many fields, including the HRI domain (Borenstein and Pearson 2010; Coeckelbergh 2010; A. Sharkey 2014; Vallor 2011). Jason Borenstein and Yvette Pearson (2010) argue that the CA can be used as a framework to evaluate the design and implementation of robot caregivers. They argue that this approach can assist designers and users in terms of the values that must be considered when designing such products. Doing this will ensure that the technologies improve disabled people's ability to interface effectively with their physical and social environments.

In responding to criticisms against the use of assistive technologies in healthcare, Mark Coeckelbergh (2010) proposes that the CA can be adopted as a set of principles to evaluate assistive technologies and healthcare. This proposition is essential in that adopting the list of capabilities, as stated by Nussbaum, would assist in preserving the dignity of all. This holds especially considering that the most vulnerable can be within the healthcare paradigm and need a framework to protect and preserve their dignity. Following Borenstein & Pearson (2010) and Coeckelbergh (2010), Shannon Vallor (2011) proposes that the CA can also be employed "to help us to conceptualise the goods internal to caring practices for caregivers, goods that might be lost should a person choose to surrender some or all of such practices to a carebot".

Additionally, Amanda Sharkey (2014, pp. 67–68) lists two strengths of the CA: firstly, that "the CA articulates the relationship between human rights and human dignity more clearly than other rights accounts. Secondly, that the CA "grounds rights claims in human birth and minimal agency, not rationality or any other property". Sharkey (2014) also applies the CA as the framework to assess the possible effects of eldercare robots on dignity.

She (ibid., p. 74) argues that "[s]ince the CA offers an account of what is required for a life worthy of human dignity, its use as a framework for evaluating the effects of robots on the dignity of older people equates to assessing the ways in which robots can increase or impede their access to the central capabilities". From her evaluation, she concludes that "the CA permits a valuable and balanced perspective on the relationship between robots and the dignity of older people, highlighting some positive consequences that could result from the careful deployment of certain forms of robotics and warning of potentially negative ones" (ibid.).

From the discussions above, it is evident that the CA is another good approach to evaluating AI technologies' impact on human dignity for several reasons. Firstly, it resolves the discrimination issues evident in Kant's conception and, in the process, extends the scope of dignity to many other humans, making it more inclusive. It also outlines a list of key capabilities that can be applied as a framework to measure whether someone's living conditions guarantee a life worthy of dignity. Given the strengths of this conception, it makes sense why it enjoys such prominence in the HRI domain.

In this section, I highlighted Nussbaum's conception of human dignity and showed how it relates to the capabilities approach. I argued that although Nussbaum's conception resolves some major issues levelled against the Kantian conception, it has its own problems that would make it unsuitable for evaluating the impact of AI technologies on human dignity in an African context. One of these issues is that its scope is too wide, and it provides dignity even for animals like mice. This would not be an acceptable conception because dignity in an African normative context is a value reserved for human beings. Another issue is that her conception prizes individuals over the community and relationality. It can, however, be more easily applied in the HRI domain than in the case of Kant for the reasons given above by writers such as Sharkey.

Thus far, I have critically discussed both Kant's and Nussbaum's conceptions of human dignity as the conceptions that are salient in the HRI domain. In comparison to Kant's conception, it is evident that there are many similarities between these conceptions. For instance, both these conceptions are based on capacity, i.e. an agent must have some capacities for them to have dignity. For Kant (1998), this capacity is rationality. For Nussbaum (2008), it is basic capabilities such as sentience, imagination, emotions, choice and reason. Another example is that both these conceptions are based on individualistic normative frameworks. By this, I mean that they place an individual as a moral arbiter and attach entitlements to them.

Where they deviate from one another is on the basis for which dignity is ascribed. For Kant, individuals have dignity because they have rationality and autonomy. For Nussbaum, rationality is not the only basis on which individuals have dignity. Dignity can be ascribed based on other basic capabilities such as sentience, imagination, emotion, and the capacity for reason. Extending the basis for dignity makes room for other individuals who are otherwise not catered to in Kant's conception. Her scope, however, seems too inclusive that it provides dignity even for other non-human animals, which could be construed to have the basic capabilities in question. I have discussed these issues at length above. What is important here is to show that Nussbaum's conception is able to cater to humans whose rationality could sometimes be put into question, such as adults with cognitive disabilities and infants. This is especially important in contexts where patients are in need of robotic assistance. In such contexts, her conception would be more desirable, like in the context of this thesis.

In addition, although both conceptions are strong in their different ways, they are not accessible to all societies. As I have shown in my concern for Nussbaum's conception, associated Western individualist conceptions tend to prioritise an individual over the collective. With that, values of autonomy, rationality and individual rights take precedence.

This in itself is not an issue; the issue is the claim that such values hold a universal claim. It is the case that, for example, every individual ought to have equal rights. However, it is not the case that such rights take precedence over other values in all societies. In most African and Asian societies, family, community and social harmony are often prioritised. With this, certain values, such as duty to the community and obligation towards the community, take precedence (Bell 2015; Gyekye 2002; MacIntyre 1984; Matolino 2009; Mattson and Clark 2011; Menkiti 1984). So, the individualist concepts that are prevalent in Kant's and Nussbaum's conceptions and those in the evaluations that follow them are not representative of all societies and their values. In the next chapter, I will discuss communitarian conceptions of dignity, which are salient in African societies. The aim is to identify or develop a conception that would be suitable for evaluating the impact AI technologies could have on human dignity in communitarian contexts.

2.5. Conclusion

In this chapter, I discussed the complexities associated with the concept of human dignity. It highlighted some key challenges with this conception within the HRI domain. Regarding the concept of dignity, it argued that although complexities exist, they do not devalue the nature and importance of the concept. The fact that it is debated so vigorously is indicative of its importance. The issue is the failure to discern between a concept and a conception of dignity. Discerning between the two provides the necessary clarifications on whether it is the term in question or the resulting theories about it.

Regarding Kant's conception, I argued that although Kant's conception is one of the most important and widely applied, one of its key challenges is that it bases dignity on rationality. This move excludes people who lack rationality (for example, people with severe cognitive disabilities), those who are yet to gain it (children) and those who have lost it (older people because of deteriorating cognitive capacities). I highlighted that this is a challenge,

especially in the healthcare domain, because the larger section of the population for whom the technologies are meant would be those who fall into these vulnerable groups. Most care robots are created to cater for them.

I have also shown how Nussbaum's conception of dignity avoids Kant's pitfall by extending the scope of people with dignity beyond just capacity for rationality. She (2008) argues that human beings have dignity because of several basic capabilities, such as sentience, imagination, and emotions, and not merely choice and reason. This argument resolves the controversies linked to Kant's conception and makes room for more people like children, people with disabilities and older persons. I highlighted the main issue I have with Nussbaum's account; firstly, her scope of dignity is so wide that it accommodates other species besides human beings. Given that rats also have sentience and can be construed as behaving in a goal-driven in Nussbaum's articulation, they would also have dignity. I agree with the idea that all animals are worthy of respect because of their morally worthy characteristics. This, however, does not grant them dignity, but it grants them a moral status.

I concluded that both these conceptions have their strengths and challenges and thus resonate in different domains of AI application, especially in societies that subscribe to individualistic normative frameworks. I, however, want to propose a conception of dignity that prizes other values, such as relationality or community, as another plausible conception to evaluate the impact AI technologies could have on dignity, especially in environments that do not prize an individualistic normative lens, e.g., African communal societies. This would be a novel contribution in this domain because the key evaluations of the role of human dignity in regulating the HRI and AWS domains have largely been based on Western normative frameworks. Taking this direction introduces a different paradigm to consider when evaluating this impact in an African context and other contexts that subscribe to a communitarian worldview. In what follows, I will explore African conceptions of dignity to

find one that could be suitable for evaluating the impact AI technologies could have on dignity in the domain of care robots and AI in the military from an African perspective.

CHAPTER 3: AFRICAN CONCEPTIONS OF HUMAN DIGNITY

3.1. Introduction

In this chapter, I will discuss conceptions of dignity that are salient in sub-Saharan Africa. The aim is to provide a plausible communitarian conception of human dignity as an alternative to Western individualistic conceptions that are at play in evaluating the impact AI technologies could have on human dignity. The primary motivation for this is that the conceptions of dignity that are prevalent in evaluating AI impacts on human dignity have some shortfalls. Kant's conception excludes all human beings that lack rationality. By basing his conception on rationality, he excludes all those who do not have it, such as children and people with cognitive disabilities.

On the other hand, Nussbaum's conception is too inclusive, and her conception grants dignity even to animals that might have 'basic capabilities' such as sentience or a sense of community. Additionally, these conceptions are based on a Western individualist normative framework. I discussed these conceptions in chapter two. Although both have differences, they frame morality from an individualist perspective. Given that not all societies imagine morality from this individualist perspective, I will explore communitarian conceptions of dignity that are prevalent in African normative thought in this chapter.

I will first briefly outline four conceptions of dignity that are salient in African normative thought: vitalist, communitarian, based on spiritual nature and based on personhood. Following this, I will discuss communitarian conceptions as argued by Metz (2021, 2012) and Ikuenobe (2018, 2016). This will be followed by a discussion and comparison of both conceptions and the Western conceptions discussed in chapter two. From this comparison, I will show that Metz's conception is more compelling than Ikuenobe's in

that it avoids the pitfalls levelled against Western conceptions without falling into pitfalls levelled against African conceptions. Its only challenge is that it excludes persons with a partial moral status from consideration of dignity. This is problematic because not all persons for whom these AI technologies are meant have a full moral status. So, excluding them would exclude them from the associated benefits reserved for persons with dignity. This exclusion in Metz's conception excludes people with severe cognitive disabilities and infants who do not yet have a full moral status. For this reason, Metz's original formulation does not offer a plausible alternative conception.

To resolve this, I reformulate Metz's conception to include human beings with the capacity for communal relationships but with a partial moral status. I argue that human beings with a partial moral status but with a capacity for communal relationships have dignity because they have a higher moral value than any other entity with a capacity for communal relationships as objects. Additionally, we *recognise* and *relate* with them more meaningfully than we do other entities because we are of the same species. Lastly, all human beings already have equal human rights and all their associated benefits. This reformulation, coupled with Metz's principles for right action, provides a strong conception to evaluate the impact AI technologies could have on dignity. I will use this concept to evaluate the impact of AI technologies in healthcare robots (chapter four) and autonomous weapons (chapter five). I begin by explaining the salient conceptions of dignity in African moral thought.

3.2. African Conceptions of Human Dignity

The concept of dignity suggests that beings with dignity have an intrinsic value. Having dignity means that beings with dignity ought to be treated with respect. This idea of dignity is common everywhere, including in African normative thought. The differences emerge in the different conceptions of dignity based on one thing or the other. Literature In Africa, normative thought suggests that vitality, community, spiritual nature, and personhood are

some of the popular bases on which human dignity is conceptualised (Metz 2012, 2014a; Molefe 2022). This chapter focuses on the communitarian conceptions. I will, however, briefly explain all of them for the benefit of the reader.

The vitalist conception argues that "our special worth is constituted by human life in some way" (Metz 2012, p. 23). For this reason, humans have dignity because they have a higher degree of *life force*, which is "a valuable, spiritual or invisible energy that inheres in physical or visible things", unlike other organisms such as plants and animals. For others who champion this view, (Bikopo and van Bogaert 2010; Metz 2014a; Mkhize 2008; Molefe 2019b).

Another conception of Dignity in African normative thought bases dignity on spiritual nature. In this conception, it is argued that "human beings have something akin to a soul, an immaterial substance that will survive the death of their body; they are the most special things on the planet and hence deserve respect in the form of universal entitlement to life, liberties and resources and the like" (Metz 2014, p.312). For others who argue for dignity from this perspective, (see Deng 2007; Ramose 1999; Wiredu 1996a)

The communal view is another perspective that is prevalent and arguably the most popular in African normative thought. This view argues that our communal nature makes us the most important beings in the world. Its popularity is because the prizing of the community is popular in African normative thought (Gyekye 2002, 2023; Ikuenobe 2016; Matolino 2009; Menkiti 1984; Metz 2012, 2021). According to Gyekye (2023) , the foundations of African morality are **humanistic**; that is, they originate from considerations of human welfare and interests, not from divine pronouncements. As a humanistic morality, African morality naturally positions itself as a 'social morality which is enjoined by social life itself'. A social morality contends that individuals naturally attach themselves to other individuals. As such, they are part of a community. They are naturally related to or oriented

toward other people and thus have relationships with them. Human beings' natural sociality or relationality prescribes a social ethic rather than an individualistic one, such as those prevalent in western normative ethics. Gyekye (2023) continues to argue that as opposed to Western individualist ethics, African social ethics recognises the importance of the values of mutual help, goodwill, and reciprocity as opposed to human self-sufficiency concerning talents and capacities. The social character of African morality requires that the individual member of the society, ever mindful of his interests, adjust those interests to the interests and needs of others. Therefore, individuals necessarily embedded in a human community have a *dual* moral responsibility for themselves and other community members with whom they share certain basic needs and interests.

Others underscore this importance of the community by highlighting the importance of the community over that of a person. For example, Ifeanyi Menkiti (1984, p. 171) argues, "the African view of man denies that persons can be defined by focusing on this or that physical or psychological characteristic of the lone individual". This statement is a direct challenge to many western ideas which prize individual capacities over that of the community. He (ibid.) continues to argue that "[r]ather man is defined by the environing community". Essentially, for Menkiti, the community is ontologically prior to the individual.

In response to Menkiti (1984), Kwame Gyekye (2002) admits to the idea of African societies being communitarian but challenges the idea of the ontological priorness of the community. He argues that it is the individuals who make up the community. As such, their rights Aare maintained even if their actions are not morally praiseworthy. He, however, emphasises that in African morality, duties trump rights. The natural relationality of the individual immediately involves them in some social and moral roles in the form of obligations, commitments, and duties (or responsibilities) to other members of their community, which the individual must fulfil. For other communitarian conceptions, see (M. O. Eze 2008; Matolino 2009).

Regarding the communitarian conception of dignity particularly, Metz (2012) and Molefe (2022) argue that there are two versions of community that are worth considering: the traditional and Metz's conception. The traditional view argues that "our dignity is constituted

by our existing relationships with others, including spiritual persons such as ancestors" (Metz 2012, p. 26). For example, Ikuenobe (2016) argues that dignity has to do with the moral use of one's capacity for the promotion of love, friendship, positive identity and active solidarity. Other proponents of this view include Tutu (1999) and Bujo (2001). The other view is developed by Metz (2012) as a way of coming up with a conception that does not consider the supernatural. In this conception, it is not a person's relationships with the community or ancestors that grant them dignity but their capacity to form said relationships.

Another conception of dignity located in African normative thought is one based on personhood and was recently introduced by Motsamai Molefe (2022). Molefe (ibid) argues that one of the conceptions that finds resonance in African philosophy is the African conception of personhood. Personhood in African philosophy refers to a human person behaving in a way that is considered morally praiseworthy. Because of this virtuous behaviour, they are ascribed status in their community(Gyekye 2002; Menkiti 1984; Molefe 2022). According to Molefe, his conception and that of Ikuenobe (2016) fall under this category. Ikuenobe (2016) bases his conception of dignity on personhood that dignity is ascribed based on the moral actions of a person to ensure communal harmony. Molefe (2022) argues that dignity is the capacity for virtues, i.e. a person having the ability to act in a praiseworthy manner. (Gyekye 2002)

These four conceptions of dignity are argued to be the most salient in African philosophy. I will not endeavour to discuss all these conceptions. I will pay attention to the community-based conceptions of dignity. This is because, firstly, they enjoy more popularity, given that most African societies are considered communal. Secondly, the other conceptions still primarily prize the individuals and their capacities, which is not so different to Western conceptions. Thirdly, I agree with Metz (2012, 2014a) that other conceptions will find it difficult to ascribe human rights. Human rights are important even in African normative

thought because although African societies are communal, most of their political formations are that of liberal democracies, making them sensitive to human rights. I will primarily discuss Metz's conception, which prioritises the individuals' capacity for communal relationships as opposed to community membership and Ikuenobe's conception, which prioritises moral actions as a way to earn dignity. I begin with Metz's conception.

3.3. Metz's Conceptions of Dignity

In this section, I discuss Metz's communitarian conception of human dignity, which is expressed in his various works (Metz 2012, 2014a, 2021). His view argues that human dignity is based on the capacity of humans to be part of or party to loving relationships. In the discussion, I will also highlight and respond to criticisms levelled against this conception. I will show that Metz's conception of dignity is comprehensive because it addresses some shortfalls identified in the Western conceptions as well as the communitarian conception identified here. In terms of the shortfalls of Western conceptions, Metz's conception is based on communal values of relationality. In terms of the communitarian conceptions discussed in the next section, other communitarian conceptions ascribe dignity on the grounds of community membership. The shortfall in question is that such communitarian conceptions do not account for individuals who are unable to or choose not to use their capacities. In this section, however, I will not endeavour to highlight said communitarian conceptions, as I take it for granted that Metz has conducted a comprehensive report of the said conceptions (Metz 2012, 2014a, 2021).

According to Metz (2012, pp. 25–26), "[o]ur communal nature makes us the most important beings in the world", which means that most African cultures place normative priorities on the community, with a communal account of dignity being one philosophical expression of such an orientation. However, this does not mean community membership is obvious and open to all. Certain conditions must be met before one qualifies to enter a

community. Metz (2012, p.26) argues that "the capacity to enter into a community with human beings grounded in African thought is well construed in terms of a combination of two logically distinct kinds of relationships, identity and solidarity". Identity refers to people thinking of themselves as "members of the same group – that is, to conceive of themselves as 'we', as well as for them to engage in joint projects, coordinating their behaviour to realise common ends" (ibid.) and sharing a way of life. Solidarity is when "people care about each other's quality of life" (ibid.) by engaging in mutual aid, acting in beneficial ways, and caring for one another. So, a community based on African normative values is based on identity and solidarity.

Metz (2021) provides two examples to highlight the appeal of a community grounded on both these relationships. Firstly, he argues that the combination of these two ideas captures how family members ought to engage; that is, families should enjoy relationships that exhibit a sense of togetherness, cooperation and collective striving for one another's good. The second appeal of this combination is by considering notions of love and friendship. Metz (2012, p. 26) argues that "a loving or caring relationship more or less is one in which parties think of themselves as 'we', engage in common activities, act to one another, and do so consequent to sympathy and for the others sake". Based on these examples, it is evident that a sought-after community that embraces African communal values embraces relationships with a combination of identity and solidarity. Metz constructs his communal conception of dignity from these ideas about the community.

According to Metz (2012, p. 27), "we have dignity in virtue of our capacity for loving relationships" (see also Metz 2021). This refers to one's capacity to be part of or be a party to these relationships instead of being in them. Merely having the capacity to be in one means one has a superlative non-instrumental value. For Metz (2021), there are two ways one can be part of or party to a communal relationship: either as a subject or as an object. He (ibid.,

p.107) writes that "[a] being can be a subject of communal relationship insofar as it can think of itself as a 'we', cooperate with others, help others, and act for their sake out of sympathy. In contrast, a being can be an object of such a relationship insofar as characteristic human beings could think of it as part of a 'we', advance its goals, benefit it, and act for its stake out of sympathy".

As a subject of communal relationships, a being has the capacity to befriend others. This requires identifying with others and exhibiting solidarity with them. As an object, the being is recognised by others as an entity that can be befriended, meaning that other subjects can identify with them and exhibit solidarity towards them. The disposition of an object does not require the entity in question to respond to another's advances for friendly relationships. Essentially, the capacity to be befriended does not require any action from the object. What is involved in being capable of being a subject or object of a communal relationship is simply "*being able* in principle, i.e., without changes to a thing's nature" (Metz 2021, p.107)). Essentially, a being can be a subject or object in a communal relationship if their moral status remains unchanged.

However, there are limitations regarding who can be subjects or objects of a communal relationship. There are what Metz (2021) calls' merely contingent inabilities', which are not permanent because they do not alter a being's moral status. These, for a subject, are if they are "asleep, having drunk too much alcohol, electing not to sympathise, being ignorant of what would benefit others, and having been put into solitary confinement" (ibid., p. 107). Although these states could hinder one's decision to befriend, they do not necessarily permanently alter the nature of a person. Thus, they cannot alter their moral status. At any point where these states could change, the entities continue to have the capacity to be part of communal relationships. For objects, contingent inabilities include "the fact that a given person is unaware of the being or is scared of it" (Metz 2021, p. 107).

The permanent limitation to being a subject in a communal relationship is if a being's "biological constitution has no concept of a mind distinct from its own or that cannot trade off its own ends or interests for the sake of those of others" (ibid.). By this, Metz means that a being cannot be a subject in a communal relationship if they are solipsistic, i.e., they fail to recognise that there are any other beings except for themselves. For objects, Metz (ibid.) argues that beings that lack "*goal-directed behaviour or sentience,* [for example] a rock, shrub, bacterium, or pen", cannot be objects of a communal relationship. This suggests that any entity that by its nature has goal-directed behaviour or sentience is capable of being an object in a communal relationship with us. By this logic, even animals with goal-directed behaviour or sentience can be considered objects of communal relationships.

According to Metz (2021, p.108), a principle of moral status that is based on major gradations of capacity entails that "many animals, such as cows, cats, owls, and mice, have a partial moral status" because we can, in principle, identify with and exhibit solidarity towards them. He argues that the more capacity one has to relate communally, the greater one's moral status. So, a being has a *full moral status* or dignity if they can be both a subject and object of communal relationships. A being that is capable of only being an object has *partial moral status*. An entity with partial moral status still matters for its own sake. However, for Metz, if there is a conflict of interest with an entity having full moral status, it should lose out since it is not as important from a moral point of view.

Additionally, for Metz, only entities with full moral status have dignity. This is because "those beings that in principle can be friendly and be befriended have a dignity, while those that in principle cannot be friendly but can be befriended have *a standing*, albeit one less than a dignity" (ibid., p. 152). Essentially, such a principle that is based on major gradations of capacity entails that an overwhelming majority of human persons have full

moral status. That way, they can all enjoy equal regard since they have equal capacity to commune.

In sum, our capacity to be part of or party to communal relationships based on identity and solidarity grants us a superlative non-instrumental value that warrants respectful treatment. One can have this capacity as a subject[13] of communal relationships. Important in this conception is that one has dignity by virtue of *having* the capacity to be a party to or part of a communal relationship, not by being in said relationship. Hence, being an object of such a relationship does not require any action on the part of the being. Respect is also owed by virtue of having such capacity. Respecting others can be in the form of respecting their human rights - such respect can be construed as respecting one's capacity to love and be loved. Violating others' human rights would mean degrading this capacity.

Having outlined Metz's conception, I now discuss criticisms against him, specifically by Ikuenobe. I will then respond to the criticism and offer my view of Metz's conception. Ikuenobe (2016) challenges Metz's communitarian conception of dignity on two grounds. Firstly, he challenges Metz's idea that dignity, in a communitarian sense, derives from having the capacity for loving relationships. For Ikuenobe (2016, p. 454), "dignity is not just mere capacity for harmonious communal relationships, but *the proper use and actualisation of one's capacity* as manifested in virtuous character and good behaviour within the context and help of the community". Essentially, although having the capacity to be part of or party to relationships is necessary for human dignity, it is not sufficient. One must actualise that capacity by acting in ways that guarantee that those relationships are developed and maintained. For Ikuenobe, capacities alone have no inherent moral worth; they are just

[13] A person who is a subject in a communal relations can also be an object of communal relationships but not all entities who are objects in a communal relation can be subjects.

instrumentally good. The moral worth of capacities depends on how they are used to promote the moral good of the community.

Ikuenobe (2016) also challenges the features that Metz ascribes to his concept of dignity. For Metz (2012), dignity is a superlative non-instrumental value devoid of social construction. It is typical of individual members of the human species and does not vary based on one's status. As a value, it warrants respectful treatment. Ikuenobe (2016, p. 455) argues that these features "represent the standard passive 'having dignity' view of dignity in Western discourse, which indicates only some aspects or use-senses; there are other use-senses that these features do not capture [such as] the active use of one's capacity in a communal context, the responsibility and desert aspects that engender respect, the degrees of respect, and the prescriptive and ascriptive or bestowal aspects". For Ikuenobe, all these features make up dignity in an African communal context. Moreover, according to him, Metz's features fail to explain how they connect with their active use and responsibility components. In essence, Ikuenobe argues that one only has dignity when one uses said capacities for harmonious relations. And this dignity is bestowed upon them as a result of their satisfactory actions. By basing dignity merely on capacity, Ikuenobe argues that Metz does not capture the essential qualities prominent in African conceptions.

Ikuenobe (2016) raises interesting points regarding Metz's (2012) conception. However, there are a few points I wish to make in defence of Metz's conception. Firstly, Ikuenobe points out that Metz's concept represents the standard 'having dignity' aspect of dignity prevalent in Western discourses and ignores the use and responsibility aspect prominent in communal discourses. I have two responses regarding this. Firstly, the definition of dignity does not vary much between the West and Africa. There is mostly a consensus on what a concept of dignity is. What Ikuenobe dubs the "standard 'having dignity' aspect of dignity" many tend to accept as a definition of the concept of dignity, i.e.,

dignity is an intrinsic value that humans have (see section 2.2). It is this value that guarantees human beings' respectful treatment. The other ideas, such as active use, responsibility, and prescriptive values, are well captured by talks of 'ubuntu', 'virtue', or 'personhood'. We need not use the term 'dignity' to make sense of it. Second, for Metz, one does not have to act in any way to have dignity. Merely having the capacity grants them dignity. So, what Ikuenobe argues to be the use and responsibility values he places on dignity Metz deals with them as different conceptions such as 'right action'.

Although Metz does not prescribe action as the grounds for dignity ascription, he does encourage it as an important feature in successful communal relationships. In his Afro-communitarian conception, as stated already, Metz emphasises that dignity is the capacity to be part of or party to friendly relationships based on identity and solidarity. In defining solidarity, Metz (2012, p. 25) emphasises that solidarity means "engaging in mutual aid, acting in ways that are expected to benefit each other, and caring". This definition of solidarity highlights that forming a loving relationship requires action, hence the prevalence of the verbs' caring', 'engaging' and 'acting' – in the definition. So, although Metz's Afro-communitarian conception of dignity is based on merely having a capacity, if one enters a communal relationship, one will have to act in a way that illustrates solidarity with the persons they identify with, not to have dignity but to respect their dignity.

In the Afro-communitarian sense, there is no dispute that dignity requires action, but the action is a matter of rightness or justice, not dignity. One must act to ensure harmony in a community and respect others' dignity. Metz might not state it outright in his definition, but his articulation of solidarity does highlight the necessity of acting in a way that could be construed as friendly. The difference would be that one does not lose dignity due to inaction. They might lose favour in the community and the associated benefits, but having the capacity to be a party to relationships of identity and solidarity is sufficient to warrant their dignity.

This point highlights that inaction would not necessarily revoke one's dignity, although it might lose them favour in a community. So, action is encouraged, and there are benefits for acting, but one has dignity whether one acts or not.

Suppose we accept Ikuenobe's suggestion that dignity entails the active use of one's capacity for the betterment of the community. This notion would suggest that only those belonging to a community have dignity. For one's action to be recognised, they would have to be recognised as members of a community in which those actions are committed. Metz (2021, 2012) is critical of this conception of dignity because grounding dignity based on community membership means that those not members of the community, e.g., a hermit or those incarcerated and in solitary confinement, would not have it.

Regarding a prisoner in solitary confinement, Ikuenobe (2016) argues that his ideas on imprisonment may respond to this argument. Regarding imprisonment, he argues that it would be justifiable not to treat prisoners and others who violate the values of a community with respect because dignity and respect are not unconditional. He (2016, p.462) argues that "a community must evaluate and determine the degree of respect that is earned and deserved, and the extent of responsibility that people ought to have toward an individual, which is conditional on the quality of his moral behaviour and virtuous character". This implies that those who have transgressed against the community have diminished their dignity through anti-social behaviour. Thus, the community can judge them in terms of how they should be treated. In this regard, a community can deny an agent a degree of respect if they deem their actions as contrary to the community's values.

Similarly, the community can award dignity if people's actions are favourable and aligned with the idea of holding people accountable for their actions. Ikuenobe (2016, p.464) emphasises that "individuals' moral dignity and respect by others are partly dependent on how they use their capacities to care for our neighbours". In essence, given that someone

ideally ends up in prison due to their actions and would have caused disharmony in their respective communities, it is justified to violate or mistreat them.

I am not convinced that a community should have such a right. As I have argued in section 2.2, dignity is an intrinsic value humans have for one reason or another. For that reason, dignity is unconditional to all who have it. This is why it is the grounds on which human rights are based. Imprisonment suspends certain rights of the prisoner for a certain amount of time, such as liberty. This suspension of liberty due to someone being found guilty does not suggest any prohibition of other rights, such as the right to life. That is why no one has the right to kill prisoners, even in times of war. Since all rights are grounded on human beings having dignity, suspending certain rights does not mean that one's dignity has been lowered.

Because of prisoners' inherent dignity, there are standards in terms of their treatment. For example, the South African Constitution compels the Department of Correctional Services (DCS) to comply with certain sections of the Bill of Rights in terms of the treatment of prisoners. Under the DCS mandate, prisoners have the following rights: equality, human dignity[14], freedom and security of a person, freedom of religion, health care services, children, education, and humane treatment[15] . There is even a running joke that, in South Africa, one must be a convicted prisoner to get free education. Prisoners cannot be ill-treated because their actions have wronged the community. So, Ikuenobe's ideas about imprisonment do not address the question of the dignity of a prisoner in solitary confinement. According to his logic, such prisoners lack dignity as they cannot be part of a community and thus act in a way that could ensure harmony. By Metz's (and my) logic, prisoners in solitary confinement

[14] In the south African constitution dignity is both the foundation for all human rights and a right (see section 7 and 9 of the constitution)

[15] See the Department of correctional services website: http://www.dcs.gov.za/?page_id=172

have dignity because this temporary state does not alter their moral status. They maintain the capacity to be part of a communal relationship as both subjects and objects.

Regarding a hermit, Ikuenobe's (2016, p. 465) response is that "it depends on his degree of seclusion from society": whether he voluntarily made a choice to be a hermit and for what purpose – say religion or whether the hermit exhibits some form of self-respect and respect for others. If a hermit exhibits self-respect or respect for others, they have a sense of moral dignity that warrants a degree of respect from others. Suppose they are a hermit by virtue of other reasons, such as a mental illness. In that case, they deserve unconditional respect granted to those who cannot use their capacity for harmonious living.

Ikuenobe's response does not explain how the choice of being a hermit affects one's dignity. He also does not explain how a hermit would exhibit respect for others. How can we judge one's respect for others, or how could they even act respectfully towards others if they exist in solitude? A hermit does not engage with others meaningfully to determine whether to respect others. Like the example of a prisoner in solitary confinement, Ikuenobe's logic suggests that a hermit would lack dignity since they are not using their capacity for communal harmony. In contrast, Metz's (and my) logic would argue that a hermit, having the capacity to commune with others, has dignity.

From the two examples, it is evident that Ikuenobe does not make provisions for those who have capacities but do not apply them (see also (Molefe 2019a). By his logic, it is evident that those who have capacities but do not or are unable to use them lack dignity and thus do not deserve respectful treatment.

Thus far, I have responded to Ikuenobe's (2016) criticism of Metz's communitarian conception of dignity. I have argued that Ikuenobe's criticism of Metz's characteristics of dignity does not pay attention to the distinction Metz makes between a concept and a conception of dignity. This distinction shows that what Ikuenobe (2016, p. 455) considers the

"standard passive 'having dignity' view of dignity in Western discourse" is common in many places, including Africa. It is from this common view that multiple conceptions emerge. I have also shown his (ibid.) idea that dignity is beyond having capacity, but rather the active use of said capacities excludes those with capacities but are unable to or do not want to use them, such as a hermit or a prisoner in solitary confinement. It also fails to account for Metz's notion of personhood or virtue, which encourages 'right action' from subjects of communal relationships to respect others' dignity. In what follows, I will now discuss my views on Metz's conception, some of which are critical and call for revisions to it.

Firstly, I agree with the idea of dignity based on a capacity as opposed to any action. This makes it much more inclusive than if it prescribed an action that grants it. I also agree with the basic intuitions of a gradational approach to moral status as it compares human beings' value to other animals and living organisms. Based on these intuitions, grounded on the capacity to relate communally with earthly moral agents, human beings morally matter more. So, when choosing between them and other animals, human beings would take priority. I also agree with the argument that we have the capacity to commune with entities other than human beings and hold them in high regard. People have pets and livestock, and our treatment of such entities exhibits a relationship of some sort. In many African traditions, livestock ownership, for example, is not only for subsistence but also a show of wealth and status in the community. Thus, people naturally tend to their livestock with care and love.

It becomes tricky when the comparison is amongst human beings. There is the idea in Metz's work that some human beings lack dignity because they have partial moral status and thus can commune only as objects. According to Metz, their partial moral status means they have moral worth, but it does not warrant their dignity. Such an approach excludes those in the margins, such as infants and those with severe mental incapacities, from dignity. Metz (2021, p.164) argues that "extremely autistic, psychopathic, and mentally incapacitated

human beings (even newborn babies) *lack a dignity* comparable to ours". In this regard, Metz's conception seems vulnerable to the same criticism I levelled against Kant's conception.

Metz excludes those lacking said capacities by ascribing dignity only to entities with certain cognitive capacities, such as an awareness of others as distinct from oneself. In Metz's conceptions, only persons with a full moral status have the capacity to be subjects of communal relations and are ascribed dignity on the grounds of that status. This status warrants respectful treatment. If only subjects of communal relations have dignity, what does this mean for other human beings that can be a party to communal relationships merely as objects? If it is from this intrinsic worth that subjects deserve respectful treatment, what form of treatment would befall those who lack dignity in this sense? Lastly, considering that human rights are based on dignity, what does it mean for the rights of human beings that lack dignity in this sense?

Regarding the treatment of those who are not subjects of communal relations, Metz responds that those with a capacity for communal relationships as objects have a significant moral worth despite lacking dignity. As such, we have an obligation to avoid treating them with degrading or disrespectful behaviour. For Metz (2021, p. 110), respectfully treating entities with a moral status would constitute 'right action', that is, actions that "do not merit negative reactions such as criticism, compensation, or punishment". He (ibid.) further elaborates that "there are direct duties towards a variety of entities, including not just human persons, but also humans who are not persons as well as many non-human animals". So, these duties spread across all entities with moral status. Ill-treatment of any such entity would be regarded as wrong and impermissible.

This response does well in resolving the discomfort related to the treatment of entities that lack dignity. However, there is still a limitation in terms of the treatment of those who

are not subjects of communal relationships. Metz (2021, pp. 107-108) argues that "[t]o have a partial moral status means that the being matters for its own sake, but if its urgent interests were to conflict with those of a being with full moral status, it should lose out since it is not as important from a moral point of view". This suggests that although they matter for their own sake, their interests can still be trumped if said interests conflict with those who are subjects. Another issue is that it still excludes them from the consideration of dignity, thus excluding them from certain entitlements associated with human dignity, such as human rights.

As already mentioned, human dignity is the basis for human rights in many international governing instruments, such as the Universal Declaration of Human Rights, the International Covenant on Civil and Political Rights, the African Charter on Human and People's Rights, and the South African Constitution. If dignity is the basis of human rights, then those who lack it would not have these rights. Metz argues that his conception accounts for human rights because most human persons have the capacity as subjects of communal relations and thus have dignity. This still excludes those who lack that capacity. More precisely, if human beings have human rights because they have dignity, and some people lack dignity, does it mean these people do not have human rights?

According to Metz (2021), there are certain extreme psychopaths or severely autistic individuals. Those who cannot be the subject of communal relationships but have ends and quality of life have rights not because they have dignity but because they can be objects of a communal relationship with us to a high degree, and so have a partial moral status (somewhat higher than animals). They include newborn infants and adults with permanent severe mental incapacities. Other humans, who cannot be the subject of communal relationships and do not have ends or quality of life (and hence cannot be objects of communal relationship), do not have rights. They include embryos and early-stage foetuses. Any moral reasons for treating

these beings one way or another are 'indirect'; they are not about how the treatment would affect these beings but rather other beings, such as those who care about them.

The explanations above satisfy my discomfort with Metz's conception's exclusion of other entities from dignity. Even with the exclusion, his conception does provide moral coverage to those entities that lack dignity. This makes it a more inclusive conception regarding the moral consideration of human beings, but it is still exclusive regarding who has dignity.

In this section, I discussed Metz's Afro-communitarian conception of dignity and defended it against criticisms levelled by Ikuenobe. Thus far, Metz's Afro-Communitarian conception of dignity seems to be winning as a comprehensive communal conception of dignity. In its current form, the challenge is that it excludes persons who are party to communal relationships as objects, such as newborn babies, extremely autistic, psychopathic and mentally incapacitated persons, from consideration of dignity. In what follows, I discuss Ikuenobe's conception of dignity. Having criticised Metz's conception, Ikuenobe comes up with his communal conception of dignity. He promises that his conception will address the shortfalls he had identified in Metz's conception. The following section outlines Ikuenobe's conception of dignity in a communal context.

3.4. Ikuenobe's Communal Conception of Dignity

In this section, I will discuss Ikuenobe's communal conception of dignity. It begins by outlining his notion of personhood as his basis for human dignity. Following this, it explains his conception of dignity and outlines criticisms levelled against it. I argue that his conception of dignity lacks the impartiality often associated with the notion of dignity. Basing dignity on personhood suggests that only those considered persons would have dignity, and thus, only they can be treated with respect.

Using Metz's (2012) communitarian conception of dignity as a foil, Ikuenobe (2016, p. 453) proposes a communal conception of dignity grounded on his conception of personhood. He argues that personhood in the communal sense is the positive recognition of one's behaviour in terms of how said behaviour enhances communal values of harmony, mutuality, and solidarity. His conception of dignity deviates from Metz's conception by arguing that dignity is beyond having the capacity for harmonious relationships but is instead constituted by acting to ensure this harmony. The conception of dignity that Ikuenobe proposes involves the notion of earned mutual respect based on one's behaviour. Unlike in Metz's case, one is not entitled to respect merely because of having dignity. Respect is earned based on a person's actions.

Regarding his conception of personhood, Ikuenobe (2016, p. 445) calls his conception a *robust* conception of personhood because it combines the "descriptive, biological, spiritual and metaphysical view" with the "normative, social, and evaluative view" of personhood. He argues that the normative and descriptive conceptions are two aspects of a robust conception of personhood, the descriptive view being the material conditions for the normative view. He (2016, p. 446) argues that personhood is "a function of communal recognition that involves the moral evaluation of one's character and actions based on communal values and obligations". From these bases, one can be ascribed or denied personhood or moral dignity and its ensuing respect. He (2016, pp. 446–447) writes:

> The statement that one 'is not a person', or 'has no moral dignity', is not meant to be descriptive, in terms of denying the natural, factual, metaphysical or psychological features of personhood or dignity. Rather, it is a moral ascription based on a normative judgment, which is that one has good character, behaved morally, acted or comported oneself properly, based on accepted values… [As such] the normative ascription of personhood assumes

> that the individual has the rational and cognitive capacities for agency and free choice (Ikuenobe 2016, p 446-447).

One is morally not a person solely because one has similar ontological capacities with other human beings. Human beings are persons insofar as they use their capacities virtuously and in line with communal values. To do this, humans must have the capacity for autonomy, which for Ikuenobe (ibid) means "the ability to learn, internalise, and act in ways that promote communalism and its values of caring, mutuality, solidarity, positive identity, harmonious relationships, well-being, and dignity." It is from this foundation that Ikuenobe comes up with his conception of dignity.

Like Schroeder (2008, 2010), Laitinen et al. (2019), and Mattson and Clark (2011), Ikuenobe argues for a conception of dignity that addresses multiple aspects or use-senses of dignity and their connections. In his formulation, he accounts for the three senses of dignity person x has:

> (1) x has dignity (as a fact or a natural inherent feature), (2) x ought to be treated with dignity, (3a) x acted or behaved with dignity, and (3b) x or x's character manifests or exemplifies dignity. The first sense is a factually descriptive statement of capacity as a potential. The second is a normatively prescriptive statement, while the third (a & b) are evaluative statements of a judgment of one's personhood, character, action, or comportment regarding the actualisation of one's potential (Ikuenobe 2016, p. 457).

Ikuenobe's first sense (factually descriptive) of dignity is based on spiritual, metaphysical, or psychological factors such as autonomy, vitality, life, a soul, rationality, agency, or community. It also refers to one's profession, class, position, or office. So, when people refer to someone having dignity, they refer to this sense. However, Ikuenobe argues

that having the first sense is not sufficient. One must also endeavour for the second and third senses. The second sense is based on how one uses one's capacities for the betterment of the community, whilst senses 3a and 3b have to do with exhibiting virtue. Ikuenobe combines senses 2, 3a and 3b and calls them *moral dignity*. He (2016, p. 458) argues that "moral dignity involves normative and evaluative judgment about someone based on the moral quality of her character, achievement, comportment, or behaviour". This judgment involves the respectful treatment of the person because their actions are morally praiseworthy. Moral dignity is bestowed on a person based on their actions and attitudes. It is from the use of their capacities that their treatment is established. They will be, or at least should be, treated with respect or not based on how they act. Thus, for Ikuenobe, dignity combines the descriptive and moral senses of dignity.

One can have the descriptive sense of dignity but lack the moral sense. For example, one might lack some degree of moral dignity if they are a killer or serial rapist. They still have dignity in the descriptive sense but lack moral dignity because their actions are not considered morally praiseworthy. In this regard, they have the descriptive component because they are persons but lack moral dignity because their actions do not align with acceptable behaviour in society. For Ikuenobe, the notion that moral worth is not conceptually built into the concept of dignity means that one does not have moral worth simply because they have dignity.

However, the moral sense of dignity depends on the factual sense in that "one must be seen as having the factual sense for the other two senses to be applicable" (Ikuenobe 2016, p. 459). According to Ikuenobe, when we conceive of a comprehensive conception of dignity from an African communal viewpoint, we must conceive it as "*a coin with two sides"*. On the one side, we have the factual descriptive sense (D1), and on the other, the normative, prescriptive sense (D2). Being a person of dignity would require possessing both these

senses. It is only once a person possesses both senses that we can with confidence say that they are "a person with dignity" and thus are "worthy of moral and social respect by others based on communal values" (ibid.).

Regarding respect, Ikuenobe (2016) argues that respect from others is not something one has because they are capable. These capacities are "earned and deserved based on their active and positive use" (ibid. p.460). The inherent feature of moral dignity is "acting in a way that manifests self-respect and respect for others by those who are capable, in order to promote harmonious communal living" (ibid.). So, being worthy of dignity is contingent on actively using their capacity to ensure harmonious living. If one's actions are aligned with the norms of the community they are in, then a degree of respect will or should be ascribed to them.

Capable people are also accountable for their actions. If their actions promote harmony, they gain a certain level of respect. However, they lose respect if their actions contradict the community's norms. Ikuenobe (2016, p. 463) argues that "harmonious communal living is morally prior to the absolute dignity or autonomy of, or unconditional respect for, someone who does not use his capacity properly". Imprisonment could be one of the ways such a restriction is enforced. Additionally, capable persons also have "unconditional responsibility to respect, love and care for those (children, those with mental or physical disability) who cannot use their capacity to earn respect" (ibid. p.464).

In summary, the basis for Ikuenobe's communal conception of dignity is the notion of personhood. He argues that personhood is the positive recognition of one's behaviour in terms of how said behaviour enhances communal values of harmony, mutuality, and solidarity. In his conception of dignity, He argues that, in a communal sense, dignity goes beyond having a capacity but rather entails the active use of the capacity to ensure harmonious communal living. It is from this active and positive use of one's capacities that one earns and deserves respect.

Having outlined Ikuenobe's communal conception of dignity, I will discuss some of the issues in his conception. Firstly, based on dignity, Ikuenobe bases dignity on one's personhood and argues that one acquires dignity based on one's morally praiseworthy actions in a community. Basing dignity on personhood is challenging because *not all human beings are persons* (Kant 1998; Laitinen 2007; Molefe 2019a). There are certain qualifying characteristics one would have to have to be considered a person. For Kant, it is rationality. For Motsamai Molefe (2019, p.123), it would be "the moral recognition acquired by the moral agent relative to how well she conducts herself in the light of communal standards". For Ikuenobe (2019), it is the positive recognition of one's behaviour in terms of how said behaviour enhances communal values of harmony, mutuality, and solidarity. By Ikuenobe's logic, only those with personhood would have (moral) dignity and all its associated benefits. Basing dignity on personhood suggests that only those with it would benefit from the entitlements associated with dignity, leaving out all those not considered persons.

Molefe (2019) makes a similar point, arguing that personhood fails to offer a robust conception of dignity. For Molefe (2019, p.119), personhood is an unsuitable ground for dignity because, on the one hand, "dignity is a function merely of being human, and all beings have it in *equal* measure". It is not "a kind of moral achievement or something an agent acquires [or]…the kind of property that varies among agents relative to the conduct of agents" (ibid.). On the other hand, personhood is not a feature that is equal to all. It is acquired in varying degrees from person to person. From this variation, persons are treated differently. Also, its relative and gradational features make it an ill-suited basis for dignity.

Because of this feature, personhood offers "differential respect that depends on special relationships and performance" (ibid., p.126), which means that the treatment of persons would vary based on how society regards them and their actions. Someone regarded as having moral excellence, e.g., Nelson Mandela, would be held in higher moral regard and

thus have more respect. In contrast, a rapist or murderer, e.g., Thabo Bester (a known murderer and rapist in South Africa), would be held in lesser regard and have little to no respect. While partialism is suitable for personhood, it is problematic for dignity. According to Molefe (2019), all people should have dignity because they are human. They must have this dignity on *equal* terms. The idea of dignity is egalitarian and *impartial*. Molefe (2019) also highlights that, typically, dignity is a value that all humans with certain ontological capacities have. Being a feature of all humans means that its associated privileges are thus equal to those who have them. Because of this feature, human rights are based on human dignity.

Ikuenobe might respond to this criticism by suggesting that his conception of dignity has two senses: descriptive and prescriptive. As such, human beings with certain capacities would have dignity in the descriptive and instrumental sense (D_1). It is moral dignity (a combination of prescriptive and evaluative senses of dignity) that only persons would have. So, human beings would have dignity in the descriptive sense but not in the moral sense. This response answers the question of whether people in the margins would have dignity but raises the question of how we treat those who only possess dignity in a descriptive sense. From Ikuenobe's conception, one can deduce two types of people who would have dignity only in a descriptive sense: that is, those that lack capacities or have lost them or yet have to acquire them (e.g., children, the elderly and those with severe mental illnesses), and those who have capacities but cannot or do not want to use them (e.g., a prisoner in solitary confinement or a hermit).

Given that for Ikuenobe, respect is earned based on how persons positively apply their capacities to contribute to the harmony of their communities, this would suggest that those with capacities but who cannot or do not want to use them lack respect in a communal sense because they would not have earned it. This suggests that they can be ill-treated or treated

disrespectfully since they lack dignity and its correlative respectful treatment. With prisoners, given that their actions have harmed their community, they would have lost respect from their communities. Thus, this suggests that their ill-treatment would be justified. For a hermit, it would mean that they might be ill-treated given that their choice of not being part of a community means they have not earned any respect; thus, none is owed to them.

In terms of the marginal cases (elders, children and people with severe mental disabilities), Ikuenobe's (2016, p. 464) response is that his view "implies that we have unconditional responsibility to respect, love and care for those (children, those with mental or physical disability) who lack the ability to use their capacity to earn respect". He writes that:

> If one feels a diminished sense of dignity and self-respect, in terms of lack of autonomy, loss of physical or mental capacities and self-control, based on factors beyond one's control, such as illness or natural mishap, then one still deserves a high degree or duty of respect from others. The idea of respecting unconditionally those who are not capable of acting to earn respect is supported by the moral principle of 'ought to imply can', which indicates that you cannot hold people responsible for what is impossible for them. Dignity involves respect for one's God-given capacity for communal living, but for those who are able to use such capacity freely, there is an added element of accountability, which is that respect depends on meeting one's responsibility to use one's capacity to promote loving and caring communal living (Ikuenobe 2016, p. 464).

Essentially, the burden of care and respect for those who lack the ability to earn it themselves befalls those who can. Those who are capable would do this as part of their roles and responsibilities to the community, mainly because acting this way would earn them more respect. In this sense, the thriving of the incapable in the community would be at the mercy of

the capable. In Ikuenobe's conception, both those who act to enhance the community and those who do not deserve respectful treatment. Those who act earn it, and those who cannot deserve it (perhaps because if they could, they would).

Molefe (2019, p.129) also points out that Ikuenobe's conception of dignity does not tell us from where this 'unconditional respect' would emerge. Ikuenobe's justification of 'ought implies can' bases this unconditional respect on the unknown and unprovable claim that if these agents had capacities, they would use them. Molefe argues that Ikuenobe does not provide a sufficient reason that there is an exception to the rule of earned dignity through positive action that benefits the community. Molefe (ibid. p132) continues to argue that:

> Whereas one may concede that, in this roundabout way, the young and marginal cases would be respected, the concern is that there is a persistent potential that people in society would abuse them. No direct duty to respect them in their own right as human beings exists; and they are quite obviously vulnerable. This potential is even greater in cases where abusing them does not appear to undermine the community's interests or even one's personhood.

Essentially, not having a clear demarcation in terms of what respect exists for persons who have dignity only in an instrumental sense leads to a danger of them being subjected to abuse, especially in situations where there is not any moral consequence to the community.

Ikuenobe may argue that his exception of dignity regarding children can be proven over time based on their *potential* to be contributing community members. This justification would be acceptable. Essentially, we should consider them moral patients with the potential to be moral agents. With older persons who have cognitive issues such as dementia, their unconditional respectful treatment could be justified on *retrospective* grounds – that is, when they could use their capacities, they did, so they deserve respect. With persons with severe

mental disabilities, such as being in a vegetative state, he could justify it in two ways. Firstly, if they were born that way, they can be owed unconditional respect as there is no way of evaluating their actions presently or in the future. Secondly, if they are in that state as a result of some injury, then they could be evaluated retrospectively, similar to older persons.

The three justifications I have provided could make a convincing case for why we would owe older persons, children, and persons with severe mental disabilities unconditional respectful treatment on Ikuenobe's terms. Of course, evaluating the actions of persons in a community using Ikuenobe's logic would suggest that their earned respect varies based on the merits of their actions. This raises questions about the objectiveness of communities. The power of a community to decide on the treatment of the capable based on their actions is dangerous because it relies on the community being objective, judging what treatment is commensurate with which actions, and we all know this rarely happens. In the history of humanity, ample evidence shows that human beings act with egotistical instead of altruistic motives. To quote Molefe (2019, p. 127), "[i]n a world characterised by racism, pauperism, inequality, sexism, human trafficking, the idea of dignity is revolutionary since it demands basic respect for the vulnerable/marginalised merely because they are human".

Of course, encouraging moral participation in a community is essential. However, basing one's dignity on moral action excludes those who cannot or wish not to act. When it comes to moral treatment, all those with the capacity for communal relationships must have equal respect. It should not be based on the evaluation of a community. Ikuenobe does the critical work of emphasising the role of responsibility on agents to ensure harmonious relationships. One can use this as an essential takeaway. He also greatly emphasises duties as more central than rights in communitarian contexts (see also Ikuenobe 2018). However, duties being more central than rights does not take away the significance of rights. As part of protecting

individuals' and communities' rights, people are likely to act in a way that encourages friendly and harmonious relationships.

In this section, I have discussed Ikuenobe's communal conception of dignity and highlighted some criticisms. At the heart of the criticism against his conception is his basing dignity on personhood. Basing dignity on personhood would exclude all those who lack it. I also agree with Molefe (2019) that basing human dignity on personhood is problematic because personhood and dignity are two distinct concepts, each occupying its own moral position. Although they strongly relate, personhood is not a good basis for dignity because it lacks the impartiality often related to dignity. Because of this lack of impartiality, Ikuenobe's conception denies all those not considered as persons dignity, meaning that children, people with severe mental disabilities and older people are all excluded from having dignity, and their treatment relies on the unexplained "unconditional respect" emanating from those with capacities as part of their responsibilities. Those who have capacities but cannot use them or elect not to use them might have dignity in a descriptive sense but may not deserve any respectful treatment.

3.5. Comparing the African Conceptions of Dignity

Having considered the Afro-communitarian conceptions of dignity by Metz and Ikuenobe, I will now evaluate whether these conceptions avoid the pitfalls I have identified in Western conceptions of dignity. The two main pitfalls identified were that Western conceptions of dignity are exclusionary and prioritise individuals over the community in some inappropriate ways. I argue that African conceptions of dignity, particularly my revised version of Metz's conception, address both challenges adequately and thus are plausible to apply when evaluating the impact of AI technologies on human dignity.

On the question of exclusion, I have argued that Kant's conception excludes those who lack rationality. According to Kant (1998), only persons with rationality have the capacity for autonomy and, thus, dignity. I have also shown that Nussbaum (2008) resolves this mostly by arguing that human beings have dignity because of various basic capabilities such as sentience, imagination, emotions, choice and reason. By allowing other capacities to be included, her conception becomes more inclusive than Kant's. The challenge is that her conception is *so inclusive* that it also ascribes dignity to animals, albeit differently from that of humans. The only exception of her conception is that it excludes those with severe mental challenges and anencephalic children.

The African conceptions of dignity also do not do as well on the question of inclusion. Metz's conception is also vulnerable to the exclusion problem levelled against Kant. By ascribing dignity to only agents with a capacity for communal relationships as subjects (those with a full moral status), Metz's conception excludes all humans with merely the capacity for communal relationships as objects of a communal relationship. These include infants and those with severe mental illnesses. Although these agents lack dignity, he argues that they have a moral worth, and for that, they must also be treated with respect because they have a high moral status, although they do not have dignity. Although it excludes others, Metz's conception is more inclusive than Kant's and Ikuenobe's because it is not based on a single capacity, such as rationality, personhood, or autonomy. He ascribes dignity based on one's capacity for friendly relationships. These could be any of the capacities mentioned above or a combination of them. Also, by basing dignity on merely having the capacity and not the use of said capacity, his conception includes even those who are not community members or choose not to use their capacities (e.g., prisoners or hermits).

Ikuenobe (2016) also falls into the same trapping of exclusion identified in Kant's and Metz's conceptions. By basing dignity on personhood, meaning that moral dignity can only

be acquired through actively using one's moral capacities, Ikuenobe excludes all those who lack the capacities to act and those who have them but choose not to act from consideration of moral dignity. Although he makes exceptions for those lacking capacities by insisting they deserve unconditional respectful treatment, using the 'ought implies can' principle, this exception is based on an unjustified assumption that were they able to, they would act in a way that promotes harmonious living. There is no way of knowing that the claim that they will act in ways that promote communal harmony is true. This exception also does not provide them with dignity, only consideration for moral treatment. He also does not make any provision for those with capacities that cannot be based on circumstances (solitary confinement prisoners) or choose not to use them (a hermit).

On the question of which conception is more inclusive, between the African conceptions considered here, Metz's conception is more inclusive. Based on all the conceptions considered in this paper, Nussbaum's is still the most inclusive. I now move to the question of the priority between individuals and the community.

Individual vs community

On Western conceptions' prioritising individuals and individual entitlements over that of the community, Mattson and Clark (2011) argue that communitarian cultures tend to prize values such as duties and obligations instead of rights. This argument is well supported in the literature on African ethics (Gyekye 2023; Ikuenobe 2016; Menkiti 1984). On this question, Metz's and Ikuenobe's conceptions hold opposite views. For Metz, a plausible conception of dignity must account for rights as an important sort of duty, whereas for Ikuenobe, a plausible conception must prioritise duties. Ikuenobe (2016) argues that overemphasising inalienable rights gives everyone an undue sense of entitlement, which is never balanced with responsibilities to others. Additionally, overemphasising unconditional respect gives value and merit to those who may not deserve it. The emphasis, he argues, should be on

accountability and responsibility of humans for using their capacity to ensure communal harmony because not doing so tends to cause social pathologies such as exploitation. For him (ibid., p.442), "respect or entitlements must be earned or deserved by actions".

However, there are also some perspectives in African philosophy that do not divorce rights from duties (Gyekye 2002, Metz 2021). Metz's (2021, 2012) conception of dignity prioritises individual rights while also acknowledging other duties with respect to other people and communities that are not correlated with rights. He (ibid) argues that individuals are owed moral treatment for their own sake because of their ability to commune with others. This position is similar to the Western conceptions of Kant and Nussbaum regarding the treatment of individuals. The difference is that by ascribing dignity based on the capacity for communal relationships, Metz also prioritises the community as it expresses how individuals ought to relate with others.

Metz's priority for the community is also evident in his theory of right action, where he (2021, p. 110 (my emphasis)) argues that "(a)n act is right if and only if it respects *individuals* in virtue of their capacity to be a party to harmonious ways of relating". This theory respects individual rights because it does not prescribe that individuals should have treated community members well to have dignity (unlike Ikuenobe). It only prescribes that, when they act, should they choose to, their actions should be construed as friendly, i.e., they must act in a way that promotes friendly relationships. As a subject, an individual must act rightly towards others; as an object, the individual must be acted upon in a way that respects their capacity for friendly relationships. In addition, Metz (2014b, pp. 142–144) contends that when people have formed communities, that is, networks of friendly relationships in the form of cultures or organisations, one way to treat the individuals with respect is to respect those ways of life they have created.

Regarding the question of which value between rights and duties holds primacy, the literature in African philosophy clearly shows that there are perspectives justifying both rights and duties. Between the two perspectives under evaluation (Metz and Ikuenobe), I find Metz's perspective more appealing for historical and pragmatic reasons. Historically, discrimination has occurred because of debates about who deserves certain rights and who does not. Africa has been a victim of such discrimination for many decades. Advents of colonisation and apartheid gained prominence because a certain collective would bestow rights upon themselves while denying those rights to others. Thus, to avoid such challenges occurring in the future, human rights should be inalienable, not contingent.

Pragmatically, it is the case that although many African societies subscribe to communal values, individual rights are still given priority. This is evident because many African states adopted the Universal Declaration of Human Rights (1948), are members of the United Nations and have adopted the African Charter on Human and Peoples' Rights (Banjul Charter 1981). These documents place primacy on individual human rights, and many African states are signatories of them and have based their constitutions on them. So, it is already the case that although communal values are important in many of these African societies, individual rights also take priority.

Also, prioritising rights does not suggest that duties are neglected, which echoes Ikuenobe's thinking. For instance, Metz (2021, p. 139) argues that rights are "duties on the part of agents to treat individuals in certain ways, but they do not exhaust our duties, with some duties not being correlated with any rights". Chapter 11 of the African Charter on Human and Peoples' Rights (Banjul charter) prescribes duties to African individuals. The Banjul Charter is a human rights instrument authored by African heads of state in 1979. It outlines the importance of protecting human rights and duties that must be enjoyed by African people. Among these duties are duties towards family and society, a duty to respect

and consider others without discrimination, a duty to preserve the harmonious development of the family, a duty to serve the national community, preserve and strengthen social and national solidarity, and positive African cultural values; and to contribute to the promotion and achievement of African unity. The inculcation of duties and rights thus shows no conflict between the two; both rights and duties can be realised.

On which of the conceptions prioritises the community over individuals, both Metz's and Ikuenobe's conceptions better account for a relational as opposed to an individualist perspective. What sets their conceptions apart from other Western individualistic conceptions is that, although they ascribe dignity to an individual, value is not ascribed based on some unique intrinsic quality such as autonomy or rationality (like with Kant) or a combination of these capacities (like that of Nussbaum); it is rather ascribed based on the capacity to relate (Metz) or one's actual relationships (Ikuenobe). The relational conceptions in question are, however, not the same. Ikuenobe's notion of earned respect, that is, respect ascribed only based on morally praiseworthy actions, encourages active participation in a communal context to gain certain benefits. Metz (2021, p. 152) calls this form of relationality, where "an individual has a moral status in virtue of having interacted with other beings in a certain way", holist. These conceptions encounter the pitfalls Metz (2012) has highlighted about other communitarian conceptions, such as that basing dignity on action or membership excludes those who cannot act or choose not to act or refuse community membership.

Metz's communal conception avoids this by basing dignity on capacity, not its use. Admittedly, Metz's conception places value on individuals as well and not on the relationships they can form. He writes that:

> Individuals have a full moral status or a dignity, not because they are in fact members of a communal relationship – say, are part of a family; instead, they have it because they by their nature are able to be subjects and objects of such

> a relationship. By my account, individual entities are what are ultimately owed moral treatment for their own sake, but they are so in virtue of a modal-relational property, their ability to commune with others, which contrasts with intrinsic properties such as the capacity to feel pleasure or act autonomously (Metz 2021, p. 109).

Metz's (2021) account is similar to Western individualist accounts in that moral status is given to entities by virtue of their capacity to be part of moral relationships as opposed to being in said relationships. It is also similar to holist accounts in that it does not ascribe moral status based on two entities' intrinsic properties, such as sentience or rationality. So, his relational conception benefits from individualist and holist perspectives without being constrained by their limitations. His position is the 'sweet spot' between Western individualist and African communal-based conceptions.

Thus far, I Have evaluated the two African communal conceptions to ascertain whether they avoid the pitfalls levelled against Western conceptions. The pitfalls in question are that Western conceptions discussed here have a problem with inclusion. Kant's conception excludes all entities that lack rationality, and Nussbaum's concept is too inclusive in that it ascribes dignity even to animals. Another issue is that the Western conceptions discussed here is that western conceptions prioritise the individual over the community. From the evaluation, it is evident that there are similarities between the Western and African conceptions of dignity. This is beneficial in terms of the common qualities of the concept of dignity, but it means African conceptions considered here are also vulnerable to some of the pitfalls identified against Western conceptions of dignity.

One important similarity is that the conceptions considered thus far (both Western and African) ascribe dignity based on some *capacity.* For Kant, it is the capacity for rationality and autonomy; for Nussbaum (2008, p.359), it is having some basic capabilities such as

"sentience, imagination, emotions, their capacity for reason and choice". For Laitinen et al. (2019, p. 373), it is the "central capacities of persons", such as sentience, self-consciousness and responsiveness to reason. For Metz (2021. 2012), it is the capacity to be part of or party to communal relationships. For Ikuenobe (2016), it is the exercise of the capacity for autonomy and personhood. For Molefe (2022, p.43), it is the capacity for virtue. The primary issue with capacity-based conceptions is that they seem to exclude those lacking the prescribed capacities. The conceptions considered above only vary in terms of the extent to which the exclusion spreads.

Of the two African conceptions considered in this chapter, I have shown that Metz's conception is more inclusive than Ikuenobe's. By ascribing dignity based on capacity and not action, Metz's conception ensures the dignity of even those who cannot act and those who choose not to. On the question of prizing individuality over community, I have shown that Metz's conception, although primarily prizing individuals over the community, is well positioned to highlight and avoid the pitfalls levelled against individualist conceptions and communitarian conceptions that rely on community membership as a condition to ascribe dignity. Attached to the question of prizing individuals over the community is the question of whether the conceptions prioritise rights or duties. I have shown in the arguments above that it is best to prioritise both as opposed to choosing between them. While Ikuenobe's conceptions prize duties over rights, Metz's conception is able to account for both by granting the choice of moral action to the individual; he affords the actor their rights. Moreover, by defining dignity as the capacity for communal relationships based on love and friendship, he ensures that the individual and community's rights and duties are protected through action or inaction.

Having highlighted the similarities and differences between the conceptions of dignity discussed in this thesis, it is evident that although African conceptions of dignity do avoid

some of the pitfalls levelled against Western conceptions, they are also vulnerable to some of the weaknesses of Western conceptions. However, I have shown that Metz's conception of dignity is more robust and more comprehensive. It avoids the pitfalls levelled against Western conceptions without falling into other pitfalls levelled against African conceptions. By basing dignity on the capacity for communal relationships, it avoids the pitfalls of basing dignity on one intrinsic quality, such as autonomy or rationality, as with Kant and Ikuenobe. Also, by insisting on granting dignity based on the capacity for communal relationships instead of being in said relationships themselves (like Ikuenobe), he avoids denying dignity to those who cannot be part of relationships or choose not to be part of them. Lastly, his relational approach avoids the issues of being too individualistic, like Western conceptions, and over-relying on communal relationships compared to other holistic communal conceptions.

The challenge I find in Metz's conception is that it is not as inclusive in ascribing dignity as I would like a plausible conception to be. In normal cases, this lack of inclusivity to the dignity of human beings who can only be a party to communal relationships as objects is not problematic because, according to Metz's conception, they would still have a high moral status. It is only in cases of emergencies that Metz argues that those with a partial moral status would lose out to agents with a full moral status because they matter more. Consider a case, for example, where there is a fire in a hospice, and the firefighter has to choose between saving a nurse or a patient with severe cognitive disabilities. By Metz's argument, because the nurse has dignity as a result of their full moral status, the person with severe cognitive disabilities would lose out.

A conception of dignity that bases dignity on having a full moral status would be problematic in the context of healthcare because those who need robotic interventions in those contexts would not have dignity. That would mean we cannot evaluate the impact these

technologies could have on their dignity because they lack dignity. Yes, they would have to be treated well since they have moral worth, but in case of emergencies, where choices have to be made between them and nurses with full moral status, Metz's conception suggests that it should be the chosen nurses. The unfortunate reality is that all capacity-based conceptions are vulnerable to this critique but with a ranging severity. Depending on the basis for which agents have dignity, some conceptions are more inclusive than others.

Be that as it may, Metz's conception is still the most appealing of those considered here because it is more inclusive than Kant's and Ikuenobe's conception. Although it prizes individuals, it does it in such a way that does not compromise communal values. An individual has dignity because they have the capability to act in a way that will ensure communal relationships based on love. Action or inaction on the part of the agent in Metz's conception would not sow any discord. So, it seems the one issue that is still subject to investigation is the question of inclusivity. I believe that Metz's conception of dignity can be reformulated to ensure that even human beings with the capacity for communal relationships as objects have dignity. In the next section, I will reformulate Metz's conception of dignity to ensure that it is more inclusive, making it a plausible conception to rival or serve as an alternative to the existing Western conceptions applied to evaluate AI technologies' impact on dignity.

3.6. A Plausible African Conception of Dignity

In this section, I aim to formulate a plausible African conception of dignity that would be suitable to evaluate the impact AI technologies could have on dignity. In the previous chapter, I have shown in the previous section that Metz's conception is the more robust African conception. The only challenge with it was that it was not as inclusive. By basing dignity on full moral status, his conception excludes all people who lack full moral status, like children and people with extreme cognitive disabilities. This is especially problematic in

the healthcare sector, where most AI technologies are deployed to aid people with such challenges.

I will reformulate Metz's conception to make it more inclusive. To do this, I will expand on the scope of his conception of dignity by including *human beings* with a partial moral status but with the capacity for communal relations as objects. I argue that human beings with a partial moral status but have the capacity for communal relationships as objects have dignity for the following reasons: firstly, they have a higher *moral worth* than any other entity with a partial moral status; secondly, we can and do *recognise* them as part of us more so than we do any other entity. As such, we can and do relate with them more meaningfully than we do any other entity in this category of entities; lastly, all human beings intuitively have rights of equal measure. I believe that a combination of these reasons and the fact that these human beings already have the capacity for communal relations as objects provides a sufficient reason to grant them dignity. I explain each of these reasons below.

Moral status

According to Metz (2021), entities with the capacity for communal relationships as subjects have a full moral status, and because of this, they have dignity. Ones with the capacity for communal relationships as objects do not have dignity. However, they have a moral worth because of their partial moral status. Their moral worth varies from entity to entity based on how they are graded. For example, the value of a cow would be higher than that of a mouse, so in comparison, the cow would have a higher moral worth. This, Metz calls the gradational approach to moral status.

Not all human beings have a full moral status, meaning those with a partial moral status lack dignity. These include people who are "extremely autistic, psychopathic, and mentally incapacitated" (Metz 2021, p. 164), as well as infants to a certain extent. Such an

exclusion from dignity is especially problematic in the healthcare sector because patients who need treatment, in most cases, lack a full moral status as espoused here. Also, many AI technologies proliferate in this domain for the purpose of assisting those who might not be considered full moral agents, such as older persons, people with cognitive disabilities, and infants. To resolve this concern, I argue that *human beings* with the capacity for communal relationships as objects (those with a partial moral status) also have dignity. This is because, by Metz's own account, they have a higher moral worth than any other entity with a partial moral status.

One might then ask why we are just considering human beings and not any other entity with high moral status, especially given that Metz's formulation would suggest that even some animals have the capacity for communal relationships as objects. A similar challenge was raised against Nussbaum's (2008) conception of dignity possibly extending to animals. She argues that animals could have dignity, but it would not be the same as we would have it. It would be dignity relative to animals of its species but not necessarily comparable to our dignity where, for example, animal protection needs would trump the provision of basic education or monkeys would be entitled to education similar to that of people. Their dignity would ensure entitlements that would be beneficial for them as a species.

To respond to this possible question, I currently do not think we should be regarding animals as having dignity. Them having a moral worth is sufficient to ensure that our treatment of them does not encourage unnecessary cruelty. I will not delve deeper on this matter as it is not central to my concerns. My concern is on the dignity of human beings . This is not to dismiss the conversation about animals and dignity; I am only arguing that that conversation falls outside of the scope of discussion because although there are certain common qualities between human beings and animals, there are also very distinct ways in

which human beings relate with other human beings versus how they would relate to other species.

A child is born from human parents who care for and nurse them for most years, and it is more likely to be held in higher regard than a pet dog they have owned prior. That means, in cases of extreme emergencies, there is an intuition to care for them first, regardless of the idea that other full adults morally matter more. If one's building is on fire, I do not imagine that human adults would save their pet dog before rescuing their newborn. Both these relationships would matter for the owner, but relations between them and the child would matter more. This is precisely why even Metz's (2021) gradational approach to moral status places humans with a partial moral status higher than any other animal. So, the basis for extending the scope of dignity to include people with a partial moral status would be that human beings *morally matter more* because of their various capacities.

Another reason human beings that have the capacity for communal relationships as objects must have dignity is that in addition to their higher moral, we also recognise and relate to them more meaningfully. I now move to the notion of relationality. I explain the notion of recognition and relation below.

Recognition, relationality and rights

In the last section, I provided one of the reasons why human beings with a partial moral status. I argued that they have dignity because they have a higher moral status than any other entity on the moral chain of partial moral status. I also argued that this alone is not a sufficient reason to argue that they have dignity. Here, I advance another reason in addition to the argument on moral status based on recognition. I argue that because human beings with partial moral status have a higher moral worth, we recognise them more than we do any other entity in that moral chain.

Recognition is one of the most important foundations of human coexistence. One cannot claim to have a full sense of self outside of others noticing, acknowledging, or affirming them to be what they consider themselves to be. The concept has a rich philosophical grounding with popular conceptions from Gottlieb Fichte, Georg Wilhelm Friedrich Hegel, Immanuel Kant, Paul Ricoeur, Kwame Appiah, Arto Laitinen, and others. It has to do with the relationship entities have with one another based on acknowledging the other as what they are. With said acknowledgement, it follows the treatment of others in a way commensurate to how they are viewed.

According to Iser Mattias(2019), recognition has both a normative and psychological dimension. Normatively, "[i]f you recognise another person with regard to a certain feature, as an autonomous agent, for example, you do not only *admit* that she has this feature, but you *embrace* a positive attitude towards her for having this feature" (ibid., my emphasis) (see also Laitinen 2007, Ikaheimo 2002). Admitting and embracing that other persons or agents are what they say there are demands a certain type of treatment commensurate to what they are. For example, if we recognise that an agent has dignity in the same way we do, we are obligated to treat them with respect, given that that is how we treat an entity with dignity. Psychologically, Mattias (ibid., my emphasis) argues that "to develop a practical identity, persons fundamentally depend on the *feedback* of other subjects (and of society as a whole)". This means that if one does not receive adequate recognition, this will affect how one perceives themselves. These two dimensions indicate the relational nature of relationships.

In the case of dignity, for example, acknowledging that another person has equal dignity to ours demands that we treat them respectfully. Christopher Allsobrook (2023, p. 120) advances this point by arguing that "it is not my dignity or your dignity, in a person, which we recognise in human dignity, but our dignity, between us. It is the social ontology entailed by such mutual recognition of human dignity that is foregrounded by the African

ethics of *ubuntu*". This means that other human beings can be recognised as having dignity by possessing a certain capacity that is recognisable by the recogniser. It is from these points that I, therefore, argue that one can have dignity because we recognise them as having it. Descriptively, we might recognise all other species for what they are, for example, a cow as livestock and human beings as having a capacity that can be constitutive of it, like being kin to us.

Descriptively, this is just an acknowledgement of entities for what they are, but when it comes to treatment, we are more likely to gravitate towards treating our own species or kin better than we would others. It is unlikely that, even though a dolphin might have a higher cerebral function than a person with severe mental issues, we would hold the dolphin in a higher veneration than we would the person. Thus, I argue that because we recognise other persons more so than we do other entities, we relate with them more meaningfully and treat them with a higher level of respect. It is thus more logical that they deserve dignity equal to ours.

Another reason I argue that human beings with a capacity for communal relationships as objects have dignity is that the modern discourse on universal human rights already grants equal rights to all human beings. In this discourse, all human beings have equal rights because they have dignity. These rights also do not extend to other species except for humans. The justification for these rights, for example, in the International Declaration of Human Rights and the South African constitution, is that human beings have intrinsic dignity, and because they all have it, they must be treated equally. I would emphasise here that this is not my case; they have these rights because they have some capacities that are recognisable as worthy of dignity.

Thus far, I have provided reasons why human beings with partial moral status have dignity. I have shown that it is because human beings have equal rights and hold a higher

moral status (even human beings with partial moral status) than other entities. We also recognise them as closer to our as having dignity than any other; thus, we can relate with them more meaningfully than other entities. With these additions, Metz's conception of dignity would be more inclusive and thus a suitable conception to evaluate the impact of AI technologies on human dignity. This conception is especially important in cases where the evaluations in question apply to cases where the moral status of human beings varies, such as in the case of healthcare and warfare.

Being a capacity-based formulation, I concede that this reformulation does not include all human beings. The challenge with capacity-based conceptions is that they tend to exclude someone, so the point is to try and be as inclusive as possible. My formulation would exclude those human beings who lack the capacity for communal relations even as objects. These are human beings that lack any cerebral functions, such as persons in permanent vegetative states or anencephalic children[16]. These people would lack dignity.

From this formulation, it is evident that a plausible African conception of dignity suitable for evaluating the impact of AI technologies on human dignity must be inclusive, based on capacity, protect individuals' rights, promote communal values, and ensure respect for all with capacities. The conception must be capacity-based because dignity, as defined in the various conceptions discussed in this thesis, is based on different capacities. I do not prescribe what those capacities must be; only those capacities can be used by the agent to be part of or party to communal relationships. As objects, the capacities must be recognisable to others. Regarding inclusivity, I argue that a suitable conception for evaluating AI technologies, especially ones in the HRI domain, must be inclusive because most conceptions of dignity tend to exclude those who, because of their conditions, are more susceptible to

[16] This concession is like Nussbaum's (2008) concession in her conception of dignity.

abuse without any form of recourse. So, including as many people as we reasonably can ensures that violations of people are minimised because having dignity prescribes respectful treatment. Rights and duties are important in African societies. That is why they both find expression in the Banjul Charter and most constitutions worldwide. So, a conception that encapsulates both rights and duties makes it applicable in many societies, not only African ones.

Such a conception would be able to respond adequately to the challenges of AI applications on dignity in most societies without compromising the values of said societies. Lastly, although there are clear conceptual distinctions between societies and the values they prize, almost none of the societies exists in a purist individualistic or communitarian sense. Societies largely exist between these values, so a conception that favours the values of both societies would be more easily applied wherever the need arises. Therefore, this reformulated version ticks all the boxes, thus making it a plausible conception suitable for evaluating the impact AI technologies could have on dignity.

Although the reformulated conception is now more inclusive, we must remember that this conception ascribes dignity based on having a capacity for communal relationships instead of using said capacities. Respecting people's dignity in this regard would mean people simply minding their business and avoiding interactions. However, we know that this is an impossibility in communal contexts. People are likely to interact with each other, and said interactions will likely create discord, which might lead to people intentionally undermining or violating the dignity of others. Thus, in addition to finding a suitable conception of dignity, we also need means to evaluate interactions between persons - or between persons and robots. This is to ensure that such interactions do not undermine or violate dignity. To accomplish this, I rely on Metz's conception of 'right action' as a means

to provide guidelines on evaluating actions of entities to ensure that such actions ensure harmonious living.

Right Action

Metz's (2021) relational moral theory, specifically his conception of 'right action', highlights the importance of acting in acceptable ways and how we can go about evaluating such actions. In his relational, moral theory, which he terms 'rightness as friendliness', Metz's (2021) is composed of his principles of right action.

According to these principles, an action is right or permissible if it does not "merit negative reactions such as criticism, compensation or punishment" (Metz, 2021, p. 110). Metz (2021) goes on to highlight four principles of his conception of right action as follows:

> [1] An act is right if and only if it respects individuals in virtue of their capacity to be party to harmonious ways of relating. [2] An act is wrong insofar as it degrades those with the capability of relating communally as subjects or objects. [3] An action is permissible if it treats beings as special in accordance with their ability to be friendly or to be befriended. [4] An action is impermissible to the extent that it disrespects beings with the ability to be part of relationships of identity and solidarity (Metz 2021, p.110).

According to these principles, it is important to note three things. Firstly, right and permissible actions do not come in degrees. However, wrong actions do because some kinds of actions are morally worse than others. Secondly, these principles are not only applicable to persons. Lastly, respecting an entity's capacity for communal relationships means treating them as the most important and having non-instrumental value.

Additionally, one has positive and negative duties when it comes to these principles. In terms of negative duties, Metz (2021, p.109) argues that "a moral agent is obligated to

avoid degrading, disrespecting what has moral status, namely those beings in the world capable of being party to communal relationships". This can be done by "being alienated from people's projects and indifferent towards their good" (p. 111). However, he argues that distancing oneself from others would fail to honour dignity in a communal setting, necessitating some cohesion. This means that dignity in a communal setting is honoured when one acts in ways that encourage cooperation and harmony. Acting in harmony and cooperation cannot be fulfilled by isolating oneself and not caring about other's interests. For this reason, Metz's relational theory also prescribes positive duties or 'right action'. I will now highlight the implications of the principle of right action in the above context in terms of constraints/goals, guilt/innocence, and partiality/impartiality.

The primary constraint Metz (2021, p.113) mentions is that "negative duties may not be overridden for the sake of fulfilling positive ones". This implies that one should not act unfriendly even if doing so might have a good end. This is the case as this may be construed as treating an agent merely as a means, not an end in themselves. Since there are degrees when it comes to wrong and impermissible actions, killing a patient and harvesting their organs without their consent, for example, would be wrong because the patient has a negative right not to be killed, even if doing so might improve others' positive right to live or have their lives saved.

In terms of guilt and innocence, Metz (2021) emphasises that it would be degrading to innocent parties' (those who have fulfilled their duty not to act unfriendly towards others) capacity for friendliness if they are treated in unfriendly manners, even if that would somehow be beneficial for future friendliness or harmony. However, Metz argues that when it comes to guilty parties, it need not be degrading of their capacity for friendliness to act towards them in unfriendly ways. So essentially, acting unfriendly towards a guilty party to resolve unfriendly behaviour now or as reparation for past unfriendliness is permissible.

Regarding partiality and impartiality, Metz (2021, p. 117) submits that while other African conceptions of dignity position themselves either from an impartial perspective, "that every human has a dignity, and so is equal from a moral point of view", or a partial perspective, that "we owe more to those related to us than we do those who are not", and his view elegantly integrates both perspectives. His view is impartial in that it ascribes dignity to anyone with the capacity to relate communally regardless of whether they act on said capacity or not. It also has partial considerations in that it calls for all who have been related to take up the task of maintaining and enriching said relationships by participating in more helpful ways in those relationships than they had not or been.

The above principles will serve as an important framework for evaluating actions among persons or persons and robots. This is because, by having principles that evaluate agents' positive, the responsibilities that emerge as a result of their actions, we can avoid the harms that are often related to actions. For example, when deciding to use assistive robots in a healthcare setting, we would use these principles to determine permissible actions, be it from an agent or patient perspective. In this sense, the use of AI technologies that would enable or improve the agent's or patient's capacity for communal relationships will be welcomed, while those that undermine said capacity will not. These principles also go further to provide ways of acting toward entities whose actions could be considered unfriendly.

3.7. Conclusion

In this chapter, I discussed the salient conceptions of dignity in sub-Saharan Africa with the aim of finding a plausible alternative conception of dignity to use in evaluating the impact AI technologies could have on dignity. I have shown that four conceptions of dignity are prevalent in the literature on African normative: the vitalist, communitarian, based on spiritual nature and based on personhood. The communitarian conception is the most popular of the four because most African societies are communal. I discussed two communitarian

conceptions of human dignity: one provided by Metz (2012, 2021) and the other by Ikuenobe (2016). For Metz, dignity, in a communal sense, lies in the capacity to form loving relationships. Ikuenobe challenges Metz's communal conception of dignity because, according to him, it is a Western rights-based concept that does not pay attention to the communal duty to act. Ikuenobe argues that dignity, in a communal sense, goes beyond having capacities to using those capacities to ensure harmonious relationships. I argued that Metz's conception is more attractive than Ikuenobe's conception because it avoids some of the pitfalls levelled against Western conceptions while avoiding pitfalls levelled against African conceptions. The only challenge with it is that it is not as inclusive. By ascribing dignity only to agents with a capacity for communal relationships as subjects, it excludes those people who do not have the capacities as subjects, such as those with cognitive disabilities.

To resolve this, I reformulated Metz's conception to ensure that people with the capacity for communal relationships as objects have dignity as well because although they might have a partial moral status, they have a higher moral value than any other entity that has the capacity to be a party to communal relationships as objects. Additionally, we recognise and relate with them more meaningfully than we do other entities. Extending the conception to include human beings with the capacity for communal relations as objects means even people with cognitive disabilities and infants have dignity and its ensuing benefits. This reformulation provides a sufficiently inclusive conception, which makes it plausible to use in robotics domains aimed at vulnerable people.

In addition to the conception of dignity, I also introduced his principles of right action to consider when evaluating the actions of people in a communal relationship. I have argued that combining my reformulated conception of dignity with Metz's take on principles of right action provides a strong framework to evaluate the impact AI technologies could have on

dignity. This is because this conception ascribes dignity based on capacity, not action, and so, in situations where actions have an impact on dignity, they will not be sufficient. So, the theory of 'right action' ensures that dignity is still protected when agents interact with each other and will provide a guideline on how AI technologies ought to interact with agents or patients. In the following chapters, I will evaluate the impact of AI technologies in healthcare and warfare and evaluate this impact based on the view of human dignity I elaborated above.

CHAPTER 4: HUMAN DIGNITY AND CARE ROBOTICS

4.1. Introduction

The thesis's main question under investigation is: Under which conditions do AI technologies undermine human dignity? In the pursuit of answering this question, chapters one and two explored some mainstream concepts of dignity and the various conceptions that emanate from it in the West and in Africa. It found that dignity is an intrinsic value that human beings have because of certain capacities warranting respectful treatment, particularly in the form of observing human rights. From this concept, various conceptions of human dignity arise.

Chapter two focused on the conceptions of dignity salient in the human-robot interaction domain (HRI) – which, for the purposes of this thesis, includes AI in the military domain. The conceptions at issue include Kant's, Nussbaum's and Laitinen et al.'s conceptions. The chapter discussed the strengths and limitations of each conception. The most common finding amongst these conceptions is that they are all based on Western values and norms. Although that is not necessarily an issue in itself, I argue that the discourse is polarised and thus not representative of philosophically plausible views from other parts of the world—a missing part of the discourse to which I seek to contribute.

Chapter Two explored African conceptions of human dignity. These are conceptions that are predominantly based on values and norms that are prevalent in Africa. I explored three conceptions, Metz's, Ikuenobe's and Molefe's. My favoured approach was the Metzian conception because it is more inclusive and respects communal values whilst appreciating individuals' rights, but I elaborated on Metz's conception to be more inclusive. Thus, to answer the question above, I argue that *in a communitarian sense, AI technologies in various domains would undermine dignity if their use inhibited one's capacity to be part of or party to communal relationships either as a subject or an object of that relationship. Conversely,*

they would typically enhance dignity if, in use, they assist one's capacity to love or be loved. The following chapters are aimed at evaluating the impact AI applications could have on dignity and considering whether the conception of dignity I formulated in the previous chapter would be able to withstand some of this impact.

I will evaluate the impact care robots could have on human dignity. It starts by explaining what care robots are and discussing their ethical challenges. The aim is to provide the reader with an overview of the technologies and the ethical arguments prevalent in the discourse. This will provide the reader with a clear understanding of the impacts of already existing technologies, especially as their use pertains to dignity. Following this, I outline the impact care robots could have on the dignity of patients and caregivers. It will specifically outline evaluations by Sharkey (2014) and Laitinen et al. (2019). Following this, I will conduct my own evaluation from an Afro-communitarian perspective.

I aim to answer two questions: What are the demands for dignity between human beings in this domain? What are the demands for AI technologies? The answer to this question will determine what conditions for dignity the interaction with robots employed in the care domain should satisfy. I argue that there are clear benefits and negative consequences associated with using AI technologies in the care domain. Robots that enable patients' capacity for communal relationships enhance dignity, whilst those that hinder that capacity undermine dignity. Robot technologies that enhance dignity, for example, those that assist patients with mobility, enhance communication and entertainment and serve as companions, would be welcome, whilst those that undermine dignity in that their use humiliates, objectifies, deceives, infantilises the patient, and creates a loss of social interaction, would not be welcomed.

4.2. Background into Care Robotics and Ethical Issues

Care robots (sometimes referred to as healthcare robotics) refer to robot technologies that are used to provide assistance and support in taking care of those who need it. The people in need include persons of all ages, with much focus on assisting younger and older persons, as well as persons with disabilities. For the purposes of this chapter, we will name persons assisted by care technologies patients. For over a decade, a large part of the literature has been focused on robotics in eldercare. Although the literature is mostly based on eldercare, and thus most examples will be based on that, the issues raised and how these robots are used are common among vulnerable groups. These technologies can be used in controlled spaces such as hospitals and nursing homes or at the patient's home. Robots currently in use "include robots that support human carers/caregivers (e.g., in lifting patients or transporting material), robots that enable patients to do certain things by themselves (e.g., eat with a robotic arm), and robots that are given to patients for company, comfort and companionship (e.g., the "Paro" robot seal)" (Muller 2021).

Like with any other technologies, there are benefits and ethical issues associated with the use of these robots. The advantages include increased quality of life for the patients. Robots do not become tired, stressed or overworked; they cannot become cruel or act out of anger towards the patient. Essentially, their lack of sentience makes them optimum for functionality. Another advantage is that they would improve the efficiency of care: over time, with the ubiquity of robots, the cost of their purchase will be cheaper, making them accessible to more people. Also, the use of robots will provide some leeway for the caregivers to do other things, and having such would enable the patients to have some independence since they do not always have to rely on caregivers. These robots can thus also allow patients greater independence, minimising the need for institutional care. They have also been shown to be useful for patients with special needs, such as children with learning

disorders or patients with cognitive disabilities. (A. Sharkey 2014; A. Sharkey and Sharkey 2011, 2012). Additionally, certain monitoring robots ensure patient security, and others, such as companion robots, reduce loneliness, thus making up for a shortage of human contact in care environments.

There are many ethical challenges levelled against robots in care. In general, the ethical issues that have to do with the deployment of robots include law and liability, privacy and data security, consent and autonomy, and the replacement of human beings by robots (Stahl and Coeckelbergh 2016; Wullenkord and Eyssel 2020). Within the HRI domain, ethical issues come into play when robots interact with patients either as assistants to the patient directly or as assistants to caregivers. The main issues are that such interaction with robots could replace caregivers, could be deceptive, minimise human contact, and their use could spark feelings of objectification, infantilisation, and loss of agency, which, viewed cumulatively, seem to be undermining the dignity of patients since such interaction could be construed as disrespectful. (Coeckelbergh, 2016; A. Sharkey & Sharkey, 2011, 2012; Sparrow, 2016; Sparrow & Sparrow, 2006). I will discuss these ethical concerns briefly in terms of three broad categories provided by Wullenkord and Eyssel (2020): discrimination of users and robots, dehumanisation of users (and robots), and deception by robots.

Regarding discrimination, the ethical issue raised is that if humans programmed robots may victimise, they may fall prey to the same biases known to cause problems in human-human interaction (Wullenkord & Eyssel, 2020, p. 87). A robot that is trained on data sets that include racial biases could not only exhibit the same biases in the application but, in fact, amplify them due to the nature of machine learning methodologies. Wullenkord and Eyssel (ibid.) also highlight that the design of some robots might discriminate against others if their features are Euro-centric or overly feminized (see also Lamola 2021).

Another ethical and social issue raised is that the proliferation and successful application of these robots would likely minimise human contact (A. Sharkey, 2014; A. Sharkey & Sharkey, 2012; N. Sharkey & Sharkey, 2012; Sparrow, 2016; Sparrow & Sparrow, 2006). The fear is that robots will ultimately replace human social relationships, resulting in the dehumanisation of the human (elderly) user by society, in that they would feel infantilised, objectified and disrespected by the treatment they receive that is devoid of human care.

Another issue has to do with deception. The issue here is that "[r]elationships between humans and robots might even be considered deceptive by their very nature, as they can only simulate a connection that resembles a human–human relationship. There is no consensus in terms of whether robot deception is morally problematic, with some arguing that it is (Sparrow 2006) and others saying such deception is okay as long as it is for the benefit of the affected patient (Sharkey and Sharkey 2012, Coeckelbergh 2016).

Put together, the issues raised above have an adverse consequence on dignity. As Vincent Muller (2021) argues, "While AI can be used to manipulate humans into believing and doing things, it can also be used to drive robots that are problematic if their processes or appearance involve deception, threaten human dignity, or violate the Kantian requirement of 'respect for humanity'". The issues raised above all seem to undermine dignity. Issues such as dehumanising treatment, discrimination, and deception all seem to encompass dignity being undermined in one way or another.

I agree with Laitinen et al. (2019) that having dignity is one of the core qualities of human existence. Thus, dignity being violated would be a bigger affront than any other ethical issue raised thus far. Dignity is also a complex concept, carrying with it a plethora of other associated concepts such as autonomy, respect, recognition, relationality, and privacy, to name a few (I discussed dignity in detail in section 2.2). Dignity can be defined as an

intrinsic value that human beings have because of some capacity (e.g., autonomy, relationality). Because human beings have dignity, the only worthwhile treatment is respect, and given that this value is intrinsic, they must also be treated as ends in themselves and not mere means to others' ends.

In This section, I aimed to highlight some of the salient issues that have been debated concerning the use of robots in care scenarios. It does not cover the full spectrum of the issues, but it provides sufficient information to highlight that there are indeed issues and benefits with the prevalence of such technologies. For fantastic and detailed reviews on the ethics of artificial intelligence, see (Gordon and Nyholm n.d.; Müller 2021; Vandemeulebroucke et al. 2018for a detailed review of the ethics of robots in care)). In addition to the abovementioned issues, some authors (Sharkey 2014; Laitinen et al.2019) have delved deep into evaluating how different robots could affect human dignity. In the next section, I outline those evaluations. Later, I will compare them with mine.

4.3. Dignity and Care Robots.

One of the arguments raised against the use of robotics in healthcare is that such use undermines the human dignity of the patients (see Sharkey 2014. Laitinen et al. 2019). In this section, I will discuss these claims as articulated by Laitinen et al. (2019) and Sharkey (2014). From these discussions, there are cases where the use of care robots enhances the patient's dignity and cases where they undermine their dignity. The most harm seems to be located in instances where the care would deprive the patients of human interactions, deception for profiteering, and dehumanising treatment. Also, the prospect of having exclusive robot-based care is not a possibility that would generally be welcomed regardless of its benefits since the assumption is that humans will prefer to be cared for more by humans than by non-humans. I begin my discussion with an evaluation of Sharkeys (2014) view.

4.3.1. Evaluation # 1 – Amanda Sharkey

Amanda Sharkey (2014) investigates how introducing robots in eldercare would fare in terms of human dignity. She conducts this evaluation using Nussbaum's 'Capabilities Approach' as a framework to evaluate whether the use of care robots – assistive, monitoring, and companion – has an impact on human dignity in eldercare. In her view, dignity is the capacity human beings have based on some basic capabilities. Because they have dignity, they deserve a life worthy of dignity, that is, one that ensures, at a minimum threshold level, ten basic capabilities: life; bodily health; bodily integrity; senses, imagination and thought; emotions; practical reason; affiliation (A&B); other species; play; and control over one's environment (A&B)[17]. She argues that the capabilities approach provides a balanced way of evaluating the impact robots could have on eldercare. Although this thesis is provided for robotics in eldercare, the findings are easily applicable to vulnerable patients. I explain these impacts based on the categories outlined above.

Assistive robots are designed to aid patients with tasks that are difficult for them to undertake or complete based on their conditions. These robots can help older persons by picking them up and placing them in bathtubs, helping them reach toileting facilities, dispensing medication, or helping them with mobility. These include robots such as 'My Spoon', an automatic feeding robot that enables people with limited motor control to feed themselves. The Sanyo bathtub robot provides an automatic washing facility (Sharkey 2014, p. 69). There has also been a growing prominence in the development of exoskeletons. These can "improve the mobility of frail older people or could help their carers to have the strength to lift and move them" (ibid.). One of the recent robots introduced in this category is called 'Pillo', a home health robot that manages and personalises a family's healthcare regimen.

[17] For a detailed discussion of Nussbaum's capabilities approach, refer to section 1.4.

'Pillo' can dispense medication to any member of the family. It can dispense the appropriate medication to the appropriate person using its facial recognition technology. It can also store medicine and refill one's prescription (Girling 2021).

Regarding assistive robots, Sharkey (2014, p. 69) argues that it is possible to identify how such technologies could positively and negatively impact human dignity. For example, "[e]xoskeletons and driverless cars such as ROPITS[18] could provide fragile older people with improved '*Bodily Integrity'* and increased ability 'to move freely from place to place'" (ibid.). Assistive robots that increase older persons' mobility could enable them to have other capabilities, such as affiliating with others. It would also encourage their sense of agency as they would have less reliance on their caregivers. Essentially, by improving their movement, they would have access to a wider range of capabilities that could be seen as leading older persons to a more dignified life. These robots could also assist caregivers with the heavy work of lifting patients. This would essentially minimise the workload of the caregivers, thus minimising the pressure of work and encouraging more socialisation.

These robots could also have adverse effects. For example," robots' insensitive lifting of older persons could make them feel objectified. Such could make them feel humiliated and thus reduce their self-respect. This would affect their capability for *affiliation (b)* – "having the social basis of self-respect and non-humiliation*… "* (Nussbaum 2011, p. 34). Another example would be that if the patients meant to use exoskeletons do not use them safely, they might be subject to injury. In effect, this would negatively affect their *bodily health and bodily integrity*. Lastly, the increased use of assistive robots could mean less human contact, affecting the patient's sense of *affiliation*. Having a robot to assist older persons might create

[18] Robot for Personal Intelligent Transport System (ROPITS) is an assistive robot by Hitachi designed to assist the mobility of older and disabled people.

less incentive for family members or caregivers in care homes to check in on them with the assumption that they are fine.

Monitoring robots are used to monitor the behaviour of frail older people. They can also serve as mobile entertainment and communication devices. Monitoring robots are also assistive robots; the difference is that with them, others (family members or caregivers) outside the patient can connect and access them and even communicate with the robots or with the patients through them (Sharkey 2014). These robots can send emergency messages if anything goes wrong. Most of them have cameras, meaning that others can observe the person being cared for, even from a distance. For example, Gecko Systems has developed a personal robot called 'CareBotTM' that can follow an older person in their home. This robot can also deliver medicine, conduct remote video monitoring, and give verbal reminders at predetermined dates and times. There is also the 'Kompai robot' developed by a company called RoboSoft, which can navigate around the home, respond to simple spoken commands, and even connect family members to the person being contacted via Skype. Another exciting development in this category is a robot called 'HECTOR'. This mobile companion robot interfaces with a 'smart home' and offers support services such as fall detection, diary management, reminders about taking medicines, and remote video conferencing with family members.

Sharkey (2014) argues that these technologies would enhance the dignity of older persons by expanding their range of capabilities in a few ways. Reminding patients when to take medications and eat would ensure a healthier older person, thus enhancing their *bodily health*. Robots such as 'CareO-bot', which can "act as an intelligent walking guide" (ibid., p. 70), would enhance older persons' abilities to move freely and safely, thus enhancing *bodily integrity*. They can also assist older persons with communication since some can connect to the internet and social networks. This would enable older persons to interact with others, thus

enhancing their capability of *affiliation.* Access to the internet would also ensure that older persons have access to information that can be used creatively, thus enforcing the capability of *senses, imagination and thought.* These robots can also improve the lives of caregivers by not having to constantly monitor older persons since they would be alerted whenever there is a problem. Such assistance could also benefit families as they would not have to live in constant anxiety and stress caused by worrying about whether their loved ones are okay.

The potential negative impact of monitoring robots on older persons' dignity would be that monitoring is intrusive. Although monitoring older persons could provide peace of mind to family members and caregivers, it might be unpleasant to the older person who would constantly be under surveillance. No one wants to live under constant surveillance. Such surveillance would undermine the patient's right to privacy. Another possible adverse effect would be that the effectiveness of such robots might minimise social contact. If caregivers are certain that one is doing well and are monitored, they are less likely to do it themselves. This would, in essence, reduce the older persons' capability for *affiliation.* Another issue would be that such robots could prohibit specific actions because said actions are 'not safe'. By so doing, they would be limiting the older person's freedoms.

Lastly, companion robots are robots built to create some form of companionship for the user. Many such robots have been built, and some are already used in care facilities. Some can even be bought for personal use (Sharkey 2014). The most famous robot in this category is called 'Paro', a seal-like robot that provides the benefits of having an animal pet that can be used in places that are not pet-friendly or by persons who are not able to care for a pet but who are lonely (Girling 2021).

Regarding companion robots, Sharkey (2014) argues that these robots could address the commonly held criticism against robots in care – that they would take away older persons' need for social interaction (Sharkey and Sharkey 2012; Sparrow and Sparrow 2006).

Having companion robots would fulfil the need for interaction with other people and thus could fulfil the capabilities of *affiliation, emotion,* and *play*. Having a robot pet would mean the older person has a companion with whom they can play, converse and interact. Such social contact has been seen to reduce stress and anxiety in older people. Sharkey (ibid., p.71) argues that "all of these benefits are likely to be especially important for those with dementia or other cognitive impairments that make regular social interaction with other people more difficult". This argument is evidenced by several studies that use Paro for therapeutic purposes.

For example, Takanori Shibata (2012) conducted a study where they observed the use of Paro in therapy for patients with dementia. They found that robot therapy was effective in positively affecting the mood of patients. The same study indicated how these robot pets seem to have advantages for those who care for patients with dementia as well. It was found that stroking the robot also had relaxing effects for the carers. Similarly, Sharkey &Wood (2014, n.p.) argue that "a consideration of the robots' effects in the context of the Capability Approach makes it possible to identify ways in which it could be seen as enabling because it can increase the access of people with cognitive impairments to a wider range of capabilities than would otherwise be possible".

Sharkey also highlights the risks levelled against care robots by Robert Sparrow(2016); for example, these robots would reduce social contact, and their use to care for patients is akin to deception. Regarding the issue of social contact, Sharkey (2014) argues that such concerns may be unnecessary based on the evidence provided above that shows using care robots as a solution to social interaction. Regarding deception, the most popular reference is that of Robert and Linda Sparrow (2006), who argue that robots are incapable of meeting older people's social and emotional needs. They (ibid., p.148) argue that using robots in older persons' care is akin to deception because the thought that robots have human

qualities such as sentience and consciousness might make older people happy. However, this happiness would be a result of manipulation.

In an earlier text, Amanda Sharkey and Noel Sharkey (2011) challenge this idea by arguing that, although much of AI depends on creating illusions and thus is a form of deception, the conclusion that such deception is unethical is extreme, especially considering that much of the deception is based on the "anthropomorphic contribution of the user" (ibid., p.34). Essentially, it is not the robot itself that deceives; the patients are deceived by the human qualities they ascribe to the robot. People generally ascribe human features such as consciousness to objects and then expect said objects to behave based on said features. In this sense, the suggestion of deception would be self-imposed. Another thing to consider is that anthropomorphism is not a problem unique to robots. There are many other objects that human beings ascribe human-like features to, such as people naming their cars or children treating their toys as though they are alive. In many of these contexts, it is not that the patients are not aware that the entity in question does not have the features that they ascribe to them (A. Sharkey & Wood, 2014).

The Sharkeys (2011, 2012) argue that instead of focusing on all forms of deception, we should focus on those with the potential to have negative consequences. For them, two negative consequences of using robots in older persons' care warrant attention. Firstly, it increases anxiety for the older person because they think they have a human relationship with the robot and thus have to take care of it, even at the expense of their well-being. This might lead observers and relatives of a confused older person looking after a robot pet to think that the robot is depriving their relative of dignity and infantilising them (2011, p.35). Secondly, "the presence of a robot might result in a reduction in the level of social interaction an elderly person experiences" (ibid.). There might be fewer incentives for other facility members to

visit if they know their loved one has company. These issues are well addressed in this section under the different categories of robots that have been explained above.

So, when it comes to the issue of deception, Sharkey 2014 is not convinced that the issue is as important as her other counterparts make it out to be. She (ibid., p.1) argues that "[p]eople can enjoy playing along with an illusion and behaving 'as if' a robot was able to understand them". Sharkey argues that we should consider, however, that a case could be made about the difference between real and virtual relationships. Unfortunately, she argues that the CA does not seem to have a provision for such a distinction.

There is also a risk that companion robots would undermine the capability of *affiliation* (having the social bases of self-respect and non-humiliation, being able to be treated as a being whose worth is equal to that of others). Sharkey (2014, p.72) argues that it is possible that if one is neglected and only has access to a pet robot, they would engage with it but feel ashamed or humiliated. This might create an impression among family members that older persons experience a loss of dignity as a result of being humiliated. However, this is a bit of a stretch in that it is based on the impression of others that the older person would lose their dignity and not the older person themselves.

Sharkey (2014, p. 74) concludes that "the CA permits a valuable and balanced perspective on the relationship between robots and the dignity of older people, highlighting some positive consequences that could result from the careful deployment of certain forms of robotics and warning of potentially negative ones". Coeckelbergh (2016) holds a similar view to that of Sharkey. He (ibid., p. 461) argues that "[g]ood elderly care, and especially preserving human dignity in older person care … requires that care seeks to maintain or enhance older people's capabilities".

The capabilities approach to human dignity is an interesting way to evaluate the impact of robotics in care because it provides one with a list of capabilities that could be used as a checklist to check if there are benefits and challenges. In the case of Sharkeys (2014) evaluation, it is interesting to note that with all the robot technologies evaluated, there is evidence of an enhancement of capabilities, thus dignity and a possibility of the risks to certain capabilities, which could undermine dignity. For example, with assistive robots, there are ways in which these robots could enhance the capabilities of patients by improving bodily integrity and affiliation, whilst she also shows how they could challenge one's affiliation by invoking feelings of objectification or challenging bodily integrity by causing injury. With monitoring robots, she also shows how their use can improve bodily health and integrity by reminding patients to take their medication and being able to use them as communication and entertainment devices. She also shows the benefits and consequences regarding companion robots – that they could improve patients' affiliation, emotion and play and how they could risk capabilities of affiliation. This provides a balanced evaluation of care robotics, showing benefits and consequences.

Regarding the issues raised, the main concerns prevalent in Sharkeys (2014) evaluation are that using such robots might lead to dehumanising treatment, diminish patients' human interaction, and undermine the right to privacy. She also mentions the issue of deception but does not seem convinced that it is as pertinent an issue as the Sparrows (2006) and others make it out to be.

She is not the only one who holds this view regarding deception. Growing research highlights that robot deception might be beneficial and lead to good consequences. One such example is Coeckelbergh's (2016) paper titled 'Care Robots and the Future of ICT-mediated elderly care: a response to doomsday scenarios. In this paper, he (ibid.) responds to the challenges levelled by the Sparrows (2006) against the use of robots in older persons' care as

they pertain to deception. On the claim that suitable care is necessarily incompatible with deception, he (ibid., p. 456) argues that the worry about deception is "problematic if it assumes that good care should never involve any form of deception [because] while there may be a prima facie obligation to treat people as autonomous persons, caregivers also have an obligation to take into account the particular empirical situation and capacities of the older person and in some cases the result might involve 'deception', without this being necessarily morally problematic". For example, an older person with dementia might find it challenging to interact with others but might benefit from a robot pet. Also, a child with severe autism might benefit from developing a relationship with a robot companion. In these two cases, deception does not seem to be harmful. Some even go as far as to say that robots need to deceive in order for them to be beneficial (e.g., Isaac & Bridewell, 2017).

According to Muller (2021), the idea of deception is problematic in cases where the design of the robots is deliberately deceptive for profiteering purposes. An example is Hanson Robotics 'Sophia', a humanoid robot purported to have human qualities. Here, we can say that deception is problematic because the company uses manipulation tactics to draw popularity and profit. Rule four of the five ethical principles for robotics, published by Margaret Boden and Colleagues (2017, p. 127), states that robots "should not be designed in a deceptive way to exploit vulnerable users; instead, their machine nature should be transparent". This rule prohibits the deliberate trickery of robot technologies and proposes that everyone using the robot must be able to understand that its intelligence is artificial. Another problem with humanoid robots like 'Sophia', for Noel Sharkey (2018), is that Hanson Robotics is using such robots "as a platform to represent the current state of artificial intelligence falsely and to actively deceive the public into believing that we have AGI (Artificial General Intelligence) or are very close to it". The company is currently creating a

false impression in terms of the current capabilities of AI. These impressions create impossible expectations of the capabilities of the technology.

Regarding human contact, the idea of robot companions and assistance reducing human contact is well supported (A. Sharkey & Sharkey, 2012; Sparrow, 2016; Sparrow & Sparrow, 2006). The Sharkeys (2012, p. 30) argue that "reduced social interaction can have a measurable impact on the health and well-being of the older persons and reinforces the idea that depriving them of such contact is unethical and even a form of cruelty". The Sparrows (2006) also make a case that if robots were used for tasks such as cleaning floors, this would remove a valuable opportunity for social interaction between a human cleaner and the senior citizen and, as such, should not be encouraged. Elsewhere, Robert Sparrow (2016) describes a future scenario where older persons would be cared for entirely by robots and argues that a future like that would be dystopian. For Sparrow (2016), using robots in older person care would deprive them of the two important objective elements of welfare, recognition and respect, which are important to the practice of aged care.

Although there is a need for social contact and risk associated with deploying robots as assistants (replacing human assistants) and as companions (reducing family contact), there are also cases made about robots alleviating loneliness that results from lack of human contact (Khosravi and Ghapanchi 2016; A. Sharkey and Sharkey 2012). In a review of the effectiveness of assistive technologies in senior care, Pouria Khosravia and Amir Hossein Ghapanchi (2015, p. 20) argue that "[t]hese technologies have [a] positive impact on social and emotional well-being and reduce the social isolation of seniors by increasing social connectedness". They highlight that interactions with the robot seal 'Paro' have reduced social isolation in seniors. The Sharkeys (2012) highlight a similar point with a robotic dog called 'AIBO'. Another study shows how social robots have had a positive impact on the quality of life for older persons by reducing agitation and anxiety (Pu et al., 2019).

Sharkey & Sharkey (2012) also highlight some studies that show that the loneliness scores of older persons interacting with robots reduce loneliness, and their quality-of-life assessment improves. The Sharkeys (ibid.:37) thus conclude that "robot companions for the elderly could offer positive benefits in terms of improvements in health and welfare, although these are risks in terms of dignity and loss of social contact" (see also Sharkey & Sharkey 2011).

Another related issue is objectification, that "[r]obots designed as replacement nurses or carers that carry out some of the same tasks of feeding, lifting etc., may make their charges feel like objects" (2012, p.30) (see also Decker, 2008; Sparrow & Sparrow, 2006). These types of technologies undermine the autonomy of the older persons. Sometimes, the introduction of such technologies is for the benefit of the caregivers or the company in question since automation increases productivity and cuts costs. In this regard, the patients are treated as mere means and not as ends in themselves. It sometimes leads the patients to feel like they have lost control or agency. Monitoring robots could be deemed an invasion of privacy, thus undermining the autonomy of the patient. Most of these issues are prevalent if these technologies are provided for the patients without their consent or consideration.

Lastly, the notion of privacy is tricky in the context of care. It goes without saying that privacy in terms of patient information should always be protected. In this regard, such a violation would not be acceptable. This is well covered in health professionals' principles of good practice (see HPCSA, 2021). As to privacy regarding surveillance, this would vary based on the condition of the patients and what interventions are required. For example, a patient whose condition could cause harm to themselves or others might need to be monitored—however, one whose condition is less severe, not so much. Essentially, privacy challenges vary from patient to patient, with others benefiting from it and others not needing it. I do agree that deception is problematic when the aim is to deceive for profiteering as

opposed to care. By this, I mean that if a patient can benefit from deception by robots, then the deception is permissible. It is in cases where the deception is not for a social utility that I find to be problematic.

Regarding the other issues that are prevalent, I agree with Sharkey that if the use of robots results in the inhumane treatment of patients, it would be an affront to dignity. By virtue of patients having dignity, they are entitled to respectful treatment. They are also entitled to certain rights, such as privacy, granted that privacy would depend on the severity of the condition of the patient. Whilst some patients might need constant monitoring, others might not. So, the use of these robots would be beneficial insofar as their use is commensurate with the needs of the patient. Laitinen and colleagues (2019) also wrote a paper in which they conducted a similar evaluation using a different conception of dignity. I now turn to their evaluation.

4.3.2. Evaluation #2 – Laitinen and colleagues

Laitinen and Colleagues (2019) conducted a similar investigation to that of Sharkey (2014). To answer the question of how robotic care would support or challenge dignity, they argue that the answer to that depends on two things: firstly, whether the care at stake is robot-based, robot-assisted or teleoperated; secondly, the distinction between demands and realisations of human dignity. They first systematically explain what they mean by dignity and then analyse the impact of robots on older persons' care.

On the concept of dignity, they take a similar stance to that of Schroeder (2008, 2010) and Ikuenobe (2016). Laitinen et al. (2019) conceptualise dignity as having two aspects: the inalienable aspect (D_1) and the variable aspect (D_2). The inalienable aspect refers to human dignity in terms of a "source of strictly undeniable, stringent, and unconditional normative claim; dignity is something that everyone possesses automatically and equally merely

because they are humans or persons" (ibid, p. 372). This conception of dignity is similar to various authors' conceptions covered in this thesis thus far. It is what Molefe (2019, p. 118) calls a "standard philosophical understanding of dignity". It is based on the Kantian conception that sees persons as having an intrinsic worth that deems them ends in themselves. Like Nussbaum, Metz, and Molefe, this intrinsic worth is based on one's central capacities. Regarding this aspect of dignity, there are strict moral boundaries in terms of how one can be treated. According to Laitinen et al. (2019, p.373), the boundaries are that "the inalienable dignity is to be respected, not violated, and any interactions must take place within the boundaries for respect for dignity".

Laitinen et al. (2019) argue that the demands for respectful treatment of moral agents by other moral agents take the form of *negative and positive* duties. Essentially, for human dignity to be upheld, specific actions or inactions must be carried out. They argue that the negative duty could be met by doing nothing. One cannot violate any value of a person through inaction and lack of interaction. The positive duties, on the other hand, demand some form of action, such as helping people in need. They (ibid., p.373) highlight *three norms* in which positive duties to inalienable dignity could be met: we should "protect each other as vulnerable beings with needs (H1), we [should] respect each other as autonomous agents (H2) and [we should] respect and engage with each other as beings with sophisticated inner lives, as rational thinkers, emoters and subjects of experience (H3)".

These three norms constitute the three targets of respect for human dignity, and they are all relevant to older people's care and can assist in assessing the ethical acceptability of robot care. H1 would refer to older persons as care receivers, including lifting and cognitive, recreational support and therapy robots, e.g., *Paro.* The relevant robots for H2 would include "intelligent rollators, robots that can simulate the role of care-recipient and activation robots" (ibid., p.374). For H3, "the cognitive aspect of autonomy and respecting an older person's

personal experience and opinions are relevant; how they are acknowledged or accepted by others will affect their self-relations" (ibid.). From these three norms, they (ibid., p. 374) formulate three positive and negative duties for all moral agents:

Negative duties	**Positive duties**
1. Not harming others	1. Taking due care of human needs and protecting vulnerabilities
2. Not blocking people's autonomous agency	2. Aiding and supporting people's autonomous agency
3. Not blocking people's rational thinking and subjectivity	3. Aiding and supporting people as thinkers and subjects of experience

If one breaches the negative duties, it is a *violation* of human dignity, and if they succeed in meeting the negative duty, then it is *respect* for human dignity. Breaches to positive duties are called *neglect* of human dignity, whereas successfully meeting positive duties creates *positive support* for human dignity. These distinctions are important because they highlight moral agents' duties as a result of the dignity of moral patients. Such a distinction assists in ascertaining what is considered appropriate actions in the agent-patient relationship.

Laitinen et al. (2019) continue to argue that if robots were considered moral agents, they would have the same duties. However, robots are not moral agents and thus cannot have duties. Because they cannot have duties, Laitinen et al. (2019, p.374) argue that they should "be built to be such agents so that they function; accordingly, they *ought to be* such that the dignity of moral patients is not violated but is supported". They must be developed in such a way that, in use, they do not violate or neglect the dignity of the older persons (or any other

moral patient); rather, they support it. In terms of the above norms, robots "ought to be built so that they do not harm yet protect vulnerable humans and help meet human needs" (Laitinen et al. 2019, p. 375), meaning that they ought to ensure that the targets for respect, H1- H3, are met. These are the demands and requirements of dignity based on D_1.

How can these demands be met in practice? For Laitinen et al. (2019), how these demands are responded to constitutes the variable realisation-aspect of dignity (D_2). D_2 is "contingent and gradable in how it is fully realised in actual living" (Laitinen et al. 2019, p. 372). This form of dignity can be realised in varied degrees based on one's "*own* actions, emotions or self-relations" or in "*interactions* with others or in one *living conditions*" (ibid.). A person who behaves well in their community or has done well in their community would be held in higher esteem than someone who does not. If one is a doctor and assists community members with ailments, their regard in the community would be of higher value than someone who steals from a community. This regard of the two would not take away from their intrinsic value (D_1), but D1 would be undermined or promoted through treatment. Laitinen et al. (2009, p. 376) outline three aspects where dignity could be realised in varying degrees: "[o]ne's own attitudes and comportment can be more or less dignified; the way others interact with and recognise one can realise dignity to a higher or lower degree; and the cultural, material, and institutional background conditions can be consistent with the demands of dignity". Essentially, the extent to which one's dignity is realised is contingent on factors other than that they have it as an inalienable status.

How people go about their everyday lives and interact with others determines how they are treated. Although they always ought to be treated with respect, such treatment would vary from one person to the other based on the impression they give or others have of them. Also, the environment plays an essential role in how people interact and treat each other. For example, older persons are held in higher regard in African cultures than young people.

Additionally, some societies are economically better off than others, meaning their residents will have access to a better quality of life than in other places. Another one is that certain jobs are thought to be more prestigious than others, so people are treated differently based on their jobs. Laitinen et al. (2019) argue that in the same way that fulfilling positive duties would enforce and promote human dignity in moral agents, fulfilling such duties in the case of robots in terms of robots' ought-to-be norms might promote human dignity too. Failure of this by robots may lead to neglect of moral agents that could have otherwise been fulfilled by moral agents or not-so-smart technologies.

Regarding the impact that robots could have on the dignity of older persons in care, Laitinen et al. (2019) apply the three norms that emanate from positive duties, i.e., vulnerability, agency, and experiential and cognitive subjectivity, to evaluate three types of robot technologies: robot-based, robot-assisted and teleoperated robots. The robot can be near the patient or in the background with robot-assisted care. In cases where the patient has certain capabilities, the robot can assist them or assist the caregivers in providing care for the patient. In robot-based care, the robot is in direct contact with the older person. It is the robot that provides care without any assistance. The third category is robots, where the caregiver teleoperates. These robots fall in between the two other categories.

Below is a diagram that shows Laitinen et al.'s (2019) three categories and examples of robots.

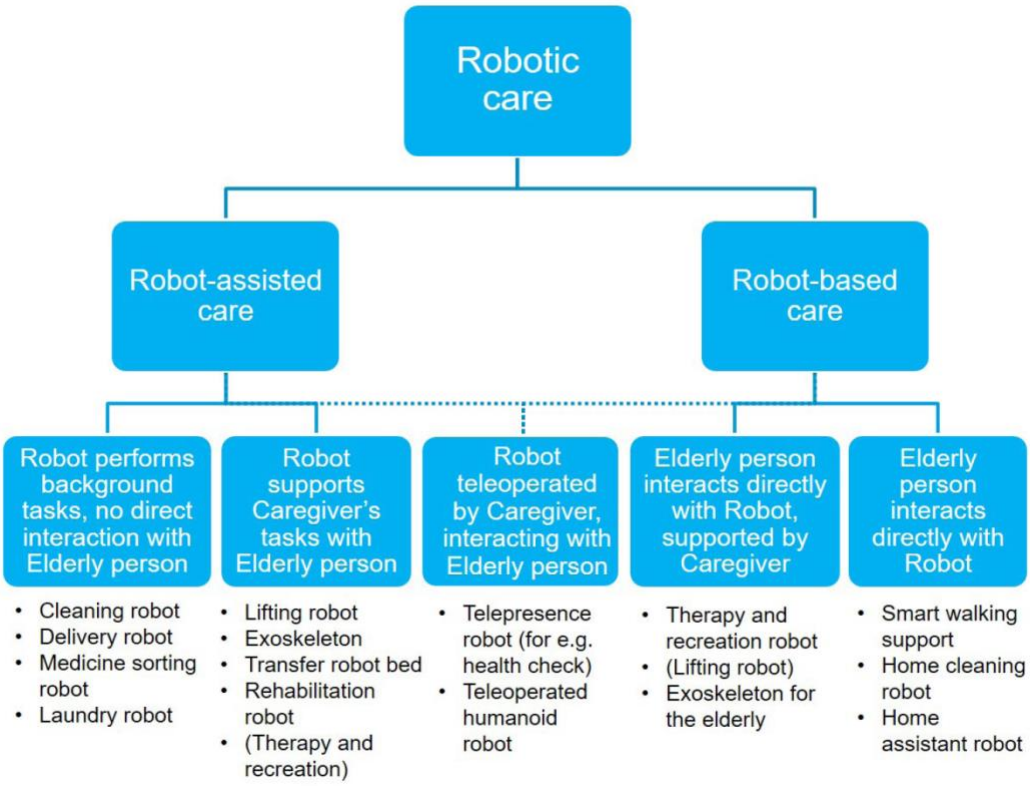

Figure 1. Categories of Robot Care (Laitinen et al. 2019)

The main difference between robot-assisted and robot-based care is that it is hypothesised that in extreme cases, this would be a scenario where only robots provide care without the presence and assistance of human caregivers. Robert Sparrow (2016) theorises such an event and argues that a future where care for the vulnerable is solely conducted by robots would be dystopian.

Laitinen et al.'s findings are summarised in the table below*:*

Dignity realised in ↓	Concern for →	Vulnerability and needs	Agency	Emotion and cognition
The inalienable status		(Negative duty) not harming others; and (positive duty) taking due care of human needs and protecting vulnerabilities	(Negative duty) not blocking people's autonomous agency; and (positive duty) aiding and supporting people's autonomous agency.	(Negative duty) not blocking peoples rational thinking and subjectivity; and (positive duty) supporting people as thinkers and subjects of experience

Own action **(Robot-assisted self-care)**	Positive: Robotic assistance for protection of safety in one's own activity. Violation: self-injury in using robots. Neglect: neglect of ones needs in one's own action.	Positive: Increasing physical, agentic capabilities. Violation: (robot-assisted) disregard for one's dignity in one's action Neglect: neglect of ones (future) needs in one's own action.	Positive: Increasing cognitive capacities. Violation: (robot-assisted) disregard for one's dignity in one's action. Neglect: neglect of ones (cognitive) needs in one's own action.
Robot-based "interaction" in care	Positive: needs met. Violation: risk of injury. Neglect: cannot provide human contact for social needs.	Positive: make more capable, strengthen and motivate. Violation: risk of disability Neglect: cannot provide real interaction.	Positive: make more capable, challenge and stimulate. Violate: risk of disability. Neglect: cannot provide real recognition and communication.

Robot-assisted interaction in care	Positive: helping a nurse treat a person safely. Violation: nurse causing pain or anxiety to a person by robot use [it is argued here that nurses are better equipped at telling if a patient is in distress merely by movement of facial expressions, robots on the other hand cannot] Neglect: sometimes may not provide human contact for social needs.	Positive: supporting a person's own participation in treatments. Violation: preventing a person's own participation by robot use. Neglect: sometimes may not provide real interaction.	Positive: robot taking care of hard work (e.g., lifting) enables nurses to concentrate on emotional needs of a person. Violation: objectification of a person by robot use. Neglect: sometimes may not provide real recognition.

Teleoperated interaction in care	Positive: client needs met with technological solutions. Violation: possible injury and insult via teleoperation. Neglect: cannot provide equally good human contact; some technologies restricted to communication, fewer aspects of well-being (e.g., smell, general view of an apartment) may be recognised.	Positive: supporting agency by strengthening and motivating. Violation: possible violations of autonomy via teleoperation. Neglect: cannot provide equally embodied interaction; mere instructions instead of action.	Positive: enables meaningful conversations for the lonely. Violation: possible denial of cognitive autonomy in teleoperated interaction. Neglect: sometimes may not provide real recognition; lack of comprehensive human interaction.
Background conditions of living	Smarter technology may be better, but may be more dangerous?	Smarter technology may be better, but will it require too much physically?	Smarter technology may be better, but will it dumb us down? Will it require too much cognitively?

Table 1: Realisation of dignity in robotic care regarding vulnerability, agency and cognition (Laitinen et al. 2019, p. 378-379)

What is key about Laitinen et al. (2019) findings is that the major controversy regarding robots in older persons' care comes from robot-based care. They (ibid., p. 377)

argue that "the challenge with robot-based care is that it is unclear whether interaction with or recognition from robots can be directly constitutive of human dignity; whether robot-based care could directly provide the needed recognition for patients"). With assistive and teleoperated robots, "it is easy to see how robots could help support dignity by assisting the agent's own actions, and how they can be a smart part of living conditions consistent with human dignity" (ibid.).

Another central challenge highlighted is that the sole or increased use of this technology would reduce human contact. The most common response in this regard is that robots are not meant to replace human beings. The aim is mostly to ease the burden of caregivers in varied ways. It has also been shown above that there is, in fact, evidence of the introduction of robots such as *Paro* that actually address the common problem of loneliness in caring for the vulnerable.

Another issue that Laitinen et al. (2019) highlight is that of bodily integrity and the need for the human touch in older persons' care. They argue that in this regard, robotics has both positive and negative effects; assistive robots may help older persons go to bathrooms or feed themselves; however, the consequence could be the deprivation of touch that they would have received had the caregiver been human.

Another concern they raise is inequality, essentially being the issue of whether the benefits of robotics will be available to the better off. Laitinen et al. (2009, p. 391) conclude that the issue of the impact of robotics on dignity *depends* on whether the care is robot-based, robot-assisted or teleoperated, "each contributes to realisations of human dignity both positively and negatively, and it will expectably depend on institutional and cultural settings whether positive or negative effects dominate".

Having highlighted the two analyses of the question of the impact of care robots on human dignity, I wish to highlight a few key similarities between both. Firstly, both sources agree that there are identifiable benefits to assistive, teleoperated and monitoring robots. Their instrumental use makes them less likely candidates for fundamental harm to human dignity, especially when used with human caregivers or by patients with some discernment capabilities. Although there could be harm from using these robots, that harm does not pose much threat to human dignity. The injury resulting from these can easily be covered in risk and liability principles found in other technologies such as cell phones or cars.

The challenge for these thinkers emerges when these assistive technologies are used exclusively as caregivers. Robot-based care might be a challenge because it takes away human contact, which is already a challenge in the care environment. Of course, Sharkey (2014) clearly remarks that companion robots such as *Paro* have been created to alleviate this very problem. There should, however, always be a conscious effort to reinforce companionship, not to substitute it. Robot-based care is also frowned upon by others, such as Sparrow (2016), who argues that a future where care is exclusively robot-based is dystopian. Be that as it may, there are some advantages to it. If the patient in question still has certain capabilities, then such care would advance their agency and alleviate having to feel like one is a burden. However, if they are completely reliant on the care, it might perpetuate loneliness and, thus, unhappiness.

These two analyses show a clear case about the benefits and consequences of using robots in care - their differences lie in their approaches to human dignity. Sharkey (2014) uses the CA to conduct this evaluation, while Laitinen et al. (2019) use the dual-aspect conception of dignity. Both conceptions make for good frameworks of evaluations in terms of conducting such evaluations in Western individualist-based societies. Given that not all societies' moral foundations are based on Western individualist approaches, my approach

focuses on a relational approach that would be attractive to societies that place communal values as foundational, such as many African societies. This is a new approach, given that evaluations are based on individualist conceptions.

The next section will conduct a similar evaluation using an Afro-communitarian conception of human dignity (proposed in Chapter 3). In principle, I agree that assistive technologies used with caregivers do not fundamentally threaten human dignity. I also agree with Laitinen et al. (2019) and Coeckelbergh (2016) that although robot-based care could be beneficial, we should not aspire for it. In my conception, as shown below, we would only consider robotic technologies when there is a clear indication that their functioning improves the existing friendly relationships. That is why technologies that work in tandem with human care are attractive, and those that seek to replace or substitute human caregivers are not.

4.4. Afro-Communitarian Evaluation

In this section, I evaluate the impact that care robots could have on dignity. Following the evaluation by Sharkey (2014) and Laitinen et al. (2019), I will conduct a similar evaluation of my African conception of human dignity, considering assistive robots, monitoring robots, and companion robots. I argue that one has dignity insofar as one has the capacity to be part of or party to communal relationships as subjects *or* objects (those with a partial moral status because we recognise and can relate with them in a more meaningful way). Their possession of dignity entitles them to respectful treatment by other agents. This treatment can be evidenced by acknowledging an agent's capacity for communal relationships or interacting with agents in a way that is harmonious. In essence, agents must act in such a way that they respect other agents' capacity for friendly relationships.

I aim to answer the question: under which conditions would care robots undermine dignity, supposing an African perspective? On this view, given that robots are designed to

serve a particular purpose, my conception of dignity would necessitate that they are designed in a way that, in their use, would enhance the capacity for communal relationships. In the case of healthcare givers, they would ensure that the nurses are aided in providing the highest quality of care. For the patients, they would enable their capacity for communal relationships by enabling them to participate (should they wish) in harmonious relationships meaningfully.

So, any AI technology that would enhance the capacities of the agents in a healthcare space to relate with others would be welcome. Also, any technologies that would assist the caregivers (professional or family member) in providing meaningful care would be welcomed. Thus, the response to the above-stated question is that robots would undermine the patients' dignity if their use were aptly construed as unfriendly either by the patient or any agent observing, that is, if their use undermines the agent's capacity to commune with others or others to commune with them. I will now evaluate assistive, monitoring and companion robots from this point of view.

4.4.1. Assistive robots

As stated above, assistive robots are robot technologies that work to aid the user (patient or caregiver) with tasks that they struggle such as lifting, moving, and feeding them. They could positively or negatively affect patients' dignity in a communal setting. Using them could enhance the dignity of the patient if their use can help the patients connect with others by enabling their mobility. One of the challenges with vulnerable patients is the need to move from one place to another. With many patients in need of care, this is not an easy feat. For patients to be able to move, they often need assistance from others, such as caregivers. Depending on the severity of the cases, this can range from helping someone to the wheelchair to assisting them in toileting facilities.

Robots that can assist patients in moving around and help them eat, use toilet facilities, and bathe themselves would be beneficial in that they would provide the patient with a sense of relief, knowing that they do not have to rely on others overly. Although reliance on others is not an issue in and of itself, a patient's capacity to do things by themselves improves their capacity to relate in the sense that they can navigate difficult tasks by themselves. This also helps those who would otherwise have to care for the patient to do it much easier than they otherwise would have or not have to do it at all. Such assistance would ensure that the way those involved engage is in a cooperative manner. Being able to do these things with the assistance of robotics would mean that they are less reliant on the help of others. By this, I do not mean to say that others would not be happy to help in communal societies. I am merely saying that it would be beneficial for patients to have the capacity to do what many consider rudimentary tasks by themselves so that they can participate in communal relationships more meaningfully. The use of assistive technologies for this purpose has been welcomed in the past. Thus, there should be no reason to resist robots in the same category but with advanced features.

Wheelchairs, for example, are one of the technologies that have empowered many people with mobility worldwide. For those who have lost the ability to walk as a result of injury or medical conditions, wheelchairs have assisted with the means to move around. Over time, their functions have expanded to be able to assist even patients with more complex conditions, such as muscular dystrophy - a rare neuromuscular disease that causes progressive weakness and breakdown of skeletal muscles over time (Emery 2002). Wheelchairs have, in effect, drastically improved the quality of life of their users. Thus, more advanced mobility technologies would do the same, even on a greater scale.

Lately, exoskeletons, for example, have enhanced patients' capacity for movement without the aid of others. These technologies help patients with severe physical impairments

to mobility to an even greater extent than wheelchairs could. In a movie called "The Fundamentals of Caring", the actor named Trevor, with Duchenne Muscular Dystrophy, has a wish to urinate while standing. While this might seem like an easy task for the reader, it is an impossibility for Trevor, given his condition. To make Trevor's wish possible, Ben (Trevor's caregiver), with the assistance of paramedics, straps Trevor onto a stretcher and lifts him so that he can urinate, standing upright. Using exoskeletons would make it possible for Trevor and others with similar conditions to do such tasks by themselves. This would minimise the need for additional caregivers at home or the need to admit such patients to permanent care homes.

Similarly, self-driving cars[19] could assist visually impaired patients in travelling to their desired destination without having to rely on others. The ability to move freely could mean patients can participate in community activities in a meaningful way. A person who feels empowered and self-sufficient is more likely to be a productive community member and open to relating with others. As my preferred conception shows, although agents have dignity only insofar as they have the capacity to relate to others, distancing oneself from others would fail to honour this dignity in that dignity in a communal setting is honoured when one acts in ways that encourage cooperation and harmony. So, agents who have gained the capacity to connect with others with the assistance of these technologies will more likely be willing to do so.

Other assistive robots such as 'Pillo', 'Robert the Robot', 'My Spoon', and others would help them with bathing themselves, accessing toileting facilities, or eating, thus

[19] Self-driving cars are another frontier technology that have their own ethical challenges, that I might briefly highlight in the chapter on autonomous weapons, just to show what issues are involved. But I will not delve much in those issues of autonomy. I am assuming that their implementation will only happen when they have met all the necessary ethical criterion.

alleviating the humiliation that could emerge from relying on others for such basic needs. The use of assistive robots could benefit the patient and, thus, the rest of the community. Such help could mean alleviating the burden on their loved ones who would otherwise have to do so. If people do not have to take care of their loved ones on a full-time basis, they will have more time to do other things. This enhances capacities for harmonious living. Enabling patients to have such capacities would ensure they can contribute to and benefit from communal relationships without being compelled to be in them because of their needs. This would, in effect, enhance their dignity.

The main ethical challenges raised against assistive robots are that such technologies would reduce human contact and that their use would make the patients feel objectified. With the issue of reducing human contact, the ethical challenge is that using these technologies would reduce the level of social interaction, causing loneliness for patients. Sharkey and Sharkey (2012) argue that loneliness is one of the main issues when adults get older and their children start their own lives. There is sufficient research that affirms this claim. For example, Hansen & Salsola (2016) highlight that a study on late-life loneliness conducted among older adults in 11 countries across Europe found the prevalence of loneliness common in France, Norway, Russia, Czech Republic, Bulgaria, and Georgia.

Although there is evidence of the correlation between old age and loneliness, most of this research is prevalent in Europe, Asia, and America, and the issue is not as prevalent in African societies. Here, the convention is that most of us live with our older persons at our homes, and the process of *mutual care* continues in a different way. As Michael Kpessa-Whyte (2018, p. 405) poignantly argues, older people mostly live in the same household with other family members to help take care of grandchildren and other relatives, enabling the economically active population in the family to focus on their work. The working generation provides support for them. Therefore, the older people continue to contribute to the family's

overall welfare by helping to raise, educate, and transmit time-honoured Indigenous wisdom and traditional values to the younger generation in their care while supporting minimal housekeeping duties where health conditions allow.

This traditional means of mutual care is common among many African societies. The sites of care are usually in our homes and are not a service that is outsourced, meaning that there is less prevalence when we take our loved ones to care homes. The act of sending a loved one to care homes is considered by most to be cold, impersonal and a form of neglect. It is thus shunned, and those who do it are seen as an abomination. The act of caring for one another in this way is what Metz (2012, 2021) correctly captures in terms of the two central pillars that make our communities successful: identity and solidarity, as explained above (see section 3.3).

Thus, the criticism that assistive robots have the potential to minimise social contact and thus cause loneliness is prevalent in other parts of the world, but it is not likely to be the case in African communal societies. Here, it is even very unlikely that the older persons would be sent to nursing homes, and second, they would be neglected if they were to be sent there. The more likely case is that families would even opt to hire a nurse as an additional caregiver at home rather than send a family member to care institutions. So, it is likely that if assistive robots are to be deployed, they will avoid having to send our loved ones to care homes, similar to how we opt to hire a nurse as opposed to sending our loved ones to care homes. In this regard, these robots would not be a substitute for humans or human contact but would serve as a value add as and when the patient needs it. Having such robots would only enhance already existing networks of care.

However, some emerging studies highlight the changes in traditional structures of care in Africa and, thus, the risk of lowered social interaction and the resulting loneliness among older persons (Kpessa-Whyte 2018; Phaswana-Mafuya and Peltzer 2017). Some of

the reasons cited for this emergence are urbanisation, socioeconomic development and globalisation. Although the prevalence of loneliness is unlikely, the fact that it is slowly creeping in as a result of changing traditional social structures is a source of concern. If there were robot technologies that would further drive the wedge between the patient and other human beings, such a technology would be unwelcome. Reducing social contact, and thus a technology that reduces social contact, would dishonour dignity because anything that would distance one from their community would undermine dignity. Dignity in a communal setting is honoured when people act in ways that encourage harmony and cooperation. In this sense, I also agree with Metz (2021, p.94) that "one who identifies with others enjoys a sense of belonging or is glad of the presence of others, in contrast to feeling lonely or isolated". We are more likely to want to be with one another than apart from one another.

So, to avoid this challenge with care robots, we would have to be intentional that their deployment aims to entrench social connections as opposed to driving a wedge amongst people. This can be achieved by ensuring that their deployment is in tandem with already existing networks of care as opposed to replacing them. Dignity in a communal context is guaranteed by our capacity to be part of friendly relationships. Anything that risks this capacity would not be welcomed. As such, unless it is the patient's choice, solitude should not be imposed.

Regarding the ethical issue of objectification, that the handling of the patients by some of these technologies may cause feelings of being 'thingified', resulting in humiliation and loss of self-respect, my conception advocates for respectful treatment of persons. I share Metz's (2021) sentiments that respectful treatment for moral agents entails the negative duty of avoiding treating them in degrading ways, but also and importantly, acting in ways that promote friendliness and harmony. So, the use of robots that can be construed by the patient as humiliating or objectifying would be disrespectful to the patient. Such an act can be

deemed as unfriendly and thus undermining the patient's dignity. Thus, robots ought to be designed and used in such a way that their use respects the patient's capacity for communal relationships. This means that in their use, they should not induce negative feelings such as humiliation or objectification. In this regard, any treatment of the patient that would trigger or be construed to trigger a sense of humiliation on the part of the patient would not be acceptable. Such actions would be construed as undermining the patient's dignity. Thus, any robotic use that would be regarded as demeaning or infantilising would be considered disrespectful and thus unacceptable.

The use of assistive technologies should be for the advancement of the dignity of the persons who use them. Even if it is not demeaning to the patients but invokes these feelings in other members of the community, it would not be acceptable, or at least there would be some moral reason to avoid it. Unlike in Western conceptions, where the focus is only on the patient or user, the communal conception also holds regard for how these technologies would affect other members of a community.

In addition to the potential ethical challenges, there are risks, such as injuries or improper use. These risks are not so novel that they would require new risk mitigation processes or undermine the patient's dignity. Importantly, in this regard, it would be ensuring that there is due care to ensure minimal risks. This means that the patient or the caregiver must be well acquainted with how the robot works so that they can use it properly and stop using it when they feel discomfort. Using these technologies carefully will also be empowering to the patients and minimise feelings of objectification and infantilisation that they would otherwise have suffered at the hands of their caregivers.

From an Afro-communitarian perspective, there seems to be very little danger in the use of assistive robotics to care for our loved ones. The benefits show that increased mobility enables patients to be involved and engaged in communal activities in meaningful ways.

Here, there is a similarity with Sharkey's (2014) evaluation that increased mobility ensures a patient's capacity to affiliate with others. Where our approaches seem to veer off is the point where such mobility ensures self-reliance. In my conception, self-reliance is not as pertinent a value as interdependence. So, I argue that increased mobility means increased connections in ways that are not strenuous to those involved.

Regarding the challenges, objectification and loss of interaction are not crucial concerns in communal contexts. The fact that such contexts encourage interdependence as opposed to individuality makes it almost organic for agents to gravitate toward each other, not the other way around. So, fears of objectification by the robots are unlikely since these technologies will be used collectively by the patients and caregivers. Injuries could be mitigated more easily because, unlike Laitinen et al. (2019), the notion of 'self-care' would not be desirable as it would not honour dignity in a communal context. The same can be said with the notion of isolation. Thus, the risks of assistive robots do not outweigh their benefits. Having assistive technologies is like having an extra hand to do what is essential and, in fact, maintain or even add to a patient's dignity to improve their ability to maintain communal relationships.'

4.4.2. Monitoring robots

Regarding monitoring robots, like assistive robots, there are identifiable ways in which their use would enhance or undermine dignity. As Sharkey (2014) has shown, monitoring robots would enhance patients' dignity because their use assists them in ensuring that they take their medications on time, fill up their prescriptions, remind them to eat, ensure that they feel safe knowing that someone is watching over them, and enable them to communicate with their loved ones, even from a distance. Assisting patients with these functions expands their range of capabilities, including 'bodily health', 'bodily integrity', 'affiliation', and 'sense, thought

and imagination'. From a capabilities perspective, these advanced capabilities would ensure that the patient is self-sufficient.

My preferred conception also acknowledges that these enhanced capabilities by robots would enhance patients' dignity, however, for different motivations. These capabilities enhance the patient's dignity by ensuring a wider and effortless connection with loved ones. Having a robot to remind them to do the things they need to do and assisting them in doing them would ensure that the patients are in good health, are able to reach out to loved ones (even at a distance), have a sense of safety knowing they are looked after and have access to entertainment. Having such assistance would minimise the load for the caregivers, especially because, in most cases, the caregivers are family members.

The task of taking care of loved ones might make the patient feel like they are a burden to those who care for them. Having such assistance robots would ensure that caregivers do not always have to be present and in the vicinity of those they care for, thus assisting patients in caring for themselves and thereby helping their families. These robots can also send emergency messages to caregivers in times of danger. Like assistive robots, monitoring robots would provide patients and other members of the family peace of mind, knowing that their loved ones are safe. The loved ones and caregivers would not have to be physically present all the time to care for their loved ones. Thus, the cost of caring for loved ones would not require them to forfeit their ambitions and interests.

Two ethical issues regarding monitoring robots are that they would infringe on the patient's privacy and minimise social contact. Regarding the issue of reduced social interaction, like with assistive robots, that is less likely to be an issue in communitarian contexts. As I have explained above, the more desirable scenario is that these technologies will be used in tandem with already existing networks of mutual care. In the case where they encourage the distancing of social networks, they would be unwelcome.

On the issue of privacy, the argument is that constant monitoring can be construed as intrusive and undermining a patient's privacy (Sharkey 2014). This is especially the case if such monitoring robots are used without the patient's consent. Such application undermines the patient's dignity in that it undermines their autonomy. In a communal sense however interdependence is prized beyond independence or autonomy. In that context privacy would be negotiated among all members and not only the patient. The notion of privacy is referred to here, not as it is often referred to in AI and big data debates that speak to data ownership, data mining or data sharing as it is controversially referred to as crises of social media. In the context of eldercare, it has to do more with the right to be left alone or not to be surveilled, especially in cases where the patient does not consent.

The notion of privacy is tricky because, on the one hand, in my conception of dignity, I advocate for the protection of individual rights, but on the other, I advocate for the promotion of communal values. On the point of protecting individual rights, such surveillance would undermine the patient's right to privacy. Using robots in a way that violates the patient's rights would not be acceptable. The right to privacy, as captured in Article 12[20] of the Universal Declaration of Human Rights and Section 14[21] of the South African constitution, states that every individual has a right to privacy, which must be respected. Thus, as a right, the privacy of the patient must be respected, and the use of such robots should be in a way that is acceptable to the patient. It would be considered an

[20] "[n]o one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks".

[21] Everyone has the right to privacy, which includes the right not to have a. their person or home searched; b. their property searched; c. their possessions seized; or d. the privacy of their communications infringed.

unfriendly act if the use of robots would infringe on the patient's right to privacy. Such use would be construed as disrespecting the agents' rights.

Of course, the right to privacy in such cases is often challenged on the grounds of safety. The argument usually goes that certain monitoring technologies and robots are installed for the safety of vulnerable patients. Such technologies make it possible for caregivers to respond with haste in times of emergency. Although it is the case that such technologies would enable rapid response in the cases of emergencies, in a liberal formulation where rights take priority, it would still be necessary that the patients are made aware and consent (as far as they are able to) to the surveillance. It would have to be the responsibility of the caregivers to explain the benefits and risks of such robots to their patients.

If there is informed consent from the patient, then the use of such robots to care for the patients would be regarded as friendly. Such friendliness would be an acknowledgement of the patient's agency and respect for their dignity. In cases where the patient cannot fully consent because of their condition, it is still the caregiver's responsibility to explain to the best of their abilities what they are doing and how it is beneficial to the patient. Even if the patient cannot evaluate the benefit themselves, such benefit must be evident even for any onlooker.

The argument of the right to privacy and consent is well captured in Kantian and Nussbaum's principles that advocate for the autonomy and liberty of agents. Such ideas are also expressed in the South African constitution because it is based mostly on international instruments such as the UNHR. However, from a normative perspective, I agree with Metz (2021) that an African relational moral theory would not place as much weight on considerations of privacy as Western moral theories do because:

> [I]ndividuals are, by a characteristically African moral perspective, normally understood to have weighty duties to aid specific others, particularly family members, it is not merely up to the individual what she does with her body and mind. Since others have a stake in the individual being healthy, some have held that they ought to be aware of her illness and play a role in discussing how she ought to treat it. Metz (2021, p. 193)

Essentially, when it comes to the well-being of a patient, others, such as family members, have a stake in terms of what happens to them. It is not the sole right of the patient in question. Related to this is the notion of consent. Whereas Kantian and Nussbaum's conceptions would prioritise agency and thus consent as individual decisions, the relational moral theory I advocate for would specifically advocate for a consent that is attached to solidarity (see Metz 2021). By this, I mean the consent of the patient is acceptable if it is in tandem with the interests of loved ones.

So, in the case of monitoring, it would be necessary for the caregivers to make the patient aware of the benefits and risks of the surveillance. Imposing such a technology without the patient's knowledge and acknowledgement would be unfriendly. It is also necessary for the patient to choose a course of action that considers the interests of others apart from themselves. This position is balanced in that, although it acknowledges the rights of the patient, it also highlights how some rights are secondary to the duties we have in our communal relationships.

In Metz's (2021) words, "[f]riendly relationships, of the morally attractive sort that include joint projects, require not only transparency between actors about their goals but also willingness on the part of each to achieve them". If monitoring robots are applied in the way proposed here, their use will enhance the patients' dignity by enhancing their capacities for friendly communal relationships. Their use would respect the patient's rights to privacy in

that they will be used only in ways that the patient is comfortable with and has consented to, and in the case where they are unable to consent, their benefit would be evident to any onlooker.

4.4.3. Companion robots

There are also identifiable ways in which companion robots can promote or undermine dignity. I have explained above that Sharkey (2014) argues that such robots would minimise or alleviate the often-stated problem that technologies in care minimise social interaction and thus invoke loneliness. She argues that these robots would minimise that risk in that they would fulfil capabilities of *affiliation, emotion,* and *play*. Making an example of a robot pet known as 'Paro', She argues that having such a pet can enable the patient to have an entity from which they can play, converse and interact, which has been shown to have psychological benefits, especially with older patients with mental conditions such as dementia.

As I have already shown with regard to assistive and monitoring robots, the crisis of loss of social interaction, which seems to be one of the main concerns with robotics in healthcare, is not a major problem in communal contexts. This is because communal systems are contingent on relations with others. The same argument holds for companion robots. These robots would not minimise or substitute human contact because of the strong network of mutual care in African communities. In these communities, when the older person is unable to take care of themselves, they are invited to live with their children or relatives. If older persons had a robot companion, it would be added to the already existing network of care. In the unlikely event that the use of (companion) robots would cause a distance between the patient and their community, such would be unwelcome. Dignity, in the communal sense, is guaranteed by one's capacity to be in communal relationships either as objects or subjects. Isolating one from their community would constitute an affront to their dignity.

The other issue that is often raised is that of deception. The argument is that the successful use of these robots relies on manipulating the user. Sharkey (2014) does not seem to find issues with deceiving patients, especially if the benefits of said deception outweigh the harm. So, although deception could be an issue in itself, it seems to be justifiable in cases where deception is needed to care for someone (see also Coeckelbergh, 2016). From these explanations, it seems the notion of deception could be problematic in itself but could be justifiable or permissible based on the motivation. So, if the deception is for the purpose of helping patients, it would be acceptable. However, if the deception is for nefarious purposes like profiteering, then it would not be acceptable. This kind of consequential viewpoint considers the rightness or wrongness of deception based on whether the consequences are favourable or non-favourable.

In my African conception of dignity, which fashions itself to be deontological, the notion of deception is morally wrong, no matter the consequences. However, this is a hard-line perspective that is usually associated with Kant's ideas on lying.

A similar point is made by Metz (2021), who argues that it is immoral to deceive a person. He (ibid., p. 125) argues that:

> Deceiving another person is essentially a way of stunting another's ability to participate on a cooperative basis. When you lie to a person or otherwise get him to believe what you know to be false, you are not merely failing to share a way of life together when you could but also directly impeding the other person's capacity for that with you (Metz 2021, p.125)

In 'normal' cases where deception involves a person with full moral capacities, I agree with this stance because subjecting people to situations where they do not fully understand the conditions of what they are being subjected to is unfriendly. However, in the

case of patients who lack the capacity to decide for themselves and would benefit from such deception, then I argue it would be justified. In this case, I prefer a more subtle response to deception, which suggests that it "is always wrong to some extent, but that countervailing moral factors might sometimes suffice to make manipulation justified on balance"(Noggle 2022). From this viewpoint, one would have to ask about the intention of the deception. If the moral justification of the action outweighs the deception itself, then it would be permissible. So, in my case, an act of deception would be permissible "if it treats beings as special in accordance with their ability to be friendly or to be befriended" (Metz 2021, p.110).

So, in the case of companion robots, purchasing a robot pet for them, in cases where they have difficulty interacting with others, would be acting in solidarity with a person. People suffering from extreme cases of dementia or autism, for example, find it difficult to interact with others. That is why, in many instances, when their conditions become worse, they end up having to be institutionalised. If I were to choose on behalf of a loved one to either be placed under severe antipsychotic medication, be subjected to being shackled, or undergo therapy that involves companion robotics, I would choose the robotics option. For me, that is a better alternative to the other choices that would be presented.

One way to determine this is to ask whether the deception causes harm or benefit to the patient. In the cases shown above, where the patients have challenges interacting with people but benefit from having a robot pet like Paro, there is a clear benefit to the patient. Whether they are aware that they are being deceived might be important but not harmful. In cases where patients cannot fully comprehend what is happening, they should be made aware of it, and they should consent to the robot's use. If a patient is deemed not to have the capacity to consent fully, the decision to use these robots would still have to be made in the same complex network of mutual caring I have explained above. In any instance, the decision to use such technologies should be based on benefiting the user in a way she accepts.

Suppose the decision to deceive is made on the grounds of capital gain. In that case, where companies would deceive the patients for profitability, this act will be considered unfriendly and contrary to the harmony of the community. It will thus be unacceptable and should be avoided. In this regard, as explained above, such an act would be construed as unfriendly.

Thus far, it is evident that there are arguments to be made for the positive and negative use of care robots in an Afro-communitarian context. Having evaluated the major criticisms: objectification, human contact, privacy and deception, a few observations can be made. Firstly, the objections are legitimate in that if the use of robots would objectify, infantilise, and reduce human contact or be deceitful to the patient and the caregivers, then that can be construed as undermining their dignity in a communitarian sense. My proposed conception of dignity states that one has dignity by virtue of one's capacity to be part of or party to communal relationships. Because of this dignity, others must honour and respect this capacity. Their actions must be construed as upholding or enhancing this capacity. Any use of this technology that upholds and enhances this capacity would be welcomed, and any use that undermines it will be unwelcome.

Secondly, the use of robots will most be welcomed if it is an addition to an already existing network of care, not a substitution. This means the introduction of robotic technologies cannot be brought about by plans to substitute caregivers. Rather, they must be brought to complement them. In this sense, I agree with Coeckelbergh (2016), Laitinen et al. (2019), and Sparrow (2016) that while robot-assistive care may be welcomed, robot-based care would not be welcomed.

It appears that, with many issues, African conceptions arrive at conclusions similar to Western conceptions, although for different reasons. On the issue of humiliation, there is consensus that humiliation constitutes disrespect and thus, robotics that triggers feelings of

humiliation would be unwelcomed. While the Western conceptions focus only on the humiliation of the patient, mine focuses on other affected people in addition to the patient. Even if the use of this robot evokes such a feeling to a member of the family, then such a robot or that specific use would be unwelcomed. This also goes to the notion of privacy, where, in the Western conception, it is only the privacy of the individual in question. In my conception, the right to privacy of the patient must be conceded in tandem with the duties they have for their families. The patient would have to be open to some invasion if this invasion would allay the fears of other family members. So, the notion of privacy is not as much a priority as it is in Sharkey's evaluation.

While the concern in western conceptions about reduced human contact is that such reduction would result in the patients being abandoned and lonely, in my case, the likelihood of loneliness in unlikely because there is a network of care that involves all members of families, relatives and the community at large. So, instances of loved one being placed in care homes and abandoned are less prevalent in communal societies, so our use of these technologies adds to an already existing network of care.

In confronting contemporary technologies, one must accept that communities' contemporary formation and functions have changed from what they were in the past. People now have to work to maintain their livelihoods and care for their loved ones. Because of this, they sometimes have to rely on other means to connect and care for them. In this context, the use of robots seems to be one way this is possible. In a comparable way, productivity applications that track and submit work and help people connect virtually or e-hailing services are used for food and transportation; we could derive value from using assistive robotics. Their use would enable people's capacities, especially when they are designed and applied to the communitarian values my conception proposes.

In combination, all these technologies assist in social and productive aspects. During the full lockdown in 2020 at the start of the covid 19 pandemic, people were confined to spaces they found themselves in, with no opportunity to visit their loved ones. The only way to safely communicate with community members was through various technologies such as cell phones. Even if one was in a different country, in times of danger, they could call an ambulance for their family members, call a neighbour to check on them, or send money using electronic means. Essentially, one could care for their loved ones without being physically there. Those who could galvanise such technologies were not as isolated and in danger as those who could not. The lesson from this time is that technologies are almost always created to resolve societal issues. Of course, some would use them for malicious purposes. That is why it is always important to monitor them and ensure that they align with the values of those they are meant for in use. That way, the use of various technologies will be beneficial to society.

In this section, I evaluated the impact of care robots on dignity using my African conception of human dignity. I argued that a case could be made about robots promoting and undermining dignity. In terms of promoting dignity, I argued that assistive robots would promote dignity if they aided the patient's capacity for communal relationships. This could be done by ensuring mobility and assisting them with daily functions like bathing, eating, and moving around. Monitoring robots would ensure patients can take their medication on time, provide entertainment, and help patients communicate with their loved ones. Companion robots can assist patients who struggle to interact with other human beings due to their medical conditions. For example, a patient with dementia would benefit from having a robotic pet without bearing the responsibility of having to care for it. So, regarding the benefits of care robots, we are arriving at the same conclusions as Sharkey (2014) and Laitinen et al. (2019), but for different reasons. Where they argue that the benefits increase

the patient's self-reliance, my conception argues that the benefits increase the patient's capacity to relate with others. There are also several ways in which care robots undermine dignity. If the use of care robots objectifies, reduces human interaction, undermines patients' privacy, and is deception. In a communal setting, dignity is honoured when people act in ways that encourage cooperation.

Thus, Care robots must be designed and used to enhance the capacity for communal relationships. This means that they would ensure that they assist caregivers in providing the highest quality of care. Also, they enable the patient's capacity for communal relationships by enabling them to participate meaningfully (should they wish) in harmonious relationships. So, robots would undermine the patients' dignity if their use were aptly construed as unfriendly either by the patient or any agent observing, that is, if their use undermines the agent's capacity to commune with others or others to commune with them and they would enhance dignity if their use would enhance the capacities of the agents for communal relationships.

4.5. Conclusion

In this chapter, I evaluated the impact care robots could have on human dignity. I began by providing state-of-the-art care robots. Additionally, I highlighted the ethical challenges associated with applying these technologies in general and specifically as it pertains to human dignity. I outlined two evaluations that are present in the literature by Sharkey (2014) and Laitinen et al. (2019). Both these papers show that there are clear cases where the use of these technologies could undermine dignity. There are also clear cases where they do enhance dignity.

Following this, I conducted my own evaluation using my formulation of the Afro-communitarian conception of dignity. I argued that there are clear benefits and negative consequences associated with the application of AI technologies in care. Robots that enable

patients' capacity for communal relationships enhance dignity, whilst those that hinder that capacity undermine dignity. Robot technologies that enhance dignity, for example, those that assist patients with mobility, enhance communication and entertainment and serve as companions, would be welcome, whilst those that undermine dignity in that their use humiliates, objectifies, deceives, infantilises the patient, and creates a loss of social interaction, would not be welcomed. The next chapter moves away from the care paradigm but still considers how technology impacts society by focusing on an evaluation of the impacts of autonomous weapons on human dignity.

CHAPTER 5: HUMAN DIGNITY AND AUTONOMOUS WEAPONS

5.1. Introduction

Another domain that has received attention when it comes to AI technologies has been the military applications of AI, specifically the development and possible deployment of autonomous weapons systems. Although their applications do not fall strictly in the HRI domain, I argue that they are important to consider because, due to the conventional nature of warfare, such technologies are likely to be applied in civilian areas, meaning that there will be some interaction with civilians who are part of the war theatre but not party to it. Autonomous weapons systems (AWS) are weapons systems that, once activated, can select and engage targets without human intervention (Asaro 2012; Heyns 2013, 2017; Omotoyinbo 2023; A. Sharkey 2019; N. Sharkey 2014; Sparrow 2007; I. Taylor 2021). This implies that once programmed, AWS can determine when and how a target is executed without any human intervention. Such functionality falls out of the parameters of our current weapons in circulation and, thus, our understanding of the role and capability of weapons.

Weapons are usually operated by human combatants responsible for making the critical decisions of targeting and pulling the trigger. However, this capacity of autonomous weapons to select and engage targets is possibly what makes others attribute them to the status of 'autonomy'. The development of such weapons has been met with large contestations and debates on why they should not be deployed or, in the event that they are, how they should be regulated. Chief among the issues raised is that such weapons would increase the appetite of political actors to go to war, that in their use, they would contravene laws of war, create a responsibility gap and undermine human dignity.

In this chapter, I evaluate the impact the development, deployment, and possible use of autonomous weapons could have on human dignity. Given that autonomous weapons are

likely to be used in warfare, I will first briefly outline the principles of a just war. Following this, I will discuss arguments levelled against using autonomous weapons. Particular attention will be given to arguments against their use on the grounds of human dignity. I argue that most arguments against AWS are based on an unjustified assumption that their status in warfare would be that of combatants, given the sometimes over-emphasis or over-hype of their autonomy. If we consider AWS as an instrument, like any other weapon currently in use, most of the issues labelled are either irrelevant or resolved. The argument of dignity is, however, not as easy to resolve. As such, it will be given more attention. Using my Afro-communitarian conception of dignity, I will evaluate whether the use of such technologies undermines dignity. To do this, I will first discuss what dignity looks like in warfare and then evaluate whether the introduction of autonomous weapons undermines dignity.

I argue that dignity is the agent's capacity for communal relationships. How can this be respected in warfare and conflict situations by acknowledging and treating agents based on their moral status? For example, combatants can be targeted by other combatants. However, noncombatants ought not to be targeted as they are not a party to the hostilities. With autonomous weapons, I argue that but themselves they cannot undermine dignity. As instruments, all that can happen is that its users can commit these transgressions, but not the weapon itself. I begin by discussing the laws of war.

5.2. Laws of War

In this section, I briefly outline the principles of just war theory. This is important because it is from these principles, plus my conception of dignity, that we will determine whether the possible deployment of autonomous weapons would undermine dignity.

Within the convention of war, three traditions of thought dominate the ethics of war and peace: realism, pacifism and just war theory(Lazar 2017; Orend 2013). Realists argue

that there is no such thing as a "morality of war" (ibid.); once war has begun, the idea of morality is not applicable, "naked self-interest, survival and the pursuit of power over one's rivals" (ibid.) are all that counts. Pacifists, on the other hand, argue that "no plausible moral theory could license the exceptional horrors of war" (ibid.). The third tradition is just war theory; it seeks a middle path between realism and pacificism, seeking to justify some wars while limiting others. There is a vast amount of literature on these traditions, but for my purposes here, I will focus on the 'just war' tradition.

Just war theory is dominated by two camps: traditionalists and revisionists. Traditionalists are usually also called legalists, as "their views on the morality of war are substantially led by international law, especially the law of armed conflict. They aim to provide those laws with morally defensible foundations" (Lazar, 2017). Some of the issues they advocate are that only states can go to war and under specific conditions; civilians are not legitimate targets of war unless in cases where it is unintended or militarily necessary; and combatants are morally permitted to target other combatants (ibid.). Revisionists "question the moral standing of states and the permissibility of national defence, argue for expanded permissions for humanitarian intervention, problematise civilian immunity, and contend that combatants fighting for wrongful aims cannot do anything right besides lay down their weapons" (Lazar, 2017). They also "deny that the contemporary law of armed conflict is intrinsically morally justified, but believe, mostly for pragmatic reasons, that it need not be substantially changed" (ibid.).

The divide between the two camps does not necessarily mean that one can only be on one side and not the other to highlight the differences in opinion of some principles. I will mostly be arguing from a traditionalist perspective of just war theory because, as things stand, it is from this tradition that IHL emerges. This means that these principles are currently used to evaluate the war convention.

A 'just war' is a war that is fought to maintain some form of justice in that the hostilities in question do not regress to all-out war. Actors must adhere to certain principles to engage in such a just war. There are three main categorisations of the principles of a just war: *jus ad bellum*, *jus in bello* and *jus post bellum*. *Jus ad bellum* principles are concerned with a set of conditions that political actors must consider before engaging in war to determine whether entering the war is morally and legally permissible. There are six necessary conditions under *jus ad bellum* that a state (or other political actors) needs to fulfil before going to war. These are Just Cause, Legitimate Authority, Right Intention, Reasonable Prospects of Success, Proportionality and Last Resort (Necessity) (Lin, Bekey and Abney, 2008; Orend, 2013; Abney, 2015; Lazar, 2017).

Jus in bello refers to the principles and laws concerning the acceptable conduct of agents during a war. Unlike *jus ad bellum*, where responsibility rests on the state (or the political actor instituting the conflict), with *jus in bello*, responsibility "rests with those commanders, officers and soldiers who command and control the lethal force set in motion by the political hierarchy" (Walzer, 2015, p.39). Within the traditionalist perspective (and what was codified as IHL), there are three main principles: discrimination, proportionality and necessity (Lazar, 2017). These principles serve as necessary conditions by which combatants must abide. *Jus post bellum* principles involve the law concerning acceptable conduct following the official or declared war's end. It is a relatively new field, so there is not much literature regarding the principles categorised under it. These principles – *jus ad bellum*, *jus in bello*, and *jus post bellum* - collectively cover the institution of war from when it starts until it officially ends and must be followed for a war to be considered a just war. Brian Orend (2013) also proposes the following moral principles for consideration: proportionality and publicity, rights vindication, discrimination, punishment, compensation

and rehabilitation. He suggests these principles would benefit the countries who engaged in war after the war. As such, they must be considered part of 'just war theory'.

Recently, there has also been a development in the philosophical literature on just war theory from an African point of view (see Metz, 2019; Okeja, 2019; Ugwuanyi, 2020; Cordeiro-Rodrigues, 2021). I will briefly outline some key arguments for this viewpoint. Luis Cordiero-Rodrigues (2021) provides an overview of contemporary African perspectives on a just war. He outlines the main arguments concerning the moral justifications to initiate or abstain from war (*jus ad bellum*) by looking at works from 20th and 21st- century African philosophers. He (ibid.) argues that within African thought, there are four kinds of arguments for going to war:

> Firstly, consequentialist rationales for starting a war constitute the dominant perspective in African thought. Secondly, some African thinkers uphold the idea that to initiate or abstain from entering a war; one needs to look at the therapeutic impact that war or peace may have and, from this impact, then deliberate on whether it is justified to enter a war. A third category of arguments defends that entering war is justified if it responds to a wrong committed. Finally, some African philosophers uphold the idea that wars are justified if they have been agreed upon under certain conditions (Cordeiro-Rodrigues 2022, p. 9).

Accordingly, it is for these reasons that the bulk of theorisation in the African context has emanated. In this text, Cordiero-Rodrigues only focuses on literature that addressed the moral justifications of war (*jus ad bellum*), arguing that that is where a large amount of literature emanates. I agree that the bulk of the literature on warfare in African philosophy focuses on *jus ad bellum.* Although theories on the moral justification of war are important, they are not particularly relevant here since my evaluation is based on the use of weapons. This means that by this time, all the necessary justification has been conducted, and the

hostilities are active. There are, however, some theorists who make provision for the principles of going to war. Uchenna Okeja (2019), for example, outlines the African motivation to go to war (*jus ad bellum*) and the conduct of war (*jus in bello*) to address the question: what makes a war morally justified?

Looking at pre-colonisation wars in Africa, Okeja (2019, p.195) argues that "consensus is the basis for the determination of the morality of the recourse to war in traditional African contexts". Essentially, he argues that even in the different types of conflicts that occurred in different military organisations, what was common about these conflicts was that chiefs and council members had to consent to any decision to go to war. The motivation for this was the just cause claims that required deliberation and agreement by the parties affected or their representatives before such a war could ensue. He (ibid.) continues to argue that the conduct and means of war at this time had flexible moral requirements. Essentially, there was no legislation regarding how the conflict could be conducted or what means were permissible. All means accessible to the warring parties to subdue the enemy could be used, ranging from "charms, talismans, amulets, chief priests" (ibid., p. 194), to name a few. He thus summarises such conduct of war as an endorsement of '*limited realism*', meaning that "considerations of what is permissible or impermissible in war is confined to the domain of expediency, taboos and customs" (ibid., p. 195).

The idea proposed by Okeja articulates rules that have to do with the justification of war but not rules in terms of the means by which war will be fought. On these grounds, the use of any weapons would not be considered a violation since the conduct of war is not evaluated. Lawrence Ugwuanyi (2020) articulates the idea of war differently. Looking at wars fought in Africa post-1957, Ugwuanyi (2020) articulates an African theory of just war. He (ibid., p. 51) argues that "a just war in African thought is a war fought to protect the corporate harmony of a people who are bound and bonded together through the land,

resources, and other symbols and traditions that make them distinct". Essentially, the primary interest of African communities is peace and harmony, not war. So, even in cases where they must resort to war, they would have to abide by certain principles. He lists four fundamental principles of a just war in Africa: 1) last means of addressing conflict; 2) proportional means and ends; 3) participatory pain; 4) harmony, arguing that these principles are informed by the norms, values and principles that drive philosophical inquiry in Africa and thus constitute a just war.

Of these principles, the two that are associated with the conduct of warfare are proportional means and participatory pain. The principle of proportional means and end suggests that war is not meant to annihilate the other party. Instead, it must be "executed in a manner that would provide for and make up for the harm that caused the dispute in question and lead to what can be called a measured restoration"(Ugwuanyi 2020, p. 62). So, in fighting a war, parties involved must conduct themselves in a way that, by the end of the conflict, no one would be motivated for reprisal. Related to this principle is the principle of participatory pain. This principle emphasises that "war in Africa observed such ethics that forbade the total annihilation of an opponent" (ibid.). This means that in a way hostilities are conducted, actors must not aim at the total eradication of others. They should not use weapons that cause more harm than necessary, they should be willing to negotiate ceasefires, and the defeated should not be tempted to engage in revenge conflicts. These two principles emphasise that the conduct and means by which hostilities are conducted must always ensure the most minimal damage in acquiring whatever goals the hostilities envisage. So whatever weapon is used cannot, for example, be indiscriminate or annihilate the other party. It must be targeted with the aim of minimal harm and suffering.

Having briefly outlined the principles of a just war, it is evident that although there are greater similarities in the principles of war, even the divergent views do not dispute that

principles ought to be followed in the convention of war. There is no dispute that there are principles to be followed. Most of the principles here are codified in laws of armed conflict and are currently used to evaluate the justification and conduct of war. It is these principles that will be applied in responding to some of the arguments levelled against the use of autonomous weapons. I discuss these arguments in what follows.

5.3. Ethical Issues Against Autonomous Weapons.

In this section, I discuss the main ethical issues against using autonomous weapon systems. This provides the reader with an overview of the related discourse and controversies. The main issues that will be discussed are that having such technology would provide an incentive to go to war, that their use would undermine the laws of armed conflict, that they would cause a responsibility gap and that their use undermines human dignity. I will respond to these arguments based on the principles outlined above. Specifically, my responses will largely be focused on *jus in bello* principles. This is because the use of weapons happens in active combat after the war has been declared and justified by the parties involved. I argue that most arguments against AWS are based on a misunderstanding of the status of AWS in warfare. As instruments, there are sufficient conventions that can regulate their use. The arguments assume that AWS could have the capacity to become political actors or have human autonomy. Taken as instruments and not agents, most of the arguments fall short.

One of the arguments against AWS is that the development and possible use of these weapons would encourage political leaders to go to war, given the minimal cost of human casualties, thus encouraging other political actors to develop them. The potential diminished human cost would make states less hesitant to launch war campaigns. Such actions would cause another arms race (Sparrow, 2007; Asaro, 2012). I agree with Keith Abney's (2015) response that these arguments are not based on the weapons themselves but on the actions of political actors who have them. These actors are the ones who declare wars; weapons are

merely the means by which a war can be fought. However, in themselves, weapons cannot go to war or sanction one. Thus, no matter how sophisticated the weapon might be, they remain the means, never the actors that can declare war.

Additionally, I argue that if states develop such weapons, they could deter them and political actors from attacking each other. The reasonable prospect of success is one of the requirements in *jus ad bellum.* It aims to discourage lethal advances known to be futile by the aggressor (Orend 2013). Essentially, when the prospect of success is less, it is less likely for one to launch an attack. Such is the case in terms of countries possessing nuclear weapons. Following the bombing of Hiroshima and Nagasaki, there was a consensus amongst many nations that such weapons should never be used in warfare. In addition to this consensus that gave rise to the United Nations as we know it, states without nuclear power have no prospect of attacking others who do because doing so would guarantee the destruction of both parties. This principle is known as “Mutually Assured Distraction”, which is why there has not been and probably will not be a nuclear attack again. Thus, considering the development of autonomous weapons in this regard, such a development would lessen the prospect of going to war in the long run, as no state or political actor would want to attack a country with such military prowess.

The one threat in this regard is that due to the availability of the components, such technologies would not only be available to states, but bad actors would have access to them. Although this is a real threat, it is not a unique threat. There is evidence of many other lethal weapons at the hands of criminals. The obvious solution to this is to tighten the already existing controls. This thinking is similar to the case in which, although individual components of making a bomb are available, purchasing a basket of said components raises the necessary flags to enable timely interventions. So, this is simply an issue of tightening regulations (Müller and Simpson 2014).

Another argument against AWS is that using them would not comply with laws of armed conflict. Noel Sharkey (2014, p. 116) argues that "allowing robots to make decisions about whom to kill could fall foul of the fundamental ethical precepts of a just war under *jus in bello*", specific principles of discrimination and proportionality. Essentially, in a situation where the war is conducted in a place where people live, these weapons would be unable to distinguish between legitimate targets and non-legitimate ones. In a situation where combatants are targeted, Antonio Gutteres (2018) also argues that "the prospect of machines with the discretion and power to take human life is morally repugnant". Similarly, Christof Heyns (2013, p. 10) argues that autonomous weapons lack "human judgment, common sense, appreciation of the larger picture, understanding of the intentions behind people's actions, and understanding of values and anticipation of the direction in which events are unfolding". For him, these qualities are important for making decisions about human lives. Peter Asaro (2012, p. 700) also argues that "the very nature of IHL, which was designed to govern the conduct of humans and human organisations in armed conflict, presupposes that combatants will be human agents".

Additionally, the argument against using AWS is based on responsibility. The argument is that it is unclear who, if anyone, could be held responsible for unlawful acts caused by a fully autonomous weapon: the programmer, manufacturer, commander, or the machine itself. The accountability gap proposed here would make it difficult to ensure justice, especially for victims (Anne Gerdes 2018; Heyns 2013; Roff 2013; N. Sharkey 2014). Accordingly, the issue of accountability is significant because IHL "emphasise[s] a clear chain of accountability as a prerequisite" (Gerdes, 2018, p. 238). Heyns (2013, p. 14) also argues that "robots have no moral agency and, as a result, cannot be held responsible in any recognisable way if they caused the deprivation of life that would normally require accountability if humans have made the decision". It thus makes it difficult to ascribe

responsibility, especially considering N. Sharkey's (2014, p.117) argument that "you cannot punish an inanimate object".

In response to the arguments above, I argue that these arguments are based on an unsubstantiated assumption about the moral status of autonomous weapons. There seems to be an assumption that AWS are moral agents. This assumption, I argue, is made possible by the label of 'autonomous' given to these weapons, which creates an impression that they are autonomous in the strict sense that we understand autonomy. Strictly speaking, an autonomous agent is a self-governing agent with the capacity of will. Because of their autonomy, the agents can be held responsible for their actions. In the context of warfare, *combatants* are considered autonomous agents insofar as, although they act based on orders, they can *reject* an order if they consider it immoral or illegal. This capacity to reject orders is a sign that they can exercise their *own* judgment outside of prescribed orders. Their autonomy also empowers them with the capacity to target and kill. However, autonomy, in the sense of autonomous weapons, would refer to robots having the capacity to perform certain functions without *direct* supervision. In attack contexts, autonomous weapons can *only* act based on their programming and a certain set goal. The only thing is there would be no way of knowing when and how it will initiate this goal. So, they are autonomous in that they have *some* decision capacity. This kind of autonomy is weak and not one that is attributed to combatants. (see also Coeckelbergh 2011, p.273)

If we assume that their status is that of agents, these weapon systems would have to comply with *jus in bellum* principles; they would be capable of being political actors, required to practice judgement, held responsible for their actions, and all the other demands that we have on the behaviour of combatants in warfare. However, there is insufficient evidence to suggest that AWS can or would assume the status of agents of combatants in warfare. However, if we accept autonomous weapons as *instruments*, like other weapons,

albeit more sophisticated, then, regarding compliance with laws of armed conflict, their use would only need to comply with already existing weapons treaties[22] and conventions. Such rules would ensure they do not act indiscriminately in their missions and have higher target precision. They can be fitted with cameras and assist in recording hostilities. Other combatants' awareness of being recorded might mitigate war crimes and be used as evidence when transgressions are committed.

Regarding responsibility, as instruments, then responsibility pertaining to their use would befall the agents using them. Since AWS would be programmed based on the instructions of the commander in charge, the responsibility would lie with the commander using a principle known as 'command responsibility'. It states that:

> Commanders and other superiors are criminally responsible for war crimes committed by their subordinates if they knew or had reason to know that the subordinates were about to commit or were committing such crimes and did not take all necessary and reasonable measures in their power to prevent their commission, or if such crimes had been committed, to punish the persons responsible (Walzer 2015, p. 317)

Therefore, the commanding officer is the first locus of responsibility regarding the conduct of war. As such, they can be held responsible for the conduct of their combatants and the arsenal at their disposal. In every declared conflict, a commander is appointed and made responsible for carrying out the interests of their representative state. The commander takes charge of the entire infantry, from the combatants to medical staff and weapons.

[22] The International Committee of the Red Cross (ICRC) details the list of these treaties. For a list, please refer to https://www.icrc.org/en/document/weapons

Thus, Michael Walzer (2015) argues that nothing happens or should happen in war without the commander's knowledge. It is because of this that she bears the primary responsibility. Also, as an accounting officer, she could hold programmers responsible if they misunderstood the instructions or decommission the AWS if she realises it is faulty. Essentially, the commander is responsible for all actions in the field. They even have the power to hold others responsible. It is for this reason that they are the locus of responsibility.

Thus far, I have responded to the three arguments against using autonomous weapons. I argued that many of the arguments provided tend to misunderstand and exaggerate the role AWS would play in warfare in three ways. Firstly, autonomous weapons are not political actors. As such, they cannot encourage or discourage anyone's prospect of going to war. Secondly, although they might have some 'autonomy' only in terms of their functions, they are not autonomous in a human sense. Even if they could be autonomous in the human sense, what would be the motivation to want to make them autonomous? Thus, it is likely that their status will be as instruments, perhaps with extended capabilities, but not combatants. This implies that their use will be evaluated in the same way all other weapons are. Lastly, as instruments used in war, all responsibility related to their actions would befall their user. Although there could be a need to evaluate the existing weapons treaties to ensure that in their use, AWS does not contravene the principles of a just war, none of the arguments proposed against them is sufficient to justify a claim for an outright ban. If the weapons treaties are insufficient, we can rely on existing just war theory principles to work as a buffer. It is apparent that the arguments levelled against autonomous weapons are based on unsound claims. In fact, based on the existing arguments, it seems we would benefit from having such technologies. The trickier argument to deal with is the one on dignity, which I will now delve into next (see also Muller 2016; 2021).

Regarding dignity, several authors and organisations argue against the use of autonomous weapons on the grounds that their use would undermine human dignity. Stephen Goose and Mary Wareham (2016, p. 31) argue that "[f]ully autonomous weapons would also undermine human dignity, because as inanimate machines they could not understand or respect the value of life, yet they would have the power to determine when to take it away" (see also Docherty, 2014, and Sharkey, 2014). A report by the International Committee of the Red Cross argues that a fundamental moral line is crossed when machines make life-and-death decisions. According to this view, "it matters not just if a person is killed or injured but how they are killed or injured" (ICRC 2018, p. 10). Christof Heyns (2017) also argues that beings targeted by autonomous weapons make targeted agents merely objects that can be targeted with impunity. Such an act undermines their intrinsic capacity, which ensures their respectful treatment. When targeted by autonomous weapons, he (ibid., p. 63) argues, the agents "have no avenue, futile or not, of appealing to the humanity of the enemy, or hoping their humanity will play a role because it is a machine on the other side". By this logic, human beings are subjected to autonomous weapons, and they are reduced to objects that could be used to fulfil a particular goal. Their lives are not respected, and their value is viewed as instrumental instead of intrinsic.

Another argument against the use of autonomous weapons can be located in the works of Ozlem Ulgen (2018, p. 1), who argues that autonomous weapons would undermine dignity in two ways: Firstly, he argues that the "use of autonomous weapons denies the right of equality of persons and diminishes the duty not to harm others, [and secondly that] replacing human combatants with autonomous weapons debases human life and does not provide respectful treatment". The first argument implies that by replacing combatants with autonomous weapons, the value of the combatants in hostilities is debased since others are removed from physical risk. Ulgen bases the second argument on Kant's principle of

humanity and argues that using "[a]utonomous weapons would devalue humanity by treating humans as disposable inanimate objects rather than ends with intrinsic value and rational thinking capacity" (ibid., p. 10). By such treatment, the respect of those being targeted would be undermined. For other arguments that pertain to dignity and autonomous weapons, see Asaro (2012) and Johnson & Axinn (2013)

In response to some of the arguments above, Ariadna Pop (2018) argues that with the many arguments provided on the grounds of dignity, there is very little reflection on how human dignity is conceptualised to warrant the conclusion that allowing a machine to decide to kill a human is against human dignity. She argues that "in anti-AWS arguments, the concept is typically invoked in a rhetorical manner, in the form of an emotional charge or as a conversation stopper of sorts". Pop (2018) also notes that even among those who invoke Kant's conception of dignity, none of their arguments show how exactly the employment of AWS violates human dignity. She (ibid.) argues that "[u]nless it can be shown why the concept of human dignity mandates the prohibition of AWS; such a prohibition is not justified".

Essentially, even if we accept that being killed against our will affects our autonomy, this issue is not unique to autonomous weapons. This is the case with any other lethal weapon (see also Sharkey, 2019). Pop (2018) concludes that "[a]nti-AWS arguments are therefore either obscure by drawing unjustified inferences or if the inferences can be justified, they must be based on a different conception of human dignity than the one that seems to be endorsed". The complexity is caused mostly by the difficulty to measure dignity in warfare because "[w]ar itself takes a toll on human dignity through the intentional sacrificing of lives to achieve military objectives" (ibid.). It is further complicated by the fact that soldiers are part of the war convention, meaning that taking their lives does not constitute a violation per se. Pop (2018) recommends that AWS "should not be viewed as an inherent violation of

human dignity due solely to the weapon's autonomy" (For a similar argument see Saxton, 2016).

Another similar argument in defence of AWS is made by Garry Young (2021, p. 478), who argues that killer robots would not disrespect dignity because permitting them "does not deny combatants the opportunity to fight back against these automated weapons and therefore act as moral agents". So, a combatant can still perform their task as a combatant regardless of the entity targeting them.

Thus far, I have highlighted the arguments made against the use of autonomous weapons on the grounds of dignity and the rebuttals to said argument. In summary, the main arguments against the use of AWS based on dignity are:1) AWS, as inanimate machines, do not understand or respect the value of life, so they should not have the capacity to take a life because having such a capacity undermines the value of that life; 2) being targeted by AWS treats one's life as a mere object; 3) denies the equality of combatants. The arguments in rebuttal are as follows: 1) most appeals of dignity do not explain what dignity is; 2) even those that appeal to Kant's conception, in how they phrase their argument, the challenge to dignity is not unique to AWS; 3) the use of AWS does not in any way alter the functions of a combatant and 4) how can one measure dignity in warfare?

Firstly, I agree with the argument that autonomous weapons are inanimate machines and, thus, incapable of understanding. I would even go as far as to argue that they are incapable of other cognitive capacities, such as undermining, (dis)respecting, or violating, given that all these require an entity to have cognitive capacities. Currently, autonomous weapons do not have cognitive faculties. At best, they would have functional capabilities that resemble agency - they can do their best to initiate the task they are programmed to initiate. They would be good with the 'what' of a task (identifying a target) and the 'how' (targeting them in the most optimal way), but *not* the 'why'; autonomous weapons cannot develop

motive or intention. It is why those who argue against their use say they should not make 'kill decisions'. The claim that autonomous weapons should not make 'kill decisions' is based on the misconception I have outlined above, that there is an assumption that these weapons have moral agency. This is currently not the case. Even in debates about 'artificial moral agency' (AMA), there are conditions set up on when technologies can be regarded as moral agents, and no technology currently meets said criteria. Therefore, I want to emphasise that no AI technology can decide who to target by itself. At best, autonomous weapons can initiate decisions programmed in them, meaning that they can only act after a kill decision is made.

The commander defines the decision to target and whom to target and is programmed in the weapon. Their 'autonomy' in this sense is only in terms of how and when set orders would be executed. So, the type of understanding related to target decisions still lies in the command centre. Similar to human combatants, they do not get to define who the enemy is. This is done by political actors and implemented by commanders through them. Combatants only discern who should be targeted, as would autonomous weapons. Conceptually, it does not make sense why those who argue on these grounds would state that weapons are incapable of understanding and later accuse them of tasks that require an entity to have the capacity of understanding in order to execute them. It appears these arguments are based on the authors' inaccurate assumption that autonomous weapons have certain cognitive capacities. They, however, do not provide a reasonable justification for why they believe so.

On the issue of how combatants are targeted according to the traditionalist view of just war theory, specifically Walzer (2015), it should not matter how combatants are targeted. The convention relating to killing combatants states that it should matter how combatants are killed in warfare, which is true but not as fundamental to the war convention. I agree with Walzer (2015, p. 42) that "rules specifying how and when soldiers can be killed [are] by no means unimportant, and yet the morality of war would not be radically transformed were they

to be abolished altogether". Essentially, although it is important to evaluate methods of killing to ensure that we avoid the unnecessary suffering of those being targeted, not doing so does not undermine the necessary conditions of a just war, which is non-combatant immunity. That is why different actors apply different weapons to target their 'enemy' and to avoid unnecessary suffering; weapons that cause suffering, e.g., exploding bullets, are prohibited. So, when it comes to the rights of combatants in warfare, we must always remember that they cede their right not to be targeted.

The only prohibitions regarding combatants refer to them not being subjects of torture or subjecting others to it, as prisoners of war, being treated respectfully and treating others respectfully, and that their corpses cannot be violated in any way. For example, beheading and parading the heart on a stake is prohibited. These conditions are indicative of the principle of the moral equality of soldiers. I acknowledge that the principle of moral equality of combatants is a hotly debated issue in just war theory. For example, Jeff McMahan (1994) argues that by distinguishing between 'just combatants' – those that fight in defence and 'unjust combatants' – combatants of the aggressor, Walzer's claim of 'liability to be attacked seems implausible. Essentially, posing a threat to someone is not a sufficient condition to lose one's right to life because some threats are for good reasons. Another argument is based on Walzer's (2015) claim that combatants *consent* to hostilities. In this case, the argument goes:

> Unjust combatants have something to gain from waiving their rights against lethal attack, if doing so causes just combatants to effect the same waiver. And on most views, many unjust combatants have nothing to lose, since by participating in an unjust war they have at least weakened if not lost those rights already. Just combatants, by contrast, have something to lose, and nothing to gain. So why would combatants fighting for a just cause consent to be harmed by their adversaries, in the pursuit of an unjust end? (Lazar, 2017)

I think the argument of just and unjust combatants can be tricky, especially considering that such an evaluation is either not made by the combatants in question or is made after the fact. It is usually not the combatants who distinguish whether they are just or unjust, but rather, this distinction is made by political actors. They are motivated based on either defending the countries from imminent attack or defending their country pre-emptively. For example, in the war against Russia and Ukraine, although it was Russia that invaded Ukraine, the Russian combatants are said to be defending Russia from future hostilities. Similarly, Ukrainian combatants are defending Ukraine from an aggressor who invaded their lands. So, both parties are convinced that their cause is just, although the burden of proof lies with Russia as an aggressor. Only after the fact can an evaluation be made to see whose cause was 'just'. However, one consistent thing is that when combatants enter hostilities, they are aware of their status as combatants and the rights and limitations of said status.

Regarding non-combatants[23], the laws of war state that non-combatants should not be targeted. Unlike combatants who are participants in the hostilities, non-combatants are the unfortunate victims of warfare. That is why they should not be targeted unless it is necessary. Even then, there must be clear justification showing that targeting them was unintentional. If it is intentional, there must be proof of necessity and proportionality of the attack in relation to the political aims. So, these principles would be applicable to autonomous weapons as they are applicable to other combatants. They should be deployed only when they can discriminate between legitimate and non-legitimate targets.

[23] I use the term ''non-combatant'' to specify other actors in war who do not have the status if combatants such as civilians, medical personnel, priest etc.

So, on the point that autonomous weapons should not make life-or-death decisions, I reiterate that they do not. It is the commanders and programmers who program them that actually do. AWS only execute decisions already made. Coupled with this point is the point about AWS targeting combatants. I have argued that foundational to the war convention is more ensuring civilian immunity than worrying about how combatants are targeted. As members of a war convention who are aware of what they are getting themselves into, I agree with Young (2021) that deploying autonomous weapons does not transgress any of their roles or rights. Combatants always have the power to defend themselves when targeted. Additionally, they have the right to target other combatants or strategic sites to achieve their ambitions. Another point I wish to emphasise is made by Pop (2018) and Sharkey (2019), who say that autonomous weapons causing harm in warfare are no different from any other currently used weapon. In this regard, autonomous weapons do not pose any novel challenges to the convention of war. So, calling for their ban on the grounds of this argument would be like calling for a ban on all weapons in use in warfare.

Returning now to the core of the thesis, the two other arguments that I wish to discuss in greater detail have to do with the conception of dignity in warfare. Pop (2018) argues that most authors who evoke the idea of dignity in the context of AWS do not elaborate on what they mean about it, and those who do (mostly appeal to the Kantian conception) do not show how exactly the deployment of AWS violates human dignity. I partly agree with her because the reports highlighted above invoke the concept but do not dwell on what precisely dignity is. They only dwell on how deploying AWS could undermine dignity (see ICRC 2018, Heyns 2017 and Goose and Wareham, 2016). However, some labour is needed to explain what they mean by dignity. For example, Ulgen (2018) evokes Kant's conception of dignity whereby she argues that deploying AWS would undermine the dignity of combatants in that the combatants targeted would be treated as mere means; that replacing human combatants with

machines creates a hierarchy on the consideration of dignity among combatants in that some persons' lives are more valuable than others and thus can be protected and removed from the dangers of the battlefield and lastly that being targeted by autonomous weapons is tantamount to being treated as an inanimate object.

Although I agree with Ulgen's analysis of Kant's conception of human dignity, I would like to highlight a few issues raised above. Firstly, she gives so much credit to the capabilities of autonomous weapons that she does not realise that assuming autonomous weapons are instruments and not agents, the criticism about their harm would apply to other weapons (see Sharkey, 2019 and Pop, 2018). Secondly, it is difficult to measure the moral status of agents in warfare because they vary, and with that, so do their privileges. Political actors can wage war, but mostly, they do not fight it themselves. They are thus absolved from hostilities but bear the responsibility of every life lost, justifying the cause and equipping their soldiers to ensure a reasonable prospect of success. Commanders give instructions to combatants. They are thus mostly shielded from hostilities but bear the greatest responsibility for what happens on the battlefield. The combatants have their role of engaging in hostilities; however, the cost of such right in their lives. Combatants have the right to kill, which is different from all other actors.

These differences in status among the actors in war mean that their treatment is not the same. For example, according to laws of armed conflict, killing combatants in a theatre of war is permissible. Although it is permissible to target combatants in war, there are certain limitations. For example, the principle of 'no means mala in se' suggests that means of targeting cannot be evil in themselves. One cannot torture combatants to meet any political ends as this creates harm to bodily integrity that goes beyond the principle that combatants are liable to be killed. Also, prisoners of war cannot be subjected to inhumane treatment. The fact that their status has changed from 'combatant' to 'prisoner' suggests that their treatment

ought to change as well. These are well-articulated prohibitions in IHL related to the status of combatants. So, when we refer to the equal status of combatants in warfare, we refer to the rights and privileges they share, not the means of fighting. The rights in question here is that every combatant has the right to target others. In the same breath, they are liable to be targeted. Non-combatants also hold different statuses that grant them certain privileges. For example, they ought not to be targeted, and in cases where they are, there must be proof that targeting them is proportional to the military objective. One must justify that targeting them was necessary to justify certain ends.

What is important about Ulgen's argument is that it shows the importance of dignity in evaluating the morality of war. However, her appeal is that her accounts emphasise the means of attack instead of the fact that combatants are liable to be attacked. In the general theatre of war, she does not outline the implications of dignity, given that different actors hold different statuses, which might mean different entitlements and limitations.

As I have argued above, the idea of dignity in warfare is tricky but not impossible to evaluate. Some of the issues can be responded to by appeal to the war convention, but more systematic work needs to be done on the matter of dignity. In the next section, I am to evaluate the impact of autonomous weapons on dignity using my conception. I will first explain what we mean by dignity in warfare using my Afro-communitarian conception of dignity and show how we can measure dignity in the context of war.

5.4. Afro-Communitarian Evaluation

In this section, I wish to evaluate whether the use of autonomous weapons would undermine dignity. As I have shown, A plausible African conception of dignity that is suitable for evaluating the impact of AI technologies on human dignity must be inclusive and based on capacity, protect individuals' rights, promote communal values, and ensure respect for all

with capacities. My conception of dignity argues that one has dignity insofar as they have the capacity to be part of or party to communal relationships as subjects or objects. Their possession of dignity entitles them to respectful treatment by other agents. Respectful treatment in this regard means acting to encourage a harmonious relationship. Doing so would be indicative of respecting the dignity of others, and doing the opposite would undermine dignity. In ideal conditions like peaceful times, evaluating or imposing harmony is not as difficult. Many states have principles and laws that morally legislate how we should engage with one another, and there are punishments for transgressing said behaviour. The same cannot be said about warfare.

So, what does dignity look like in warfare? Furthermore, under which conditions would the introduction of autonomous weapons undermine dignity?

I will first endeavour to answer the question of what dignity looks like in how warfare is conducted. I will do this by evaluating whether the current principles of the conduct of warfare (*jus in bello)* guarantee dignity to the affected parties in warfare. Based on this, I argue that all agents with a capacity for communal relationships have dignity. The difference among them is based on the different ways they ought to be treated. Dignity for combatants would be best represented by the doctrine of 'moral equality of soldiers', meaning combatants treat each other with respect, knowing that they are each target to one another. Respecting noncombatants would be ensuring that they are not targeted, and in the event that they are, it is unintended or targeted as collateral damage.

Regarding autonomous weapons, I will then evaluate whether introducing them would undermine or enhance dignity. I argue that autonomous weapons in themselves cannot undermine dignity. As instruments of war, the focus of evaluation should be placed on their users and how they use them because it is from the decisions of said users that the autonomous weapon would act.

Dignity in warfare

The principles set out in 'just war theory' aim to resolve some of the dilemmas that occur in warfare. These principles acknowledge the carnage that comes as a result of war but seek to come up with some permissible actions that would ensure 'just causes to go to war' (*jus ad bellum)*, 'just conduct of fighting a war' (*jus in bello)*, and 'just means of cleanup post-war' (*jus post bellum)*. I have outlined the principles in detail above. Here, I will focus on the principles related to the conduct of warfare, specifically whether these principles guarantee dignity to the agents that find themselves in the theatre of war. Evaluating these principles first will provide clarity of what dignity could look like in warfare, thus enabling us to determine if the introduction of autonomous weapons would radically change any of that. Three main principles are prominent within the traditionalist perspective of jus in bello:

> 1) Discrimination - Targeting non-combatants is impermissible; 2) Proportionality - Collaterally harming non-combatants (that is, harming them foreseeably but unintendedly) is permissible only if the harms are proportionate to the goals the attack is intended to achieve; and 3) Necessity - Collaterally harming non-combatants is permissible only if, in the pursuit of one's military objectives, the least harmful means feasible are chosen (lazar, 2017)

These principles have mostly been codified into the additional protocol of the Geneva Conventions. As I have shown in the previous section, these principles are what mostly constitute laws of war; thus, I will rely on them for my evaluation. So, how do these principles guarantee dignity? In my conception, dignity is guaranteed to all human beings with the capacity for communal relationships. This is still the case in warfare. However, it may be difficult to evaluate whether one's capacity for communal relationships is respected in warfare because war, by its nature, applies unfriendly means to resolve conflict. So, it

means we need a principle that will justify unfriendly actions towards others. Metz's (2021) principle of right action provides a constraint in terms of how to act towards guilty or innocent parties. He (ibid.) argues that acting unfriendly towards a guilty party to resolve unfriendly behaviour or as reparation for past unfriendliness is permissible. This corollary provides us the means to evaluate actions that can be construed as respecting other agents' dignity.

Another major challenge with evaluating dignity in warfare is that the agents have different statuses, which grants them different rights and limitations. Two main actors in warfare are combatants and non-combatants. The principle of discrimination distinguishes them from legitimate targets (combatants) and non-legitimate targets (non-combatants). This is what makes it the most important principle in *jus in bello.* So, in warfare, all agents with the capacity for communal relationships have dignity. Because of this capacity, they must be respected. However, since agents in warfare hold different statuses, to act in a way that shows respect to the agent's dignity would be to respect their rights and duties as prescribed by the principles of warfare (*jus in bello*). That means treating them in a way prescribed by principles of the conduct of war. I will refer to the two most important agents in warfare, combatants and non-combatants.

Combatants have a right to kill other combatants and, at the same time, are also liable to be killed. This principle is known as 'moral equality of soldiers'. So, combatants lose their right to life but gain the right to take a life. Although combatants can be targeted and killed, if they lay their arms in surrender, they cannot be targeted. Also, if they are captured, they can be stripped of their weapons, incarcerated as prisoners of war and verbally questioned, but they cannot be tortured or used for medical experiments. They must be offered basic hygiene and medical treatment. This principle is known as 'benevolent quarantine'.

So, respectful treatment for and by combatants would ensure they only target legitimate targets for the right reasons. The most obvious reason is that combatants must be fighting in defence of something. This is quite a tricky position to evaluate in many contexts because both the aggressor and the defending actor would have the motivation they deem justified. However, it is sufficient normatively to state that there must be a 'just motive' to bear arms because it is to defend oneself imminently, pre-emptively, or as reparation for a past wrong.

In targeting others, combatants must not aim to cause unnecessary pain and suffering. So, the means used should not cause suffering on the other side. If combatants surrender, respect for the principles of the armed conflict suggests that they should not be targeted anymore. They can be captured and kept as prisoners of war. Even during that capture, they should not be subjected to ill-treatment. The moment combatants lay down their arms, they seize to be combatants. Thus, targeting them after the fact would be a mistreatment of their status and thus undermine their dignity. If, after killing combatants, their bodies are desecrated or paraded in celebration, this is a violation of their bodily integrity and goes beyond the scope of targeting combatants for political ends. Such an act can be construed as disrespectful of the role of the combatants.

Non-combatants also have different statuses and, thus, different rights. I use the term 'non-combatants' instead of 'civilian' because the former encapsulates more than civilians. Medical personnel, people who bring food or clear out bodies on the battlefield, or priests all fall under the category of non-combatants because although they are in an area of conflict, they are not engaged in hostilities. Non-combatants are every agent in the theatre of war that does not have the right to kill. Because they lack this right, they maintain their right to life. For that reason, they cannot be targeted in warfare. This is why the principle of discrimination is important: it separates legitimate from non-legitimate targets.

The primary way of respecting non-combatants is not to target them. Not targeting combatants would be respecting their capacity for communal relations. Given that there is no motive that justifies targeting non-combatants, they ought not to be targeted. Only in extreme circumstances might they be targeted and then more in the sense of collateral damage. For example, many conflicts these days occur in places where communities live, and thus, non-combatants could catch stray bullets. In this case, their targeting is an unintended consequence. There are cases where the attacker knows that there will be non-combatant casualties. In this context, the principle of proportionality and necessity must be satisfied to show that the number of non-combatant targets justifies the nature of a target. For example, bombing a convoy that is bringing weapons to others would put the drivers in harm's way but can be justified because it would dislodge or defeat the opponent and, therefore, save way more lives.

The principles of *jus in bello* provided us with a way to evaluate the respectful treatment of different agents in warfare. Now that we understand what dignity looks like in warfare, the question that follows is how the introduction of autonomous weapons would affect the dignity of actors in the context of war. The simple answer is that in themselves, *they would not.* Like any weapon currently in use, how the weapon is used could be subjected to evaluations of dignity. Even in that sense, it would not be the weapon per se that disrespects anyone's dignity but how the user uses it. I have argued throughout this chapter that based on their current definition, the status of autonomous weapons is that of instruments, meaning that in warfare, they are merely sophisticated weapons. In this regard, apart from evaluating the motives and actions of their users, autonomous weapons would just be subjected to a treaty that outlines their uses and prohibitions to ensure that their use does not undermine the agent's dignity.

Evaluating autonomous weapons based on a treaty might be tricky because no such treaty exists at the moment. Even trickier is that it might be harder to develop one because, as far as we know, no such weapon exists. However, civil society groups and some states have been trying to develop such a treaty, and recently, their efforts have been met with some success. On 1 November 2023, the First Committee on Disarmament and International Security of the United Nations General Assembly approved a draft resolution on lethal autonomous weapons systems[24]. The resolution highlights the possible negative consequences and impact of autonomous weapons systems on global security and regional and international stability. It stresses the urgent need for the international community to address the challenges and concerns raised by such systems.

Some of the main themes prevalent in the debate are that the deployment of these weapons must always have 'meaningful human control'. Essentially, their deployment must be within a predictive frame, with the deployers knowing where they are being deployed and in what way. This seems like a standard operating procedure for any weapon that is used. The agents using it are or should be aware of how it works and to whom it must be aimed. So, to ensure that their use does not undermine the dignity of the actors in the theatre of war, it must be aimed at legitimate targets and never with the aim of causing senseless pain. The existing advances in technology suggest that if autonomous weapons were to be deployed tomorrow, they would be deployed as instruments. As such, the conditions outlined above would hold.

Perhaps autonomous weapons could prove sufficient functional autonomy in the future so that they can be deployed as agents. Should this instance occur, their status would have to be reviewed and upgraded from instruments to combatants. Under this new status, they would

[24] For the resolution see: https://documents.un.org/doc/undoc/ltd/n23/302/66/pdf/n2330266.pdf?token=XhLGS7sGRozMFw8Rb5&fe=true.

have the capacity to target other combatants. In this scenario, all arguments discussed above would be worth considering again, looking at the extent to which they are autonomous versus the amount of human control. This, however, should not raise any alarms since even existing combatants are under command, just as the autonomous weapons would be, and the evaluation of its transgressions will be similar to that of combatants. Like combatants, autonomous weapons will undermine dignity if they kill indiscriminately or deliberately target non-combatants and other non-combatants. This is because such an act would be deemed disrespectful to the roles and status of other actors in the theatre of war. However, although they would have the capacity to undermine dignity, in my conceptions, they would still not have dignity themselves and their default status as combatants would make them unfriendly. This is different from human combatants, who assume the role of combatant for a certain period and can cede it.

However, one can also argue that such weapons could assist in maintaining dignity in warfare because their target precision would be higher, meaning the frequency of targeting non-combatants would be minimised. Such weapons would be fitted with cameras, meaning they can record hostilities in real-time, which can help if war crimes are committed and evidence is required (see Arkin, 2008; Müller, 2016; Müller & Simpson, 2014). Like care robots, there can be cases made about the benefits and consequences of autonomous weapons in warfare.

In this section, I evaluated the impact of autonomous weapons on human dignity if they are applied in warfare. The key revelation in this regard is the significant role that dignity plays in warfare, not in terms of who has it but how different agents are to be treated to ensure that their dignity is not undermined. Dignity in warfare is upheld if the rules and principles governing the theatre of war are respected and the status of different actors is also respected. The community of combatants has principles that should be upheld, and doing so

ensures harmony among them when the war is over. Combatants also bear a big responsibility of discerning between those they can target and those they cannot. This, in effect, reduces the number of casualties and avoids future reprisals. Autonomous weapons are merely instruments, so in themselves, they cannot possibly undermine human dignity. However, their user (the commander or programmer) would undermine dignity if they use them in instances where they do not comply with weapons prohibitions or use them in a way that undermines laws of war. Only in future cases where their status would be altered to agents and, therefore, combatants would they be capable of undermining dignity. However, this undermining is in the same way combatants do, and thus, the measures of mitigating such harm could be similar to those of human combatants or even harsher.

5.5. Conclusion

In this chapter, I evaluated the impact the development, deployment, and use of autonomous weapons could have on human dignity. I began by outlining the principles of warfare, which will be used to respond to the criticisms levelled against AWS and evaluate whether their introduction in warfare would undermine human dignity. I then discussed levelled arguments against using autonomous weapons. I argued that most arguments against AWS are based on an unjustified assumption that their status in warfare would be that of combatants, whereas all evidence currently shows them as possibly being deployed as very sophisticated weapons. As weapons, their status will be that of instruments, the same as other weapons, albeit with new capabilities; however, they are still instruments. If we consider autonomous weapons as instruments, most arguments levelled against them are either irrelevant or easily resolved.

The argument of dignity is, however, not as easy to resolve. Using my Afro-communitarian conception of dignity, I first discussed what dignity looks like in warfare. In terms of dignity, I argued that almost all actors in warfare have dignity. To promote their dignity means treating them in ways that acknowledge their status in the context of warfare

and the associated benefits. Among combatants, showing respect would mean acknowledging each other's moral equality and treating each other in ways that recognise this. For example, treat prisoners of war respectfully and never target others with vengeful motives or intention to cause pain. Respecting civilians' dignity would mean not targeting them because although they are caught up in hostilities, they are not party to them. So, combatants must not target them or even use them as shields.

Thus, the promotion of dignity in warfare would be to respect the agent's moral status and treat them in accordance with it. Such behaviour would promote friendliness among the actors in conflicts because, for example, combatants would understand that they can only be hostile to other combatants but not to civilians. Even in the hostilities, they would understand the limits of said hostilities as outlined in the principles of warfare. Actors behaving in a way that acknowledges and respects other actors' moral status would even avoid reprisal attacks, thus ensuring peace after the war is concluded.

Following this, I evaluated whether the introduction of autonomous weapons would undermine dignity. I argued that in their current status as instruments, they cannot undermine dignity. The entity that would undermine dignity would be the users of these if they use them in ways that are deemed unacceptable in a just war. In the event that they are ascribed the status of agents and, with it, combatant status, then they would undermine the dignity, firstly, of other combatants if they were to target combatants who have surrendered or who are held captive or by targeting combatants in a way that causes unnecessary pain or suffering(torture). This is because such targeting would be disrespecting the status of said combatants in warfare, which is unfriendly. In terms of non-combatants, their use would also undermine dignity if autonomous weapons deliberately target them or if such targeting is disproportionate. Targeting noncombatants would be considered unfriendly since they are not party to hostilities. So, targeting them would be disrespecting their status in warfare.

CHAPTER 6: CONCLUSION

6.1. Introduction

In this thesis, I have embarked on a journey to evaluate the impact AI technologies could have on human dignity. This is because dignity is often cited as the reason why certain technologies should not be allowed in certain domains. For example, how would the use of robotics in healthcare would undermine the dignity of patients in need of care? It is argued that using these robots subjects the patients to humiliation, objectification, infantilisation, and deception, denying their privacy and human contact. In the context of military robotics, specifically autonomous weapons, it is argued that using such robots is tantamount to treating the targets as mere objects that could be eradicated and that using such technologies subjects them to being a mere means to an end and not ends in themselves. I embarked on testing the veracity of these claims in these domains as they relate to human dignity.

In the thesis, I sought to answer the following question: *Under which conditions do various AI technologies undermine dignity?* Two prior questions emerged in an attempt to answer the main one: 1) What is dignity? 2) What conception(s) of dignity is plausible for evaluating the impact AI technologies could have on dignity? These two questions have created two main sections of my thesis: a discussion on the concept of dignity and the evaluation of the impact of AI technologies on human dignity.

6.2. What is dignity

On the question of what dignity is, I first discussed the concept of dignity to disambiguate it and come up with a working definition. I found that although the concept of dignity has its complexities, said complexities only highlight its importance. Also, what lies foundationally in its complexity is that authors often apply the ‘concept’ interchangeably with the ‘conception’ of dignity. In distinguishing between the two notions, I argued that dignity is a superlative intrinsic value that human beings have due to some *capacity*, whereas a

conception of dignity is a theory about the concept that specifies what constitutes this high value. It is based on the different bases that conceptions of dignity emerge.

I then proceeded to the second question: which conception of dignity is plausible to evaluate the impact AI technologies could have on dignity? Given that certain conceptions of dignity were invoked to claim that certain AI technologies would undermine human dignity, I discussed in detail these conceptions invoked in the literature, principally Kant's and Nussbaum's conceptions. Regarding Kant's conception, Kant defines dignity as an intrinsic value that persons have because they have rationality. It is this rationality that makes them capable of autonomy, i.e. self-governing entities. Because of their capacities, they are entitled to respectful treatment. I have shown that Kant's conception is one of the most important and widely applied. That is why his definition of dignity is widely accepted.

In terms of AI applications discussed here, Kant's finds greater attention in the domain of autonomous weapons, where, for example, Ozlem Ulgen (2018) applies Immanuel Kant's conception of dignity to argue that, firstly, the use of autonomous weapons denies the right of equality of persons since combatants are removed from physical risk by using autonomous weapons. So, by using robotics instead of combatants, other actors disregard other combatants who have to be in the hostilities physically and risk their lives. She also argues that using autonomous weapons diminishes the duty not to harm others. A similar point is made by Christof Heyns (2017), who asks whether it is not an affront to human dignity if robots have the power of life and death over humans. He (ibid., p. 49) argues that there are reasons, based on a right to life and dignity, why full machine autonomy over critical functions, such as using force against those who may otherwise be targeted, should not be permitted.

I argued that although Kant's conception is compelling, one of its main challenges is that it bases dignity on rationality. By basing dignity on rationality, this conception excludes

those who lack or are perceived to be lacking rationality. This exclusion is especially challenging in the healthcare domain because the larger section of those for whom the AI technologies are meant either lack rationality or are perceived to lack it. For example, in the context of healthcare, where not all people needing care have rationality, if the measure for their dignity and thus their treatment is based on it, then it means they could be ill-treated without any consequence.

Regarding Nussbaum's conception, I have shown that her conception resolves the shortfall in Kant's conception. For Nussbaum (2008, p. 357), "human beings have a worth that is indeed inalienable because of their capacities for various forms of activity and striving". Nussbaum essentially shows that dignity is based on more than one capacity, such as rationality. Other capacities include sentience, imagination, emotion, or the capacity for reason, which she (2011, p.24) terms 'basic capabilities', i.e. "innate faculties of the person that make later development and training possible". These faculties differ from person to person and depend on the world for their full development and conversion into actual functioning.

I have shown that her conception enjoys popularity within the AI domains because it resolves the challenge of exclusion that is levelled against Kant. By extending the scope of agents with dignity beyond just those with the capacity for rationality, she makes room for those who are typically considered to lack rationality, such as children, people with disabilities and older persons. I highlighted the main issue I have with Nussbaum's account, which is that her scope of dignity is so wide that it accommodates other entities besides human beings. Given that rats also have sentience and can be construed as behaving in a goal-driven in Nussbaum's articulation, they would also have dignity.

I concluded that both these conceptions have their strengths and challenges and thus find resonance in different domains of AI application. However, in addition to the challenges

listed above, I have argued that both these and other Western conceptions discussed in this paper (e.g., Laitinen et al., 2019; Schroeder, 2008) are based on individualistic normative frameworks. Western individualist conceptions tend to prioritise individuals over the collective. With that, values of autonomy, rationality and individual rights take precedence. This in itself is not an issue; the issue is the claim that such values hold a universal claim. While universal principles such as human rights are important in communitarian conceptions, duties and obligations towards communities are prioritised.

So, in Chapter Three, I considered other conceptions of dignity that are not individualistic in their nature. I considered African communal conceptions of dignity to try and find which one would be plausible in evaluating the impact AI technologies could have on human dignity. I discussed two communitarian conceptions of human dignity: one provided by Metz (2012, 2021) and the other by Ikuenobe (2016).

According to Metz (20212, 2021), our capacity to be part of or party to communal relationships based on identity and solidarity grants us a superlative non-instrumental value that warrants respectful treatment. One can have this capacity as a subject of communal relationships. Important in this conception is that one has dignity by virtue of *having* the capacity to be a party to or part of a communal relationship, not by being in said relationship. Ikuenobe (2016), on the other hand, uses Metz's conception as a foil to come up with his conception. He argues that dignity goes beyond having a capacity but rather entails the active use of the capacity to ensure a harmonious community. So, essentially, one must act in a way that ensures communal harmony. The basis of his conception is personhood. It is from this active and positive use of one's capacities that one earns and deserves respect.

I first argued that these conceptions provide a unique flavour to the conceptions that are already in the domain. By considering dignity from a communitarian perspective, both these conceptions immediately provided a unique way to evaluate AI technologies. I,

however, preferred Metz's conception because, compared to Ikuenobe's, it is more inclusive. By basing dignity on the capacity for communal relationships as opposed to the use of said capacities, his conception, unlike Ikuenobe's, is accommodating to even those with capacities but cannot or choose not to use them. I also prefer it because, compared to other perspectives discussed in this thesis, it is uniquely positioned as a means between the atomistic conceptions prevalent in the West and the extremely communal in most African societies.

While Western normative conceptions prioritise values such as individual rights and autonomy, and African conceptions prioritise community, obligations and duties, Metz's conception ascribes dignity to the individual but in such a way that the dignity is not for the sake of the individual, but as a capacity for communal relationships. This means that one's dignity is guaranteed as long as one's capacity for communal relationships. Essentially, there are no consequences for *inaction* regarding dignity, but *in action*, one must ensure that said actions are friendly and advance communal relationships.

Although Metz's conception is better than the rest discussed in the thesis, there was still an issue with it. In his formulation, dignity is only granted to agents with a full moral status or what he calls 'subjects of a communal relationship'. This excludes all human beings considered to have a partial moral status or what he calls 'objects of a communal relationship', for example, people with severe cognitive disabilities. This should not matter in ideal situations because most agents have full moral status. However, in the context of AI technologies in healthcare, this distinction is crucial because these technologies are mostly created to cater for people who are considered not to have full moral status. For example, most care robots currently in use are located in hospices where their patients are in need of full-time care. Another example is pet robotics such as 'Paro', which are used for therapeutic purposes for patients with dementia and extreme autism. So, we cannot evaluate the violation of their dignity using a conception that does not consider them to have it.

To resolve this shortfall, I reformulated Metz's conception to include also human beings with a partial moral status. I argued that human beings with a partial moral status but have the capacity for communal relations have dignity on the grounds that they morally matter more than any entity with a partial moral status (based on Metz's gradational scale of moral status); we have the same ontological features, as such we recognise and relate with them in a more meaningful way. So, dignity, in my case, is the capacity to be part of or party to communal relationships as either a subject or object. From my conceptions, it is evident that a plausible conception of dignity that can be applied to evaluate the impact of AI technologies on human dignity must be based on capacity, promote communal values, protect individual rights, be inclusive, and ensure respectful treatment for (almost) all. This reformulation, coupled with Metz's conception of 'right action', provides a sufficiently inclusive conception that is plausible to use in multiple robotics domains to evaluate the impact AI technologies could have on dignity.

6.3. Evaluating the impact of AI technologies

Now, to return to the main research question: under which conditions would various AI applications undermine dignity? Using my reformulated conception of dignity in the domains of care robots (Chapter 4) and autonomous weapons (chapter 5), I argued that there are clear cases where the use of AI technologies, such as care robots and autonomous weapons, would have both a positive and a negative impact on human dignity. If the use of such technology respects agents' capacity for communal relationships, then they would enhance it, but if their use impedes agents' capacities for communal relationships, then they would undermine it.

In the context of care robots, where, in many cases, they are used in controlled but accessible environments, AI technologies would undermine dignity if, in their use, they discourage social connection between the patients and their caregivers, evoke feelings of humiliation in the patient or are deceitful. For example, if the use of robots leads to patients

being distanced from their loved ones, it would dishonour dignity since dignity is honoured in a communal setting through cooperation and solidarity. Another example is that using technologies that rely on deception would be considered deceptive because the general intuition in African moral theory is that any act that is deceitful would undermine dignity because deception constitutes unfriendliness. The idea is always to appraise all parties involved in the decisions transparently. So, technologies that rely on deception would be unwelcome unless there are serious mitigating circumstances where deception is necessary.

However, care robots could also enhance dignity if their use enables patients to connect better with their caregivers and families. For example, an assistive robot that aids a patient with mobility ensures that patients can reach their loved ones with greater ease and would fast-track the means of connection between the patient and their loved ones. If, for example, AI, an automated wheelchair, will make it easier for the patient to reach their loved one, then it would enhance the patient's capacities for communal relationships.

My findings have similarities and differences with other evaluations conducted in the study. We generally come to the same conclusions about the ethical issues. The difference is in the motivations on why an issue is of ethical concern. For example, the most popular issue raised in the care sector is that their successful implementation would limit human contact and result in loneliness. I believe AI applications that encourage distancing would be unwelcome because that would dishonour the dignity that is necessitated by people coming together. It, however, is not as much an issue because, in a communal context, the default is not people distancing themselves from each other. It is actually the opposite. Another example is the question of privacy. While in Western conceptions, privacy is an issue because it violates the patients' rights, in my case, patient care is a collective endeavour, meaning that it is not only the interests of the patient that must be taken into consideration but also the interests of other family members. So, all in all, we arrive at similar conclusions

about the issues but for different reasons. Whilst Western conception would prioritise the patient and their need, my conceptions bring together the needs of the patient with those of the affected family members or community.

In the case of autonomous weapons, I argued that their use would undermine the dignity of their use and disrespect the moral status of agents in warfare. They would undermine the dignity of other combatants if they were used to target combatants who have surrendered or who are held captive or by targeting combatants in a way that causes unnecessary pain or suffering(torture). This is because the principle of warfare prohibits such attacks, and undermining these principles is tantamount to acting in enmity. They would also undermine the dignity of non-combatants if they targeted them. This is because non-combatants are not involved in active hostilities. As such, attacking them would be disrespecting their status as non-combatants. Respecting agents' moral status in the context of a place of conflict would minimise the number of casualties in active hostilities. It would also minimise future conflicts that occur as a result of reprisals. Warfare in itself causes an unfriendly environment, but doing it in a way that respects the rules and principles would minimise unnecessary enmity in the present or the future.

In my approach, I first highlighted that most challenges levelled against autonomous weapons are based on the idea that they would act in warfare as agents. I argued that as far as the literature presents currently, they are merely sophisticated weapons (instruments). In this sense, ethical issues levelled against them are irrelevant, given that they are just weapons. As weapons, the threat they pose to dignity is no different than any other weapon used in warfare. This does not mean that there is no threat to dignity in warfare or that their use would not undermine it. If they are programmed wrong or given wrong targets, such as civilian targets, their use would be undermining, but this would be the weapons per se, but the agent has programmed it. So, to avoid undermining dignity in warfare, agents need to

respect other agents' moral statuses. In this way, even if they use these weapons, they will use them correctly based on prescribed.

6.4. Contribution

In this thesis, I have contributed to the discourse of AI and human dignity in several ways. Firstly, I made a case for exploring non-Western and, specifically, African conceptions of dignity as another plausible way to evaluate challenges attributable to dignity. Having shown the challenges with both the Western conceptions by being too individualistic and focusing on atomistic features such as rights or autonomy, I explored African conceptions, which overly emphasise communal living and values such as duties and obligations.

Secondly, I provided a plausible Afro-communitarian conception of dignity that can also be used to evaluate the impact of AI applications on human dignity. My conception is a reformulation of Metz's conception, which ascribes dignity to entities with the capacity for communal relationships as subjects, i.e. agents with a full moral status. I argued that human beings with a capacity for communal relationships as objects (with a partial moral status) also have dignity because they hold a higher moral status than any other entity with a moral status. Additionally, we recognise and relate with them more meaningfully than we would any other entity because we have similar ontological features.

My conception is plausible in that it situates itself between Western individualistic conceptions and African communitarian conceptions. Being in between means it can benefit from both conceptions' qualities while avoiding their pitfalls. For example, by ascribing dignity to entities with the capacity for communal relations as objects, the conception is able to accommodate more human beings than Kant's, Metz's or Ikuenobe's conception. Protecting individual rights and promoting communal values means it accommodates more communities than the purely individualist or purely communitarian-based conception.

Thirdly, I evaluated the impact of AI technologies on human dignity in the domains of healthcare robots and autonomous weapons. Findings reveal that there are conditions under which AI technologies undermine dignity. In the case of care, if their use would inhibit the patient's capacity for communal relationships by creating a distance between an agent and their community, using deceitful means to provide care, subjecting the patients to objectification and humiliation.

In terms of autonomous weapons, my findings resolved that the greater danger to dignity in war scenarios, where autonomous weapons are espoused to be used, are agents not respecting each other's moral status. If combatants treated combats in ways that were allowed but treated non-combatants in a way that was allowed, this would minimise the number of casualties often experienced in these contexts. This would mean that even the use of autonomous weapons would only be directed at combatants on active duty. Targeting any other entity would be disrespecting their right not to be targeted.

Learning the impact of AI technologies on human dignity would assist us in better regulating the proliferation of robot technologies. Additionally, this knowledge helps us appreciate the values of the community and our duties towards it. Knowing that harmony in the community can be guaranteed by our positive response and contribution to the community will hopefully nudge us to become better community members. So, the danger posed by the possibility of certain AI technologies having an impact on our dignity ought to nudge us to become intentional in positively contributing to our community and caring for one another

6.5. Looking ahead

As we look into the current and future developments of AI technologies, it is important to remember that we create these technologies for *us*. We must thus guard against them being used for malicious reasons. This is as far as I have gone in the thesis to show that my Afro-

communitarian conception of dignity provides another way of evaluating the impact of AI technologies on human dignity. Although I have conducted the evaluation on care robotics and autonomous weapons, I do believe this conception can go a long way in evaluating the impact of AI technologies in other domains such as decision systems, sex robotics, self-driving cars, and essentially any AI technology that in its use could threaten the human dignity. These are some of the avenues I wish to explore in the future, and I recommend that my contemporaries engage with them as well.

However, just writing about the impact of AI technologies as they relate to human dignity is not enough. In addition to exploring the impact of dignity as it pertains to other domains of applications, it is important to consider how some of the recommendations made by this study find expression in contexts where these AI technologies are made, the governments who regulate them and the users. This is important in that it would help to bridge the gap that often exists between the makers of these technologies, the regulators and the users. One of the ways to bridge this gap would be to publish recommendations made on this dissertation in policy briefs and newspaper articles. This would ensure that recommendations become accessible to those in industry and policymakers. Another way to make the connection would be by looking into possible collaborations with industry actors through conferences like the World Economic Forum and United Nations agencies such as UNESCO.

In terms of additional contributions, there has been a growing interest in the field of dignity and African philosophy. It would be interesting to see how this new African literature would challenge or affirm my proposed conception of dignity. I hope this thesis will assist lawmakers when deciding what policy direction to take, to use it as a resource for understanding the ethical challenges that emerge and the importance of dignity as a yardstick from which to measure acceptable and unacceptable technologies.

In conclusion, this thesis has shown that there are clear cases where the use of AI technologies would undermine human dignity. Building on Metz's conception, I argued that human beings have dignity if they have the capacity to be part of or party to communal relations. I used this concept to evaluate the impact AI technologies could have on dignity, particularly on care robots and autonomous weapons. The conditions under which care robots would undermine dignity are if their use would distance the patient from their loved one, where their use would evoke feelings of humiliation either in the patient or their loved ones or when their use relies on deceiving the patients. Such actions are unfriendly and disrespectful to patients with autonomous conditions where the use of autonomous weapons would undermine dignity when they use them in a way that disregards the target agents' moral status. For example, they may use them to target inactive combatants or noncombatants. In addition to discovering how AI technologies could undermine dignity, this thesis showed us that dignity is an important instrument to use when evaluating moral conduct in different domains.

I hope this contribution will go a long way in shedding light on the ethical challenges associated with AI technologies, especially as it pertains to dignity. I also hope the thesis reinvigorates the appetite for the study of dignity and its importance as an equalising concept in society.

BIBLIOGRAPHY

Abney, K. (2014). Robotics, Ethical Theory, and Metaethics: A Guide for the Perplexed. In P. Lin, K. Abney, & G. Bekey (Eds.), *Robot Ethics: The Ethical and Social Implications of Robotics* (pp. 35–52). Massachusetts: MIT Press.

Abney, K. (2015). Autonomous Robots and the Future of Just War Theory. In F. Alhoff (Ed.), *Routledge Handbook of Ethics and War* (pp. 338–351). London: Routledge. https://doi.org/10.4324/9780203107164.ch25

Allsobrook, C. (2023). An African Communitarian Conception of Dignity in Mutual Recognition. In M. Molefe & C. Allsobrook (Eds.), *Human Dignity in an African Context* (pp. 125–154). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-37341-1_7

Anderson, J. A., & Rosenfield, E. (1998). Talking Nets: An Oral History of Neural Networks. *IEEE Transactions on Neural Networks*, *9*(5), 1054–1054. https://doi.org/10.1109/TNN.1998.712193

Anderson, M., & Anderson, S. L. (2007). Machine Ethics: Creating an Ethical Intelligent Agent. *AI Magazine*, *28*(4), 15–26. https://doi.org/10.1609/aimag.v28i4.2065

Anne Gerdes. (2018). Lethal Autonomous Weapon Systems and Responsibility Gaps. *Philosophy Study*, *8*(5), 231–239. https://doi.org/10.17265/2159-5313/2018.05.004

Arkin, R. C. (2008). Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture. *Development*, 121–128. https://doi.org/10.1145/1349822.1349839

Asaro, P. (2006). What Should We Want From a Robot Ethic? *International Review of Information Ethics*, *6*(12), 9–16.

Asaro, P. (2012). On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization Of Lethal Decision-Making. *International Review of the Red Cross*, *94*(886), 687–709. https://doi.org/10.1017/S1816383112000768

Asaro, P. (2016). "Hands Up, Don't Shoot!" HRI and the Automation of Police Use of Force. *Journal of Human-Robot Interaction*, *5*(3), 55. https://doi.org/10.5898/JHRI.5.3.Asaro

Bell, D. (2015). *The China Model: Political Meritocracy and the Limits of Democracy*. Princeton: Princeton University Press.

Bikopo, D. B., & van Bogaert, L. J. (2010). Reflection on Euthanasia: Western and African Ntomba Perspectives on the Death of a Chief. *Developing World Bioethics*, *10*(1). https://doi.org/10.1111/j.1471-8847.2009.00255.x

Boden, M., Bryson, J., Caldwell, D., Dautenhahn, K., Edwards, L., Kember, S., et al. (2017). Principles of Robotics: Regulating Robots in the Real World. *Connection Science*, *29*(2), 124–129. https://doi.org/10.1080/09540091.2016.1271400

Borenstein, J., & Pearson, Y. (2010). Robot Caregivers: Harbingers of Expanded Freedom For All? *Ethics and Information Technology*, *12*(3), 277–288. https://doi.org/10.1007/s10676-010-9236-4

Bringsjord, S., & Govindarajulu, N. S. (2018, April 13). Artificial Intelligence. *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/fall2018/entries/artificial-intelligence/. Accessed 13 November 2023

Claassen, R. (2014). Human Dignity in the Capability Approach. In M. Duwell, J. Braarvig, R. Brownsword, & D. Mieth (Eds.), *The Cambridge Handbook of Human Dignity* (pp. 240–249). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511979033.030

Coeckelbergh, M. (2010). Health Care, Capabilities, and AI Assistive Technologies. *Ethical Theory and Moral Practice*, *13*(2), 181–190. https://doi.org/10.1007/s10677-009-9186-2

Coeckelbergh, M. (2012). "How I Learned to Love the Robot": Capabilities, Information Technologies, and Elderly Care. *Philosophy of Engineering and Technology*, *5*(February 2012), 77–86. https://doi.org/10.1007/978-94-007-3879-9_5

Coeckelbergh, M. (2016). Care Robots and the Future of ICT-Mediated Elderly Care: A Response to Doom Scenarios. *AI and Society*, *31*(4), 455–462. https://doi.org/10.1007/s00146-015-0626-3

Cordeiro-Rodrigues, L. (2022). African Perspectives on Just War. *Philosophy Compass*, *17*(3). https://doi.org/10.1111/phc3.12808

Deng, F. M. (2007). Human Rights in the African Context. In K. Wiredu (Ed.), *A Companion to African Philosophy* (pp. 499–508). Malden, MA: Blackwell. https://doi.org/10.1002/9780470997154.ch42

Docherty, B. (2014). *Shaking the Foundations: The Human Rights Implications of Killer Robots*. https://www.hrw.org/report/2014/05/12/shaking-foundations/human-rights-implications-killer-robots. Accessed 16 August 2023

Döring, N., & Pöschl, S. (2018). Sex Toys, Sex Dolls, Sex Robots: Our Under-Researched Bed-Fellows. *Sexologies*, *27*(3), e51–e55. https://doi.org/10.1016/j.sexol.2018.05.009

Emery, A. E. H. (2002). The Muscular Dystrophies. *The Lancet*, *359*(9307), 687–695.

Eze, I. C. (2002). The Colour of Reason: The Idea of "Race" in Kant's Anthropology. In PH Coetzee & A. Roux (Eds.), *Philosophy from Africa: A Text with Readings.* (pp. 430–459). Cape Town: Oxford University Press.

Eze, M. O. (2008). What is African Communitarianism? Against Consensus as a Regulative Ideal. *South African Journal of Philosophy*, *27*(4), 386–399. https://doi.org/10.4314/sajpem.v27i4.31526

Fasoro, S. A. (2019). Kant on Human Dignity: Autonomy, Humanity, and Human Rights. *Kantian Journal*, *38*(1), 81–98. https://doi.org/10.5922/0207-6918-2019-1-4

Floridi, L., & Sanders, J. W. (2004). On the Morality of Artificial Agents. *Minds and Machines*, *14*, 349–379.

Floridi, L., & Taddeo, M. (2016). What is Data Ethics? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2083), 1–5. https://doi.org/10.1098/rsta.2016.0360

Formosa, P., & Mackenzie, C. (2014). Nussbaum, Kant, and the Capabilities Approach to Dignity. *Ethical Theory and Moral Practice*, *17*(5), 875–892. https://doi.org/10.1007/s10677-014-9487-y

Frank, L., & Nyholm, S. (2017). Robot Sex and Consent: Is Consent to Sex Between a Robot and a Human Conceivable, Possible, and Desirable? *Artificial Intelligence and Law*, *25*(3), 305–323. https://doi.org/10.1007/s10506-017-9212-y

Friedman, B., & Kahn, P. H. (1992). Human Agency and Responsible Computing: Implications for Computer System Design. *The Journal of Systems and Software*, *17*(1), 7–14. https://doi.org/10.1016/0164-1212(92)90075-U

Girling, R. (2021). Can Care Robots Improve the Quality of Life as We Age? *Forbes*. https://www.forbes.com/sites/robgirling/2021/01/18/can-care-robots-improve-quality-of-life-as-we-age/?sh=79596e6a668b. Accessed 28 January 2022

Goose, S., & Wareham, M. (2016). The Growing International Movement Against Killer Robots. *Harvard International Review*, *37*(4), 28–33.

Gordon, J.-S., & Nyholm, S. (n.d.). Ethics of Artificial Intelligence. *Internet Encyclopedia of Philosophy*. https://iep.utm.edu/ethics-of-artificial-intelligence/. Accessed 1 July 2023

Grote, T., & Berens, P. (2020). On the Ethics of Algorithmic Decision-Making in Healthcare. *Journal of Medical Ethics*, *46*(3), 205–211. https://doi.org/10.1136/medethics-2019-105586

Gutteres, A. (2018). Address to the General Assembly. *United Nations*. https://www.un.org/sg/en/content/sg/speeches/2018-09-25/address-73rd-general-assembly. Accessed 15 March 2022

Gyekye, K. (2002). Person and Community in African Thought. In P Coetzee & A. Roux (Eds.), *Philosophy from Africa: A Text with Readings.* (pp. 297–312). Cape Town: Oxford University Press.

Gyekye, K. (2023). African Ethics. In (E. Zalta & U. Nodelman, Eds.)*Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/spr2023/entries/african-ethics/. Accessed 17 March 2024

Hansen, T., & Slagsvold, B. (2016). Late-Life Loneliness in 11 European Countries: Results from the Generations and Gender Survey. *Social Indicators Research*, *129*(1), 445–464. https://doi.org/10.1007/s11205-015-1111-6

Heyns, C. (2013). *United Nations Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions*. https://www.ohchr.org/sites/default/files/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf. Accessed 17 April 2023

Heyns, C. (2017). Autonomous Weapons in Armed Conflict and the Right to a Dignified Life: An African Perspective. *South African Journal on Human Rights*, *33*(1), 46–71. https://doi.org/10.1080/02587203.2017.1303903

Hill, T. E. (1992). *Dignity and Practical Reason in Kant's Moral Theory*. New York: Cornell University Press. https://doi.org/10.7591/9781501735035

Hill, T. E. (2014). Kantian Perspectives on the Rational Basis Of Human Dignity. In M. Duwell, J. Braarvig, R. Brownsword, & D. Mieth (Eds.), *The Cambridge Handbook of Human Dignity: Interdisciplinary Perspectives* (pp. 215–221). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511979033.027

ICRC. (2018). *Ethics and Autonomous Weapon Systems: An Ethical Basis for Human Control?* Geneva. https://www.icrc.org/en/document/ethics-and-autonomous-weapon-systems-ethical-basis-human-control. Accessed 17 April 2023

Ikuenobe, P. (2016). The Communal Basis for Moral Dignity: An African Perspective. *Philosophical Papers*, *45*(3), 437–469. https://doi.org/10.1080/05568641.2016.1245833

Ikuenobe, P. (2018). Human Rights, Personhood, Dignity, and African Communalism. *Journal of Human Rights*, *17*(5), 589–604. https://doi.org/10.1080/14754835.2018.1533455

Johnson, A. M., & Axinn, S. (2013). The Morality of Autonomous Robots. *Journal of Military Ethics*, *12*(2), 129–141. https://doi.org/10.1080/15027570.2013.818399

Kant, I. (1998). *Groundwork of the Metaphysics of Morals*. (M. Gregor, Ed.). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511809590

Kerstein, S. J. (2013). Death, Dignity, and Respect. In *How to Treat Persons* (pp. 29–52). https://doi.org/10.1093/acprof:oso/9780199692033.003.0002

Kerstein, S. J. (2014). Kantian Dignity: A Critique. In M. Duwell, J. Braarvig, R. Brownsword, & D. Mieth (Eds.), *The Cambridge Handbook of Human Dignity: Interdisciplinary Perspectives* (pp. 221–229). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511979033.028

Khosravi, P., & Ghapanchi, A. H. (2016). Investigating the Effectiveness of Technologies Applied to Assist Seniors: A Systematic Literature Review. *International Journal of Medical Informatics*, *85*(1), 17–26. https://doi.org/10.1016/j.ijmedinf.2015.05.014

Killmister, S. (2010). Dignity: Not Such a Useless Concept. *Journal of Medical Ethics*, *36*(3), 160–164. https://doi.org/10.1136/jme.2009.031393

Korsgaard, C. M. (1996). *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9781139174503

Kpessa-Whyte, M. (2018). Aging and Demographic Transition in Ghana: State of the Elderly and Emerging Issues. *The Gerontologist*, *58*(3), 403–408. https://doi.org/10.1093/geront/gnx205

Laitinen, A. (2007). Sorting Out Aspects of Personhood: Capacities, Normativity and Recognition. In H. Ikaheimo & A. Laitinen (Eds.), *Dimensions of Personhood* (pp. 248–270). United Kingdom: Imprint Academic.

Laitinen, A., Niemelä, M., & Pirhonen, J. (2019). Demands of Dignity in Robotic Care: Recognizing Vulnerability, Agency, and Subjectivity in Robot-based, Robot-assisted, and Teleoperated Elderly Care. *Techne: Research in Philosophy and Technology*, *23*(3), 366–401. https://doi.org/10.5840/techne20191127108

Lamola, M. J. (2021). An Ontic–Ontological Theory for Ethics of Designing Social Robots: A Case of Black African Women and Humanoids. *Ethics and Information Technology*, *23*(2), 119–126. https://doi.org/10.1007/s10676-020-09529-z

Lazar, S. (2017). War. In (E. N. Zalta, Ed.). Stanford Encyclopedia of Philosophy . https://plato.stanford.edu/archives/spr2017/entries/war/. Accessed 17 April 2023

Lin, P. (2015). The Right to Life and the Martens Clause. *Calpoly*. http/www.ethics.calpoly.edu/ccw_testimony.pdf

MacIntyre, A. (1984). *After Virtue* (2nd ed.). Notre Dame: University of Notre Dame Press.

Macklin, R. (2003). Dignity is a Useless Concept. *British Medical Journal*, *327*(7429), 1419–1420. https://doi.org/10.1136/bmj.327.7429.1419

Maiyane, K. (2023a). Robots and Dignity: An Afro-communitarian Argument in Eldercare. In E. Imafidon, M. Tshivhase, & B. Freter (Eds.), *Handbook of African Philosophy* (pp. 1–16). Switzerland: Springer Nature. https://doi.org/10.1007/978-3-030-77898-9_49-1

Maiyane, K. (2023b). Autonomous Weapons and the Future of Warfare in Africa. In Uchenna Okeja (Ed.), *Routledge Handbook of African Political Philosophy* (pp. 304–314). London: Routledge. https://doi.org/10.4324/9781003143529-26

Matolino, B. (2009). Radicals Versus Moderates: A Critique of Gyekye's Moderate Communitarianism. *South African Journal of Philosophy*, *28*(2), 160–170. https://doi.org/10.4314/sajpem.v28i2.46674

Matthias, A. (2020). Dignity and Dissent in Humans and Non-humans. *Science and Engineering Ethics*, *26*(5), 2497–2510. https://doi.org/10.1007/s11948-020-00245-x

Mattias, I. (2019). Recognition. *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/recognition/. Accessed 29 April 2024

Mattson, D. J., & Clark, S. G. (2011). Human Dignity in Concept And Practice. *Policy Sciences*, *44*(4), 303–319. https://doi.org/10.1007/s11077-010-9124-0

McDougall, R. J. (2019). Computer Knows Best? The Need for Value-Flexibility in Medical AI. *Journal of Medical Ethics*, *45*(3), 156–160. https://doi.org/10.1136/medethics-2018-105118

McMahan, J. (1994). Innocence, Self-Defense and Killing in War. *Journal of Political Philosophy*, *2*(3), 193–221. https://doi.org/10.1111/j.1467-9760.1994.tb00021.x

Menkiti, I. A. (1984). Person and Community in African Traditional Thought. In R. A. Wright (Ed.), *African Philosophy: An Introduction* (pp. 171–182). Lanham: University Press of America.

Metz, T. (2012). African Conceptions of Human Dignity: Vitality and Community as the Ground of Human Rights. *Human Rights Review*, *13*(1), 19–37. https://doi.org/10.1007/s12142-011-0200-4

Metz, T. (2014a). Dignity in the ubuntu tradition. In *The Cambridge Handbook of Human Dignity* (pp. 310–318). Cambridge University Press. https://doi.org/10.1017/CBO9780511979033.037

Metz, T. (2014b). African Values, Human Rights and Group Rights: A Philosophical Foundation for the Banjul Charter. In O. Onazi (Ed.), *African Legal Theory and Contemporary Problems*, (Vol. 29, pp. 131–151). Dordrecht: Springer. https://doi.org/10.1007/978-94-007-7537-4_7

Metz, T. (2019). An African Theory of Just Causes for War. In L. Cordiero-Rodrigues & D. Singh (Eds.), *Comparative Just War Theory. An Introduction to International Perspectives* (pp. 131–155). London: Rowman and Littlefield.

Metz, T. (2021). *A Relational Moral Theory*. New York: Oxford University Press. https://doi.org/10.1093/oso/9780198748960.001.0001

Mkhize, N. (2008). Ubuntu and Harmony: An African Approach to Morality And Ethics. In R. Nicolson (Ed.), *Persons and Community: African Ethics in a Global Culture* (pp. 35–44). Pietermaritzburg: University of Kwazulu-Natal Press.

Molefe, M. (2019a). *An African Philosophy of Personhood, Morality, and Politics*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-15561-2

Molefe, M. (2019b). African Metaphysics and Religious Ethics. *Filosofia Theoretica: Journal of African Philosophy, Culture and Religions*, *7*(3), 19–37. https://doi.org/10.4314/ft.v7i3.3

Molefe, M. (2022). *Human Dignity in African Philosophy: A Very Short Introduction*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-93217-6

Moor, J. H. (2006). The Nature, Importance, and Difficulty of Machine Ethics. *IEEE Intelligent Systems*, *21*(4), 18–21. https://doi.org/10.1109/MIS.2006.80

Müller, V. C. (2016). Autonomous Killer Robots are Probably Good News. In E. Di Nucci & F. Santoni de Sio (Eds.), *Drones and Responsibility* (pp. 67–81). Farnham, Surrey, UK England ; Burlington, VT : Ashgate, 2016. | Series: Emerging technologies, ethics and international affairs: Routledge. https://doi.org/10.4324/9781315578187-4

Müller, V. C. (2021). Ethics of Artificial Intelligence and Robotics. *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/ethics-ai/#toc. Accessed 14 April 2022

Müller, V. C., & Simpson, T. W. (2014). *Killer Robots: Regulate, Don't Ban*.

Noggle, R. (2022). The Ethics of Manipulation. *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/ethics-manipulation/. Accessed 29 April 2024

Nussbaum, M. (2006). The Moral Status of Animals. *Chronicle of Higher Education*, *52*(22).

Nussbaum, M. (2008). Human Dignity And Political Entitlements. In *Human Dignity and Bioethics: Essays Commissioned by the President's Council on Bioethics* (pp. 351–380).

Nussbaum, M. (2011). *Creating Capabilities: The Human Development Approach*. Cambridge: Belknap Press: Harvard University. https://doi.org/10.4159/harvard.9780674061200. Accessed 5 June 2023

Okeja, Uchenna. (2019). War by Agreement: A Reflection on the Nature of Just War. *Journal of Military Ethics*, *18*(3), 189–203. https://doi.org/10.1080/15027570.2019.1690115

Omotoyinbo, F. R. (2023). Smart Soldiers: Towards A More Ethical Warfare. *AI & SOCIETY*, *38*(4), 1485–1491. https://doi.org/10.1007/s00146-022-01385-3

Orend, B. (2013). *The Morality of War* (Second.). Toronto: Broadview Press.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems*. *Probabilistic Reasoning in Intelligent Systems*. Elsevier. https://doi.org/10.1016/C2009-0-27609-4

Phaswana-Mafuya, N., & Peltzer, K. (2017). Prevalence of Loneliness and Associated Factors among Older Adults in South Africa. *Global Journal of Health Science*, *9*(12), 1. https://doi.org/10.5539/gjhs.v9n12p1

Ramose, M. (1999). *African Philosophy Through Ubuntu*. Harare: Mond Publishers.

Riley, S., & Bos, G. (n.d.). Human Dignity. In *Internet Encyclopedia of Philosophy*.

Roff, H. M. (2013). Killing in War: Responsibility, Liability, and Lethal Autonomous Robots. In F. Allhof, N. G. Evans, & A. Henschke (Eds.), *Routledge Handbook of Ethics and War: Just War Theory in the 21st Century* (First., pp. 352–364). New York: Routledge.

Rosert, E., & Sauer, F. (2019). Prohibiting Autonomous Weapons: Put Human Dignity First. *Global Policy*, *10*(3), 370–375. https://doi.org/10.1111/1758-5899.12691

Ruttkamp-Bloem, E. (2020). The Quest for Actionable AI Ethics. In *Communications in Computer and Information Science* (Vol. 1342, pp. 34–50). https://doi.org/10.1007/978-3-030-66151-9_3

Schönecker, D., & Schmidt, E. E. (2018). Kant's Ground-Thesis. On Dignity and Value in The Groundwork. *Journal of Value Inquiry*, *52*(1), 81–95. https://doi.org/10.1007/s10790-017-9603-z

Schroeder, D. (2008). Dignity: Two Riddles and Four Concepts. *Cambridge Quarterly of Healthcare Ethics*, *17*(2), 230–238. https://doi.org/10.1017/S0963180108080262

Schroeder, D. (2010). Dignity: One, Two, Three, Four, Five, Still Counting. *Cambridge Quarterly of Healthcare Ethics*, *19*(1), 118–125. https://doi.org/10.1017/S0963180109990302

Schroeder, D. (2012). Human Rights and Human Dignity: An Appeal to Separate the Conjoined Twins. *Ethical Theory and Moral Practice*, *15*(3), 323–335. https://doi.org/10.1007/s10677-011-9326-3

Sensen, O. (2009). *Kant on Human Dignity*. Berlin: Walter de Gruyter. https://doi.org/10.1515/9783110267167

Sharkey, A. (2014). Robots And Human Dignity: A Consideration of the Effects of Robot Care on the Dignity of Older People. *Ethics and Information Technology*, *16*(1), 63–75. https://doi.org/10.1007/s10676-014-9338-5

Sharkey, A. (2019). Autonomous Weapons Systems, Killer Robots And Human Dignity. *Ethics and Information Technology*, *21*(2), 75–87. https://doi.org/10.1007/s10676-018-9494-0

Sharkey, A., & Sharkey, N. (2011). Children, the Elderly, and Interactive Robots. *IEEE Robotics & Automation Magazine*, *18*(1), 32–38. https://doi.org/10.1109/MRA.2010.940151

Sharkey, A., & Sharkey, N. (2012). Granny and the Robots: Ethical Issues in Robot Care For the Elderly. *Ethics and Information Technology*, *14*(1), 27–40. https://doi.org/10.1007/s10676-010-9234-6

Sharkey, A., & Wood, N. (2014). The Paro Seal Robot: Demeaning Or Enabling? In *AISB 2014 - 50th Annual Convention of the AISB*.

Sharkey, N. (2014). Killing Made Easy: From Joysticks to Politics. In P. Lin, K. Abney, & G. Bekey (Eds.), *Robot Ethics: The Ethical and Social Implications of Robotics* (pp. 111–128). Massachusets: MIT Press.

Sharkey, N. (2018, November 17). Mama Mia It's Sophia: A Show Robot Or Dangerous Platform To Mislead? *Forbes*. https://www.forbes.com/sites/noelsharkey/2018/11/17/mama-mia-its-sophia-a-show-robot-or-dangerous-platform-to-mislead/?sh=20bb39697ac9. Accessed 10 July 2023

Sharkey, N., & Sharkey, A. (2012). The Eldercare Factory. *Gerontology*. https://doi.org/10.1159/000329483

Shibata, T. (2012). Therapeutic Seal Robot as Biofeedback Medical Device: Qualitative and Quantitative Evaluations of Robot Therapy in Dementia Care. *Proceedings of the IEEE*, *100*(8), 2527–2538. https://doi.org/10.1109/JPROC.2012.2200559

Siau, K., & Wang, W. (2020). Artificial Intelligence (AI) Ethics. *Journal of Database Management*, *31*(2), 74–87. https://doi.org/10.4018/jdm.2020040105

Sparrow, R. (2007). Killer Robots. *Journal of Applied Philosophy*, *24*(1), 62–77. https://doi.org/10.1111/j.1468-5930.2007.00346.x

Sparrow, R. (2016). Robots in Aged Care: A Dystopian Future? *AI & SOCIETY*, *31*(4), 445–454. https://doi.org/10.1007/s00146-015-0625-4

Sparrow, R. (2020). Robotics has a Race Problem. *Science, Technology, & Human Values*, *45*(3), 538–560. https://doi.org/10.1177/0162243919862862

Sparrow, R., & Sparrow, L. (2006). In the Hands of Machines? The Future of Aged Care. *Minds and Machines*, *16*(2), 141–161. https://doi.org/10.1007/s11023-006-9030-6

Stahl, B. C., & Coeckelbergh, M. (2016). Ethics of Healthcare Robotics: Towards Responsible Research and Innovation. *Robotics and Autonomous Systems*, *86*, 152–161. https://doi.org/10.1016/j.robot.2016.08.018

Sternberg, R. J. (2023). Human Intelligence. In *Encyclopedia Britannica.* . https://www.britannica.com/science/human-intelligence-psychology/Cognitive-contextual-theories. Accessed 29 April 2024

Sullins, J. P. (2006). When Is a Robot a Moral Agent? *International Review of Information Ethics*, *6*(12), 23–30. https://doi.org/10.1017/cbo9780511978036.013

Sullins, J. P. (2012). Robots, Love, and Sex: The Ethics of Building a Love Machine. *IEEE Transactions on Affective Computing*, *3*(4). https://doi.org/10.1109/T-AFFC.2012.31

Taylor, I. (2021). Who Is Responsible for Killer Robots? Autonomous Weapons, Group Agency, and the Military-Industrial Complex. *Journal of Applied Philosophy*, *38*(2), 320–334. https://doi.org/10.1111/japp.12469

Tshivhase, M. (2021). Understanding and (Re)configuring Personhood Contra Systemic Dehumanisation. In M. Molefe & C. Allsobrook (Eds.), *Towards an African Political Philosophy of Needs* (pp. 109–129). Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-030-64496-3

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, *59*(236), 433–460. https://doi.org/10.7551/mitpress/6928.003.0012

Tutu, D. (1999). *No Future Without Forgiveness*. New York: Random House .

Ugwuanyi, L. (2020). TOWARDS AN AFRICAN THEORY OF JUST WAR. *Revista de Estudios Africanos*, (1), 51–65. https://doi.org/10.15366/rea2020.1.003

Ulgen, O. (2018). Human Dignity in an Age of Autonomous Weapons: Are We in Danger of Losing an "Elementary Consideration of Humanity"? *SSRN Electronic Journal*, (January). https://doi.org/10.2139/ssrn.2912002

United Nations. (1948). Universal Declaration of Human Rights. https://www.un.org/en/universal-declaration-human-rights/. Accessed 30 April 2020

Vallor, S. (2011). Carebots and caregivers: Sustaining the ethical ideal of care in the twenty-first century. *Philosophy and Technology*, (24), 251–268. https://doi.org/10.1007/s13347-011-0015-x

Vandemeulebroucke, T., Dierckx de Casterlé, B., & Gastmans, C. (2018). The Use of Care Robots in Aged Care: A Systematic Review of Argument-Based Ethics Literature. *Archives of Gerontology and Geriatrics*. https://doi.org/10.1016/j.archger.2017.08.014

Von Neumann, J., & Burks, A. W. (1966). *Theory of Self-Reproducing Automata*. (A. W. Burks, Ed.). Urbana and London : University of Illinois Press.

Von Neumann, J., & Taub, A. H. (1964). John von Neumann Collected Works: theory of automata and numerical analysis. *Journal of the American Statistical Association*, *59*(307), 981. https://doi.org/10.2307/2283131

Wallach, W., & Allen, C. (2009). *Moral Machines: Teaching Robots Right from Wrong*. *Oxford University Press*. New York: Oxford University Press.

Walzer, M. (2015). *Just and Unjust Wars: A Moral Argument with Historical Illustrations* (Fifth.). New York: Basic Books.

Werner, M. (2014). Individual and Collective Dignity. In M. Duwell, J. Braarvig, R. Brownsword, & D. Mieth (Eds.), *The Cambridge Handbook of Human Dignity: Interdisciplinary perspectives* (pp. 343–352). Cambridge: Cambridge University Press.

Wiredu, K. (1996). *Cultural Universals and Particulars: An African Perspective*. Bloomington, IN: Indiana University Press.

Wullenkord, R., & Eyssel, F. (2020). Societal and Ethical Issues in HRI. *Current Robotics Reports*, *1*(3). https://doi.org/10.1007/s43154-020-00010-9

Young, G. (2021). On the Indignity of Killer Robots. *Ethics and Information Technology*, *23*(3). https://doi.org/10.1007/s10676-021-09590-2