

**ARTIFICIAL INTELLIGENCE AND HEALTHCARE IN SOUTH
AFRICA: ETHICAL AND LEGAL CHALLENGES**

by

ANISHA AMARAT JOGI

submitted in fulfilment of the requirements for the degree of

DOCTOR OF LAWS

In the College of Law

at the

University of South Africa

SUPERVISORS

PROF M LABUSCHAIGNE

PROF S MAHOMED (CO-SUPERVISOR)

MAY 2021

DECLARATION

I, Anisha Amarat Jogi, declare that this dissertation “Artificial intelligence and healthcare in South Africa: ethical and legal challenges” is my own work and all sources used in this dissertation have been referenced.

Signed at Pretoria on the 6th day of May 2021.

A handwritten signature in black ink, appearing to read 'Anisha Amarat Jogi', written over a light blue grid background.

SIGNATURE:

ACKNOWLEDGEMENTS

My sincere appreciation goes to my supervisor Professor Labuschaigne and co-supervisor Professor Mahomed for your support and guidance throughout my journey of undertaking this dissertation. I am greatly beholden to you for having patiently and sensibly directing my findings and to your continuous dedication in overseeing my work. You are highly regarded academics and I deeply value you as mentors.

I am thankful for the immense support of my husband and for his continuous love and faith in me. I am fortunate to have a partner and friend who has given me the strength and motivation for this endeavour.

Lastly, my gratitude goes to my work colleagues, who provided me with support at work during the challenging time of completing my research.

ABSTRACT

We find ourselves in an era of intelligent machines as technology pushes the boundaries towards the fourth industrial revolution. AI is already here in disruptive narrow forms, cutting across the globe and leaving almost no sector untouched. With numerous industries such as the healthcare sector already adopting AI technology, we continue to see increased benefits for our society as a whole. AI is developing at a rapid pace and is already capable of undertaking human tasks, and even more effectively and at a lower cost, creating novel challenges for recognised regulatory frameworks and posing serious ethical questions. AI systems are used by professional healthcare workers to diagnose problems or to take decisions on the method of treatment for patients. As such, these AI systems can be regarded as either products or services through which healthcare is delivered to patients as the consumers. The thesis therefore considers how AI ought to be regulated in South Africa, elaborating on the ethico-legal concepts that can be applicable to the branches of the law with emphasis on the healthcare sector. Two main groups of parties could be held accountable under civil law in the event of injury sustained in the application of AI systems, comprising of the producers and the end-users of the systems. In South Africa, where damage ensues due to defective AI systems in healthcare, the injured party would have a claim against the producers or professional end-users of the systems, either in delict based on negligence where the plaintiff carries the duty of establishing that the duty of care of the reasonable expert or professional has been breached or in terms of product liability as codified in the Consumer Protection Act (CPA). In law of delict in South Africa, it makes no difference whether AI systems are classified as a product or a service, since both will be actionable in terms of the *Aquilian* action or non-patrimonial loss when damage is unlawfully caused. Similarly, this distinction does not apply under the CPA. Thus, in delict the injured party is allowed greater flexibility in terms of pursuing an action either under the negligence principles or under the strict product liability principles under the CPA as opposed to the other jurisdictions. Notably quality assurance software standards such as the ISO series can be applied in establishing negligence in AI liability and such standards can motivate for the software and computer programming industry to be conferred with a professional status in South

Africa to provide a benchmark that registered software engineers and technicians must meet as globally recognised professional standards. The thesis also recognises the values that need to be protected in South Africa when considering the ethical implications emanating from the use of AI in healthcare and considers a working ethical framework based on international ethical guidelines that have already been developed for AI.

KEY TERMS: accountability, application, AI systems, AI technologies, artificial intelligence, algorithms, bias, data, design, development, deployment, doctors, ethics, ethico-legal, governance, healthcare, healthcare professionals, impact, legal framework, liability, machine learning, privacy, regulation, technology, transparency.

TABLE OF CONTENTS

DECLARATION.....	i
ACKNOWLEDGEMENTS.....	ii
ABSTRACT	iii
1. CHAPTER ONE: INTRODUCTION, SCOPE AND METHODOLOGY OF STUDY	1
1.1. Defining artificial intelligence	4
1.2. Artificial intelligence from a historical perspective	7
1.3. Current and future impact of AI	10
1.4. AI and the law	13
1.5. Background, literature review and problem statement	16
1.6. Research problem.....	24
1.7. Rationale	24
1.8. Aim of the study	24
1.9. Objectives	25
1.10. Methodology	26
1.11. Limitations.....	27
1.12. Expected contributions / outcomes	27
1.13. Chapter summary	27
1.14. Ethical clearance.....	28
2. CHAPTER TWO: BENEFITS AND CHALLENGES OF AI	29
2.1. Current and potential benefits	29

2.1.1.	Big Data and Machine Learning.....	33
2.1.2.	Outcome prediction.....	36
2.1.3.	Operational efficiency	37
2.1.4.	Precision diagnostics and error mitigation	40
2.1.5.	Application of AI for social good.....	41
2.1.6.	The impact on business	48
2.1.7.	The role of AI health services in rural communities	50
2.2.	Challenges	54
2.2.1.	Data privacy and security	54
2.2.2.	Data inequity.....	62
2.2.3.	Bias.....	65
2.2.4.	Social isolation and trust in AI.....	69
2.2.5.	Effects on healthcare providers	70
2.2.6.	Reliability and safety.....	72
2.2.7.	Transparency and accountability	74
2.2.8.	Challenges for governance	78
2.3.	Conclusion	79
3.	CHAPTER THREE: REGULATION AND ITS IMPACT ON AI	83
3.1.	A definition for regulation.....	83
3.2.	The reasons to regulate AI	86
3.2.1.	To prevent public harm and advance social goals	86

3.2.2.	To prevent market failures	88
3.2.3.	Protection of fundamental human rights	91
3.2.4.	Industry advancement	93
3.2.5.	Legal necessity to regulate	95
3.2.6.	Regulation of AI is already underway	101
3.2.7.	Conclusion	105
3.3.	Risks and difficulties of regulating AI	105
3.3.1.	The definition issue	105
3.3.2.	Understanding AI and its impacts	108
3.3.3.	AI and the existing regulations	110
3.3.4.	Disruptive sector landscape	113
3.3.5.	Technological neutrality	113
3.4.	Impact of regulation on AI	116
4.	CHAPTER FOUR: TOWARDS A GOOD REGULATION OF AI IN HEALTHCARE	123
4.1.	Criteria for determining if a good regulation is conceivable	123
	Necessity	124
	Lawfulness	125
	Legal certainty	126
	Transparency and accountability	126
	Effectiveness or flexibility	127

Targeting.....	127
4.2. A broad regulation.....	129
4.3. Civil liability in delict arising from AI systems	131
4.3.1. Conventional software versus AI software.....	133
4.3.2. Sources of system error.....	134
4.3.3. The identification of possible liable parties if AI systems cause harm	135
4.3.4. Negligence and product liability	141
4.3.5. The test for negligence in South Africa	143
4.3.6. Product liability in South African law	156
4.3.7. Strict product liability under the CPA	167
4.4. A comparative analysis on delictual liability arising from AI systems	172
4.4.1. United Kingdom.....	173
4.4.2. United States	191
4.4.3. Conclusion on AI liability based on comparative jurisdictions	211
4.5. Ethics guidelines for achieving trustworthy AI in healthcare.....	215
4.5.1. International context	218
4.5.2. Ethical guidelines for AI in South Africa’s healthcare.....	228
4.5.3. Conclusion.....	275
5. CHAPTER FIVE: SUMMARY, CONCLUSIONS AND RECOMMENDATIONS	
280	
5.1. Summary.....	280

5.2. Conclusions and recommendations	284
5.2.1. Assessment methods	285
5.2.2. Development approaches.....	288
5.2.3. Incentives	290
5.2.4. Support Initiatives	290
5.2.5. Considering the Africanisation of AI.....	291
ANNEXURE 1 – ETHICAL CLEARANCE	295
BIBLIOGRAPHY.....	300
Abbreviations	300
Journal Articles	300
Books.....	311
Internet Sources.....	318
Cases.....	339
South Africa	339
Germany	340
New Zealand.....	340
United States	340
United Kingdom	341

1. CHAPTER ONE: INTRODUCTION, SCOPE AND METHODOLOGY OF STUDY

As with the age of the internet that has significantly altered the way people have functioned in the past centuries, With the fourth industrial revolution rapidly progressing, artificial intelligence (AI), the technology of programming human cognitive abilities into machines, has the power to substantially change society in the next century The technology is capable of performing tasks that could only be undertaken by a human with certain knowledge and skills.¹ It has the ability to revolutionise medical science, and offers many prospects for the healthcare sector – from limiting onerous functions in administration, to enhancing the precision of diagnosis and improving patient care management, to accelerating drug production and distribution.²

While public healthcare spending in South Africa has remained the government's priority, this has been insufficient to meet service delivery costs due to the country's existing social-, economic-, and political crises, inefficiencies caused by corruption, and wasteful expenditure, which has constrained national healthcare coverage.³ The novel COVID-19 epidemic has also uncovered the deficiencies and inefficiencies in South Africa's healthcare service, such as vast inequalities in the accessibility and the standard of care between healthcare facilities across provinces.⁴

However, the pandemic has also intensified the awareness for prioritising structural improvements and reforms that can shape the healthcare market, particularly with regards to new technology. After World War II, in realising the importance to establish the United Nations (UN) in fear of another war, Winston Churchill famously

1 Autonomous systems can undertake advanced financial transactions, assist in legal enforcement to trace suspected terrorists using facial recognition software and perform document review and analyse data in less of the time that a human is capable of.

2 Turpitka D "The challenges of applying AI to healthcare and how to address them" <https://www.forbes.com/sites/forbestechcouncil/2020/03/31/the-challenges-of-applying-ai-to-healthcare-and-how-to-address-them/?sh=17d9d19e5557> (Date of use: 25 July 2020).

3 Rural health advocacy project Protecting rural healthcare in times of economic crises there is another way – An overview"
https://www.groundup.org.za/media/uploads/documents/Austerity%20report_FINAL.pdf (Date of use: 18 May 2020).

4 Benjamin P et al "What a people-centred response to Covid-19 would look like" <https://www.dailymaverick.co.za/article/2020-07-14-what-a-people-centred-response-to-covid-19-would-look-like/> (Date of use: 23 July 2020).

stated: “Never let a good crisis go to waste.”⁵ What was meant by this is that, a crises rouses an opportunity to do things that could not be done otherwise.

In recent years, public tertiary hospitals have experienced significant budget restrictions which has curtailed the buying-power of these institutions.⁶ Therefore, understanding how to respond to austerity measures, in a manner that protects health benefits, and which makes best use of available resources, is critical. In this regard, AI applications offer the ability to address the “iron triangle” quandary experienced in healthcare, where the interlinking elements “access, affordability, and effectiveness” are inevitably traded-off, and where attempting to enhance one element usually leads to harms the other.⁷

AI simultaneously provides a solution to rising labour shortages in healthcare and expanding healthcare reach to patients with limited resources or access to doctors, transferring time-consuming human tasks to machines, or providing doctors with the tools to carry out their tasks effectively, by enabling them to diagnose health issues more accurately.⁸ Whilst increased potential for innovative technology, is predominantly applied in the private healthcare industry in South Africa,⁹ it should not be discounted for the public sector, as these institutions stand to benefit the most from the technology due to their innate challenges.

However, despite the improvements and benefits that AI technology offers, it has also incited distress in many quarters, prompting a demand for government

5 Mutter J “Opportunity from crisis, who really benefits from post-disaster rebuilding efforts” <https://www.foreignaffairs.com/articles/2016-04-18/opportunity-crisis> (Date of use: 25 July 2020).

6 Sebulela G “Covid-19: Opportunity for Investment in South Africa’s healthcare sector” <https://www.iol.co.za/business-report/opinion/covid-19-opportunity-for-investment-in-south-africas-healthcare-sector-46372334> (Date of use: 20 July 2020).

7 Forbes “AI and healthcare: A giant opportunity” <https://www.forbes.com/sites/insights-intelai/2019/02/11/ai-and-healthcare-a-giant-opportunity/#56d362cd4c68> (Date of use: 23 March 2019). Terry NP “Appification, AI, and healthcare’s new iron triangle” 2018 *Journal of health care law & policy* 119-120.

8 PWC “No longer science fiction, AI and robotics are transforming healthcare” <https://www.pwc.com/gx/en/industries/healthcare/publications/ai-robotics-new-health/transforming-healthcare.html> (Date of use: 31 March 2019).

9 Sebulela <https://www.iol.co.za/business-report/opinion/covid-19-opportunity-for-investment-in-south-africas-healthcare-sector-46372334> (Date of use: 20 July 2020).

intervention in AI development and oversight on its implementation.¹⁰ The phenomena of fear is not unusual whenever there are advances in technology, but what stands out in the case of AI is that, frontrunners in the technology industry like Elon Musk and Bill Gates have cautioned that, AI technology needs to be regulated to manage the risks.¹¹ At the 2014 AeroAstro Centennial Symposium, hosted by MIT, Elon Musk during an interview stated:

I think we should be very careful about artificial intelligence. If I had to guess at what our biggest existential threat is, it's probably that. I'm increasingly inclined to think there should be some regulatory oversight, maybe at the national and international level, just to make sure that we don't do something very foolish.¹²

Some of the concerns stem from unemployment due to technological replacement,¹³ and the likelihood for novel technologies to be used by humans, for harm.¹⁴ The medical community is threatened by AI as they fear that this may replace them in their jobs and as a result thereof, they are reluctant to share data.¹⁵

AI will soon be making judgments across various sectors in the finance-, medical-, legal-, and other areas, which will impact on those individuals that are the subjects of decisions made by AI. It is, therefore, inevitable that an intelligent machine will appear before the courts and be legally recognised as “persons” for the sake of

10 Scherer UM “Regulating artificial intelligence systems: risks, challenges, competencies, and strategies” 2016 *Harvard journal of law and technology* 355.

11 Straub J “Does regulating artificial intelligence save humanity or just stifle innovation?” <https://techxplore.com/news/2017-10-artificial-intelligence-humanity-stifle.html> (Date of use: 31 March 2019).

12 MIT “Centennial symposium: One-one-one with Elon Musk” <https://aeroastro.mit.edu/videos/centennial-symposium-one-one-one-elon-musk> (Date of use: 31 March 2019).

13 See Russell SJ and Norvig P *Artificial intelligence: A modern approach* 3rd ed (Prentice Hall New Jersey 2010) 1034.

14 See Russell and Norvig *Artificial intelligence: A modern approach* 1034. The authors refer to how innovative technologies are used by those with power as an advantage over their rivals and stated that, “[a] science is said to be useful if its development tends to accentuate the existing inequalities in the distribution of wealth, or more directly promotes the destruction of human life.”

15 This was echoed during an interview conducted by The Batch (published in a new weekly newsletter on 2020-04-15 from [deeplearning.ai](https://www.deeplearning.ai)) with Dr Eric Topol a leading advocate for AI in medicine and a cardiologist and geneticist at Scripps research institute in Southern California. Newsletter published on 2020-04-15, DeepLearning.AI “A visionary doctor prescribes AI” <https://blog.deeplearning.ai/blog/the-batch-ai-for-medicine-special-eric-topols-planetary-health-system-discovering-drugs-diagnosing-heart-disease-predicting-infections-alexa-for-doctors> (Date of use: 31 March 2019).

upholding fairness in society, to attain equality before the law.¹⁶ The complexities of AI must therefore be understood against the broader ethical and legal considerations, in order to enable its incorporation into a clear regulatory framework, for a symphony of humans and intelligent machines to effectively co-exist, and to mitigate risks and diffuse public concerns.

The thesis discusses how AI should be regulated in South Africa, elaborating on the ethical and legal concepts, with emphasis on the healthcare sector. This thesis acknowledges that AI will impact on various areas of law and a possible regulation of the technology for the healthcare sector must have due regard to the aforesaid.

1.1. *Defining artificial intelligence*

A singular universal definition for AI does not exist. The Oxford Dictionary defines AI as the:

Theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages.¹⁷

This definition is also used by some scholars. In terms of the modern definition, AI is “the study and design of intelligent agents”, that can adapt to various circumstances previously unknown and learning through familiarity, reaching a goal and objective not conceivable to conventional computers.¹⁸ The definition of AI is important in the context of regulation. Scherer argues that, in consideration of an effective regulatory approach there should be a definition of that which is being regulated.¹⁹

16 Willick MS “Artificial Intelligence: Some legal approaches and implications” 1983 *The AI Magazine* 5.

17 Oxford Dictionary https://www.lexico.com/en/definition/artificial_intelligence (Date of use: 23 March 2019).

18 Russell and Norvig *Artificial intelligence: A modern approach* 31-52; Poole DL, Mackworth AK and Goebel R *Computational intelligence: A logical approach* (Oxford University Press New York 1998) 1.

19 Scherer 2016 *Harvard journal of law and technology* 359.

However, a vital problem in AI is that nobody really knows what “intelligence” is and how to then define the object of regulation.²⁰ The problem becomes more significant when artificial systems, which are opaque to humans, are considered.²¹ In order to formulate a set of guidelines, there must be clarity as to what is being regulated. In other words, a regulatory system must define AI.

An incorrect definition for AI can have undesirable outcomes insofar as its regulation is concerned, therefore, it is important that there is consensus regarding how it is defined. If the definition is minimalist, this may create loopholes allowing others to escape regulation that can be exploited and would create a competitive advantage competition against those that are regulated, as well as subjecting others to unnecessary harm, due to technology that is not subject to any scrutiny. If overstated, basic software and machines that may not require regulation, this will be subject to the same scrutiny as AI systems and will suffer the same restrictions. This can be detrimental by slowing the development of inoffensive software due to an all-encompassing definition, which would simply overburden the affected sectors.²²

Intelligence involves a perplexing mixture of concepts, rendering it equally difficult to define.²³ The challenges in defining AI does not have to do with the notion of artificiality, but instead with the theoretical opacity of intelligence. Human beings are generally acknowledged to be the only subjects as holding intelligence. Thus, intelligence is defined to be associated to characteristics that are akin to humans.²⁴ John McCarthy states that,

20 AI involves the understanding and mimicking of the human thought processes performing the same task. This would make the defining of AI challenging until the state of human intelligence is understood sufficiently as capable of being copied.

21 Hutter M and Legg S “Universal intelligence: A definition of machine intelligence” 2007 *Minds and Machines* 391.

22 For example, a software chess program that analysis each possible move and then selects the best one based on a scoring criterion would fall in such a category. An over-inclusive definition of AI may lead to unintelligent software codes being adopted, inhibiting the progress of simple software as it fails to demonstrate the difference between AI that learns from data and AI that can assess possible outcomes.

23 Hutter and Legg 2007 *Minds and Machines* 392.

24 Scherer 2016 *Harvard journal of law and technology* 359.

[there is no] solid definition of intelligence that doesn't depend on relating it to human intelligence as the problem is that we cannot yet characterize in general what kinds of computational procedures we want to call intelligent.²⁵

Adopting a legal definition for AI plays a significant role towards its regulation, mainly since laws and policies cannot merely function in a regulatory vacuum.²⁶ Various definitions of AI are being used by the tech and computer industry. As a result of the challenges in the definition of AI, many have recognised that the issues concerning a machine's intelligence and its objective is in the long run not a question of innovation, but rather that of decision.²⁷ What remains necessary is a solution that is well defined and universally accepted, and robust enough to promote good cooperation amongst different sectors.

How does one then go about developing the notion of intelligence that would be relevant to various types of smart technologies? A recommended definition should encapsulate the core of human intellect which prodigiously varies in terms of their reasoning capabilities, efficiency, feelings, and the circumstances in which they operate. It should also be grounded on rules which are likely to change with time. Furthermore, a definition for intelligence should be appropriately articulated, impartial, and realistically achievable, as an efficient test for intelligence.²⁸

In this thesis, I address the definitional issue by adopting the scientific definition of machine intelligence as "the study and design of intelligent agents",²⁹ that can adapt to various circumstances previously unknown and learning through familiarity, reaching a goal and objective not conceivable to conventional computers.³⁰

25 McCarthy J "What is artificial intelligence?" <http://www-formal.stanford.edu/jmc/whatisai.pdf> (Date of use: 16 April 2020).

26 Lea G "Why we need a legal definition of artificial intelligence" <https://theconversation.com/why-we-need-a-legal-definition-of-artificial-intelligence-46796> (Date of use: 6 April 2020).

27 Willick 1983 *The AI Magazine* 6.

28 Hutter and Legg 2007 *Minds and Machines* 393.

29 Kok *et al Artificial intelligence* 1095-1096.

30 Scherer 2016 *Harvard journal of law and technology* 361. According to Scherer, gobally accepted views to defining AI adopt the notion of "machines that work to achieve goals." See also Hutter and Legg 2007 *Minds and Machines* 391, 405-423, where the authors state that, intelligence is a measure of the subject's ability to accomplish goals in different environments.

I submit that, this is the most appropriate definition to be used because it seems to align with the contemporary scientific knowledge of this subject,³¹ and is commonly used in the literature.³² Such a definition is also not under-inclusive as it is not restricted to human intelligence but recognises different types of intelligence, as well as the flexible and developing landscape of AI. Furthermore, it is also not overly inclusive nor restrictive, as it differentiates advanced and complex technology from traditional systems that do not require regulation.

Basic technology, such implementing or operating a neutral network on your laptop, should not be regulated as it is not necessary to regulate the advancement of such harmless systems³³ Furthermore, it would be impractical to regulate, as opposed to other applications of AI, for instance, autonomous driving, which necessitates regulation.³⁴

The proposed definition can therefore be pragmatically applied when considering the regulation of AI. I submit that there should be an allowance for such a flexible definition in the laws aimed to establish the regulatory framework and which consequently allows regulators to determine what deserves their scrutiny.

1.2. *Artificial intelligence from a historical perspective*

The journey of understanding whether machines are capable of thinking began in the 1940s with philosophers such as Vannevar Bush and Alan Turing.³⁵ Modern AI was inspired by these conventional philosophers who endeavoured to explain the

31 See McCarthy <http://www-formal.stanford.edu/jmc/whatisai.pdf> (Date of use: 16 April 2020), where McCarthy defines intelligence to be “the computational part of the ability to achieve goals in the world” and AI to be “the science and engineering of making intelligent machines, especially intelligent computer programs.” Russell and Norvig refer to a “rational agent” towards a working definition for AI, describing the agent to be “one that acts so as to achieve the best outcome or, when there is uncertainty, the best expected outcome.”

32 Russell SJ and Norvig P Instructor’s manual: Exercise solutions for artificial Intelligence a modern approach 2nd ed (Pearson Prentice Hall New Jersey 2003) 31-52. Kok *et al Artificial intelligence* 1095-1096. Hutter and Legg 2007 *Minds and Machines* 391, 405-423.

33 For example, computer chess programs and video games are designed to achieve an optimal result based on predefined sets of rules and goals.

34 Huffpost “Regulating artificial intelligence: A look at the pros and cons” https://www.huffpost.com/entry/regulating-artificial-intelligence-a-look-at-the-pros_b_59eadb0ae4b034105edd4ec4 (Date of use: 18 April 2020).

35 McGuire B *et al* “The history of artificial intelligence” <https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf> (Date of use: 29 November 2018).

development of human thinking as the systematic manipulation of codes. Their ideas culminated in what is referred to as the “programmable digital computer” during the 1940s, a device based on mathematical reasoning. The idea behind such a system further inspired some scientists to begin considering the prospect of creating an electronic brain.³⁶

The expression “artificial intelligence” was introduced by John McCarthy during the 1950’s. McCarthy based his study on an attempt to discover how to attribute human qualities to machines such as language, cultivate concepts, resolve problems, and develop themselves.³⁷

During 1957 to 1974, AI underwent an enormous evolution. Computers became capable of storing more data and were more efficient, cost effective, and accessible.³⁸ With this, machine learning processes also developed and brought about a new enthusiasm amongst researchers in the field. Examples of this include Newell and Simon’s “General Problem Solver” (GPS), and “ELIZA” developed by Joseph Weizenbaum’s, demonstrating real advancements in emulating human problem-solving rules and embodied the “thinking humanly” approach.³⁹ These early successes, together with the activism of other leading scholars, convinced government agencies to financially support AI research.⁴⁰ However, while the basic foundation was in place, the end goals of ordinary linguistic processing, conceptual understanding, and self-identification was still to be attained.⁴¹

Studies greatly overestimated AI progress since the early days, partially caused by the public hype. Together with the staggered advancements in research, this resulted in a huge disappointment in the area. The AI industry was confronted with

36 McGuire *et al* <https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf> (Date of use: 29 November 2018).

37 Russell and Norvig *Instructor’s manual* 17.

38 Anyoha R “The history of artificial intelligence” <http://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/> (Date of use: 8 April 2020).

39 Russell and Norvig *Instructor’s manual* 18.

40 Anyoha <http://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/> (Date of use: 8 April 2020).

41 Anyoha <http://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/> (Date of use: 8 April 2020).

a point in time known as the “AI winter”.⁴² This created a negative connotation among the scientific community that exists even today.⁴³ This hindrance in the 1970s caused governments to significantly reduce their funding to researchers, ultimately leading to the closing-down of programs during the 1980s. AI researchers had to hide their identity by using different names, to enable them to continue receiving funding. Certain scientists disguised their works through using conventional names “machine learning” (ML), “informatics”, “knowledge-based system”, or “pattern recognition”, to enable them to continue developing their work during the AI winter.⁴⁴

Despite the challenges due to the lack of government funding, AI began to flourish in the 1990s and the beginning of the 21st century. This period of time saw many of the milestones for AI being attained.⁴⁵ It commenced with developments made to existing technologies like video-games⁴⁶ and web-based contextual searches.⁴⁷ These advancements led to the construction of new AI-infused tools⁴⁸ such as virtual assistants.⁴⁹

Investment in AI soared during the first part of the 21st century, when ML was used to resolve various problems in the academia- and technology field, ensuing from the

42 McGuire *et al* <https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf> (Date of use: 29 November 2018).

43 Yudkowsky E “Artificial intelligence as a positive and negative factor in global risk” in Bostrom N and Ćirković MM (ed) *Global catastrophic risks* (Oxford University Press New York 2008) 38-39.

44 McGuire *et al* <https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf> (Date of use: 29 November 2018).

45 During the 1990’s, renowned chess titleholder Gary Kasparov was beaten by IBM’s computerised chess platform. This largely contributed towards an AI intelligent decision-making system. Around the same time, Dragon’s speech recognition software was created and implemented on Windows.

46 The Digital stream “How AI in video games is furthering our knowledge of programming AIs in general” <https://andrewdouglasblog.wordpress.com/2016/09/28/how-ai-in-video-games-is-furthering-our-knowledge-of-programming-ais-in-general/> (Date of use: 8 April 2020).

47 Contextual search optimises internet-based search engine results based on field context provided by the user. Examples of search engines where contextual searches can be undertaken are Google, Bing and Yahoo.

48 IBM “Conversational AI” <https://www.ibm.com/watson/advantage-reports/future-of-artificial-intelligence/ai-conversation.html> (Date of use: 8 April 2020).

49 Apple’s Siri, Microsoft’s Cortana, Google’s Assistant, and Amazon’s Alexa are popular examples of virtual assistants.

creation of sophisticated computer hardware.⁵⁰ Some engineers, such as Elon Musk, projected the impending advent of general artificial intelligence – a machine with intellectual advancements that surpassed the abilities of humans. Algorithms were introduced as parts of the bigger systems. AI had proved to solve many complex problems and their solutions were used in various areas, such as: data mining; industrial robotics; gaming; speech recognition; financial analysis; and healthcare diagnosis.⁵¹

Thus, AI became the apex of today's technological age as it pushed the boundaries towards the fourth industrial revolution and super-intelligence mechanisms inserted into people's everyday lives. "Industry 4.0" is a term coined during 2011 by the Fraunhofer-Gesellschaft organisation and German national state, inspired by the exchange of information, automation and manufacturing technologies, in grouping the "Internet of Things" (IoT), "Cyber-physical Systems" (CPS), and "Internet of Services" (IoS), all working together and with human operators, where data is shared via the internet for machinists and end-users.⁵²

According to Schwab, the fourth industrial revolution "is characterised by a fusion of technologies that is blurring the lines between the physical, digital and biological spheres."⁵³ This suggests that, this progressive amalgamation of technologies and humans gives rise to the ethico-legal challenges and complexities that need to be addressed when AI is contemplated.

1.3. *Current and future impact of AI*

Currently, AI has proven to be able to execute some functions that, previously, only humans were capable of undertaking, leading to progressive (but noticeable) alterations in human interaction. Science fiction regularly depicts AI as robots that

50 McGuire *et al* <https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf> (Date of use: 29 November 2018).

51 Russell SJ and Norvig P *Artificial intelligence a modern approach* 3rd ed (Pearson Prentice Hall New Jersey 2010) 28.

52 Chung M & Kim J "The internet information and technology research directions based on the fourth industrial revolution" 2016 *KSII transactions on internet and information systems* 1312.

53 Schwab K "The fourth industrial revolution: What it means, how to respond" <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/> (Date of use: 30 March 2019).

possess human-resembling intelligence similar to that which we are noticing in reality. Today, AI operates in programs such as IBM's Watson, a question-answering computer system, also used in autonomous weapons. The AI of today can be labelled as "narrow" or "weak", as it can only perform limited tasks, such as operating contextual web-based searches or as a self-driving vehicle.

However, researchers are looking to create a "general" or "strong" AI. It is therefore anticipated that, while narrow AI may currently be able to match or beat humans at a specific task allocated to it (such as videogames or problem-solving equations) in its "strong" form, AI would be capable of exceeding humans at almost every intellectual task.⁵⁴

Healthcare is expected to be the next frontier for technological disruption and AI is already starting to transform this field. The potential for AI in the arena of medical research and healthcare services are infinite. From clinical trials and genomic data, to healthcare records and data from wearables, the healthcare ecosystem is now inundated with data, the essential raw material for AI practitioners.

More recently, the COVID-19 pandemic has globally intensified the burden on healthcare systems but has also highlighted the capability of AI as a key contributor in the solution to fight the pandemic.⁵⁵ By analysing a myriad of data, news, social media, and government reports, AI has played a role in everything – from identifying, tracking and forecasting the spread of the virus, to testing new treatments and producing a vaccine. Due to the surge of patients experienced by business and administrative divisions, block-chain platforms offered by companies (such as Ant Financial) expedites medical claims processing and reduces person-to-person

54 Future of Life Institute "Benefits and risks of artificial intelligence"
<https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/?cn-reloaded=1>
(Date of use: 11 April 2020).

55 A South African medical technology company has launched a new platform called "Radify" which helps doctors diagnose people with COVID-pneumonia. Usually, after radiologists take X-rays of patients, the results can take several days. The Radify platform eliminates waiting for results and can also identify which patient is a high risk for contracting COVID-19 by simply analysing an X-ray of their chest.

interaction between patients and healthcare staff.⁵⁶ Therefore, COVID-19 is likely to accelerate the uptake of AI.

In the medical field, another significant change introduced by AI can be found in robots used to perform surgery, which may be considered as “too fragile” for a surgeon’s hands to undertake, or as an aid in various types of regular surgeries.⁵⁷ In 2013 the Urology Hospital in South Africa acquired the first robotic-assisted surgical system. Since then, the hospital has performed more than a 1000 robotic procedures. The robotic system has an augmented 3D-vision system, and miniature wristed instruments that are much more flexible than the human wrist.

The robotic system, therefore, has features that allows a surgeon to operate with enhanced visualisation, accuracy and control.⁵⁸ In the United Kingdom (UK), the National Health Service (NHS) uses AI technology to assist with early detection of heart disease and cancer, in order to limit unnecessary medical procedures, assist in research (by considering the appropriate patients for clinical tests), and as support care for patients with challenging needs.⁵⁹ AI is also being used for predictive and precision diagnoses for breast cancer, where routine screening is important in order to detect the earliest symptoms of the disease.⁶⁰

These examples illustrate how human lives and the healthcare sector is already influenced by AI. Despite the many challenges encountered as the technology is implemented into new functions, these systems are becoming even more sophisticated and capable. The expectation is that AI will have a more positive

56 Marr B “Coronavirus: How artificial intelligence, data science and technology is used to fight the pandemic” <https://www.forbes.com/sites/bernardmarr/2020/03/13/coronavirus-how-artificial-intelligence-data-science-and-technology-is-used-to-fight-the-pandemic/#5552252b5f5f> (Date of use: 15 April 2020).

57 Davenport T and Kalakota R “The potential for artificial intelligence in healthcare” 2019 *Future healthcare journal* 95.

58 The Urology Hospital “What is Robotic Surgery” <https://urology.co.za/health-professionals/robotic-surgeons/> (Date of use: 9 July 2019).

59 UK GOV “New code of conduct for artificial intelligence (AI) systems used by the NHS” <https://www.gov.uk/government/news/new-code-of-conduct-for-artificial-intelligence-ai-systems-used-by-the-nhs> (Date of use: 20 May 2019).

60 Shetty S and Tse D “Using AI to improve breast cancer screening” <https://www.blog.google/technology/health/improving-breast-cancer-screening/> (Date of use: 17 May 2020). Google AI researchers have trained an AI model to detect breast cancer in scans with much greater accuracy against that of radiographers.

benefit than a negative impact for society. Therefore, AI must be established and implemented in a manner that is clear, accountable, in the interest of society, and balanced with the goal of driving innovation, as opposed to stifling it.

While the ultimate objective of AI is to benefit humanity, it must be designed so that it is not mutually destructive with the primary goals of humans.⁶¹ AI should not become so proficient at what it was developed to undertake, that it results in crossing over ethical or legal boundaries. Because of these new “intelligent” technologies, the evolutionary impact of AI on our society will have far-reaching legal and ethical consequences, which we need to be examine and prepare for.⁶² This is explained in the following section.

1.4. *AI and the law*

AI will not only impact on society, it will also challenge the legal world in term of its regulation. Labour law, delict (tort) law,⁶³ data protection law, and privacy laws are just a few examples of the fields of law which will be impacted by AI. It is becoming challenging to distinguish sophisticated computer technologies from that of humans, as computers have developed to a stage where they can exceed both physical and mental “human” tasks.

AI incites a plethora of legal questions stemming from legal recognition and rights. For instance, who or what should be led legally accountable, in cases where doctors delegate the diagnoses of medical conditions to AI systems, and the system makes an error? Will the government institutions and the judiciary be equipped to impute legal blame whenever an AI systems causes injury to patients? As technology

61 Marr B “What is the Impact of artificial intelligence (AI) on society?” <https://bernardmarr.com/default.asp?contentID=1828> (Date of use: 11 April 2020).

62 Whilst there are those who believe that human-extent AI is many decades away, majority of the AI researches who attended the Puerto Rico Conference on AI in 2015 estimated that this may likely occur prior to 2060. It is therefore, imperative to commence with research on the challenges of innovative AI, to establish what impact it will have for humanity, available at Future of Life Institute “Benefits and risks of artificial intelligence” <https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/?cn-reloaded=1> (Date of use: 11 April 2020).

63 The Anglo-American jurisdictions refer to the "law of tort" instead of “law of delict”. "Tort" is the Latin reference to *tortus* which means "wrong".

becomes more developed in performing human “intelligent” tasks, the line between operator and instrument begins to blur.⁶⁴

In *AN v MEC for Health, Eastern Cape* 2019 4 All SA 1 (SCA), damages were claimed on account of a baby who had sustained damages during labour. The question before the court was whether negligence of the hospital staff was causally connected to the child’s brain damage. The majority judgment found that causation was not established as the interruption of the blood supply to the foetus resulting in a lack of oxygen to the foetal brain was without warning and occurred at a time when intervention by way of an emergency delivery could not have been performed in time by the hospital staff to avert the damage. Thus, cases where medical negligence must be determined, it is not difficult to establish under the elements for law of delict whether the conduct of human medical staff were the cause of the ensuing damages, and to hold them accountable if they were. The situation becomes more complex when reliance is placed on AI systems when performing medical procedures as the question is whether human medical staff can be held liable for errors caused due to the use of AI systems that have no legal recognition.

It is an increasing reality that, intelligent computers will insert themselves into the legal arena through the courts, which will eventually give rise to some judicial recognition to “computer” personality and rights in a legal dispute. AI is an unsettling technology that has the potential to bring about changes in society. The existing legal framework may not be equipped to deal with changes and the actuality concerning legal gaps, which can create challenges.⁶⁵

Current regulation on delictual (tort) liability may not be adequate to mitigate the perils associated with intelligent and autonomous machines.⁶⁶ The autonomous nature of AI systems gives rise to concerns of foreseeability and control, which may make traditional regulatory delictual frameworks ineffective, posing far-reaching

64 Willick 1983 *The AI Magazine* 5.

65 Tiažkijus V “Gaps in labour law and their influence on flexibility and stability of the labour law system” 2012 *Jurisprudencija* 1551.

66 Scherer 2016 *Harvard journal of law and technology* 356.

risks. Moreover, the issue of regulation becomes complicated by the absence of a single definition for AI, hence its actual meaning.

AI also raises social and ethical questions that relate to data-bias and abuse, unemployment, and the lack of quality control. This is more so for countries such as South Africa, considering the continent's diversity and economic challenges. Social issues emanating from AI overlap with issues arising from automation and assistive technologies.⁶⁷ One may question whether AI, which is capable of replacing the human role, has a place in certain communities, especially where the element of trust and human sentiment may be lost by its introduction.⁶⁸ In *Mankayi v AngloGold Ashanti Ltd* 2011 (3) SA 237 (CC) the issue that was to be decided was whether section 35 (1) 1 of the Compensation for Occupational Injuries and Diseases Act 2 (COIDA) extinguishes the common law right of mineworkers to recover damages for occupational injury or disease from negligent mine owners. Such a case exemplifies the fact that current labour laws are designed to protect the rights and safety of a human workforce. Thus, where certain jobs become obsolete and become dependent on AI systems, the issue is how will these labour laws protect the human workforce against the impact of AI systems.

AI systems are trained and developed on the application of large amounts of data from data subjects. This gives rise to issues of such data being subject to abuse or exploitation. Many individuals are unaware of the collection of their data by companies, and hand over their sensitive data freely to obtain products and services. In *NM v Smith* 2007 5 SA 250 (CC), three HIV-positive women claimed that the respondents had violated their rights to privacy and dignity by publishing their names and HIV status. The South African Constitutional Court held that a person's HIV status must be afforded protection against insensitive disclosure in accordance with their rights to privacy. In South Africa, Section 14 of the Constitution, and the Protection of Personal Information Act 4 of 2013 (POPI), were

67 Computer facial and voice recognition allow for humans to communicate with machines for assistance. Some of the popular voice recognition software are Dragon, Siri, Goolge Now and Amazon Lex.

68 Nuffield council on bioethics <http://nuffieldbioethics.org/wp-content/uploads/Artificial-Intelligence-AI-in-healthcare-and-research.pdf> (Date of use: 31 March 2019).

introduced to protect the abuse of personal information and the infringement of an individual's right to privacy. Data collection practices must, therefore, be regulated in terms of legislation due to the dominant relationship that exists between AI and data.

Despite AI continuing to become an integral part of so many activities and subfields, there is limited academic research about AI regulation, and its rise has (to date) occurred in "a regulatory vacuum".⁶⁹ There are limited laws or regulations addressing the novel challenges introduced by AI, and the judicial system lags behind in developing standards to address who is legally liable (if the technology causes injury).⁷⁰ The benefit that AI technologies can offer against the growing awareness of legal and ethical concerns posed by it, can perhaps be addressed by designing legal frameworks and a universal code of conduct, aimed to address these risks. Legal structures can be applied to counter the societal risks that AI poses with not necessarily stifling the development of technologies.

1.5. *Background, literature review and problem statement*

The Organisation for Economic Co-operation and Development (OECD) that develops policies and standards on social and economic issues introduced AI principles in 2019, which was the first global principles agreed to by OECD member countries. The OECD "AI Principles" sets standards for AI, encompassing a value based approach for "inclusive growth, sustainable development and well-being".⁷¹ The G20 Ministerial Statement on Trade and Digital Economy has also introduced human-centred AI principles which are derived from the OECD's AI Principles.⁷² The United Nations Educational, Scientific and Cultural Organisation (UNESCO) that aims to establish peace through global education, social and cultural projects

69 Scherer 2016 *Harvard journal of law and technology* 356.

70 Scherer 2016 *Harvard journal of law and technology* 356.

71 OECD "Principles on AI" <http://www.oecd.org/going-digital/ai/principles/> (Date of use: 9 July 2019).

72 G20 Trade Ministers "G20 ministerial statement on trade and digital economy" <http://www.g20.utoronto.ca/2019/2019-g20-trade.html> (Date of use: 9 July 2019).

has also published statements following international forums on AI in Africa⁷³ and Beijing where the participant member states committed to:

[L]eading appropriate policy responses aimed at the systematic integration of AI and education to innovate education, teaching and learning, and at leveraging AI to accelerate the delivery of open and flexible education systems that enable equitable, relevant and quality lifelong learning opportunities.....⁷⁴

The aforementioned international guidelines on AI do not deal with specific approaches for the healthcare sector, which is a gap that should be addressed. In the South Africa's healthcare sector, hospitals are already utilising robotics assisted surgery for the removal of cancer.⁷⁵ Statutory regulators such as the Health Professions Council of South Africa (HPCSA), the National Department of Health (NDoH), the Nursing Council, the Pharmacy Council, and the National Health Research Ethics Council have developed ethics guidelines in line with their core mandates. However, there are no ethical guidelines specific to AI healthcare in South Africa and international ethical guidelines may provide useful guidance when developing ethical safeguards.

As mentioned, in order to determine the regulatory structure for AI it is essential to first define AI. There is currently no universally recognised definition among scientist and scholars in the field, and it is for this reason that this thesis proposes a working definition for the regulation of AI that encompasses the essence of human intelligence and the ability of AI technology to perform particular intellectual tasks.

Because humans are solely recognised as possessing intelligence, the definition of intelligence must be associated with human characteristics,⁷⁶ such as: understanding; self-perception; language usage; the ability to be taught; and comprehension skills⁷⁷ Therefore, the growing resemblance between humans and

73 Forum in Benguérir (Kingdom of Morocco) held on 12 and 13 December 2018.

74 UNESDOC Digital Library "Planning education in the AI era: Lead the leap" <https://unesdoc.unesco.org/ark:/48223/pf0000368303> (Date of use: 9 July 2019). The Outcome Document of the International Conference for AI and Education that took place in Beijing on 16 to 18 May 2019.

75 <https://urology.co.za/health-professionals/robotic-surgeons/> (Date of use: 9 July 2019).

76 Scherer 2016 *Harvard journal of law and technology* 359.

77 Scherer 2016 *Harvard journal of law and technology* 360.

intelligent machines may eventually accord legal recognition of computers as persons.

The United States of America (USA) has two stages for determining legal recognition. The first tier determines who are considered persons (i.e. fetuses). The second tier determines the nature of the rights and responsibilities vested in the legal persons premised on their legal capacities (i.e. mental incapacity; only eighteen-year-olds can vote).⁷⁸ In South Africa there is a difference divided into two groups of persons recognised by the law, i.e., natural persons (humans) and juristic persons (companies, trusts, etc.).⁷⁹ Intelligent computers will play an important role in society, but if they are not recognised as having legal personality, the issue of what rights are to be attributed to them and whether they should be entitled to rights at all, is not addressed.

Legal frameworks to deal with simple machines that make no decisions, already exist. For example, if a factory robot injures a worker, we will not consider blaming the robot. Factors such as the employer's safety protocols, the manufacturing or the design defect of the robot will be considered. These principles let us apportion blame and compensation for remedy considerations. However, characteristics of AI pose challenges for its regulation in terms of the current legal frameworks on liability.

AI has the ability to act autonomously, which creates a challenge when addressing the notion of foreseeability and causation.⁸⁰ AI systems are able to produce answers that people may not have thought of, and even their programmers or designers may not be capable of foreseeing or controlling what an AI system will do after it is no longer in their possession. If AI systems are so unpredictable and uncontrollable, can it be said that a programmer or manufacturer has been negligent or has caused foreseeable harm when self-learning systems learn to act in a way that it was not

78 Willick 1983 *The AI Magazine* 5.

79 Davel CJ & Jordaan RA *Law of Persons* 4th ed (Juta Cape Town 2005) 4.

80 Scherer 2016 *Harvard journal of law and technology* 363.

intended, whilst also leaving claimants with no recourse or remedies for their losses caused due to the system?⁸¹

The law is established on principles intended to control human behaviour and if employed to AI systems, may be inadequate.⁸² This could prove to have huge public risks in the healthcare industry, i.e., where an AI system makes an inaccurate diagnosis, causing injury or the death of a patient. Traditional delict claims in the domain of healthcare encompasses medical malpractice in the form of negligence, employer vicarious liability, and products liability. As with the factory robot that causes injury to a worker, patients would have recourse to recover damages from physicians, healthcare or pharmaceutical entities, or healthcare equipment manufacturers, if due to their failure to meet the legally accepted standards, a patient is injured.

Due to the fact that AI systems are not restricted by prescribed notions, as with humans, an AI system that cannot explain the route to its decision is regarded as a “black box”. Therefore, the current legal regimes for liability may not be adequate to deal with medical claims for negligence arising from the application of an AI black box.⁸³ As indicated by Scherer, it may not be reasonable to ascribe culpability to the creator of a component of the AI system whose work was far-detached in both proximity and time, from the system’s final design and performance.⁸⁴

In South Africa, in terms of strict product liability, producers; importers; distributors; or retailers are responsible for any injury or harm caused by products, regardless of whether the harm stemmed from negligence. Notwithstanding any contractual restrictions, if a product or its components are defective, and more than one actor is liable, they will be jointly-and-severally accountable for injury or damage, under

81 Delictual law in South Africa still considers fault to be a requirement for liability. It attributes fault through causation and foreseeability. For example, a person is liable for negligence if they caused foreseeable harm.

82 Bathae Y “The artificial intelligence black box and the failure of intent and causation” 2018 *Harvard journal of law and technology* 891.

83 Sullivan HR and Schweikart SJ “Are current tort liability doctrines adequate for addressing injury caused by AI?” https://journalofethics.ama-assn.org/sites/journalofethics.ama-assn.org/files/2019-01/hlaw1-1902_1.pdf (Date of use: 4 June 2019).

84 Scherer 2016 *Harvard journal of law and technology* 366.

the Consumer Protection Act (CPA),⁸⁵ or negligence in common law.⁸⁶ The CPA introduced statutory liability for defective products and provides that,

[...] the producer or importer, distributor or retailer of goods is liable for harm caused as a result of the supply of unsafe goods, a product failure, defect or hazard in goods, or insufficient instructions or warnings to the consumer relating to any hazard arising from or associated with the use of the goods, irrespective of whether the harm is the result of negligence on the part of any of these parties.⁸⁷

Complementing the common-law remedy relating to contract breaches, a consumer also has the right to execution of services in a way and based on an acceptable standard which they are ordinarily permitted to anticipate.⁸⁸ In accordance with the definition of “consumer”, “services”, and “goods” in the CPA, a patient is deemed to be a “consumer” when applying the CPA. Furthermore, “services” in the CPA is defined as: “any work or undertaking performed by one person for the direct or indirect benefit of another”. In the healthcare industry such services would, for example, include a consultation with a doctor or nurse, medical advice given by them, or medical procedures performed on a patient.

Therefore, the CPA envisages services as being performed by healthcare providers. This limitation would create difficulties in terms of the strict liability enforced by the CPA for medical services performed by AI that causes harm to patients. Due to the autonomous nature of AI as indicated earlier, there are also flaws in the protection provided to consumers under both the CPA and the common law.

The CPA offers a defence for distributors or suppliers of goods on the basis that, due to their part in only promoting same to the public, it would not be fair or reasonable to expect them to have unearthed the unsafeness, defectiveness, or hazard of the product.⁸⁹ In common law, the elements of delict have to be satisfied where fault and causation (which the consumer has the onus of proving) may be difficult to establish.

85 Consumer Protection Act 68 of 2008 (hereinafter referred to as the CPA).

86 The CPA has significantly moved from the common-law position and has adopted a strict product liability approach that excludes fault-based negligence.

87 Section 61.

88 Section 54(1)(b).

89 Section 61(4)(c).

The complexities of AI also bring to the fore issues of data and privacy. In terms of the colloquial definition referred to earlier, AI is directly connected with Machine Learning (ML), encompassing the ability to mimic cognitive human functions and to learn from human behaviours. Therefore, such technology is intrinsically dependent on personal data processing (mainly from consumers) allowing for tracing, observing, and profiling individuals, as well as forecasting behaviours.

AI ML allows for personal data processing to be executed in various approaches, and for objectives other than those that it was initially established for. This may result in the absence of control, manipulation, or abuse of personal information, and privacy rights.⁹⁰ The question is: how will industries or governments balance the voracious need for data with the protection of data and privacy rights?

The European General Data Protection Regulation (GDPR) was introduced in May 2018. It was the first major endeavour to provide consumers with greater protection of data under the law.⁹¹ In South Africa, an attempt has been made to set rules regarding control and administration of personal information, through the passing of the Protection of Personal Information Act (POPIA).⁹² The POPIA seeks to promote the security of personal information which is handled by all entities. The information regulator adopted by the POPIA published the final POPIA regulations on 14 December 2018.

However, the regulations have shortcomings as they appear to be more administrative in nature and do not assist organisations to practically interpret the Act. It has been left up to businesses to undergo the inexhaustible process of discovering, understanding and classifying their data, and based on that implement controls. The information regulator is required to develop the codes of conduct, for the various sectors relevant to the operation of the parties involved, but a particular

90 Russell and Norvig *Artificial intelligence* 3rd ed 1021. It was highlighted by the authors that, many people were not aware that they would be speaking with a computer and this posed a problem for legal authorities because of the capability to manipulate people into offering personal information allowing for their identity to be stolen or appropriated.

91 General Data Protection Regulation (EU) 2016/679 (hereinafter referred to as the GDPR) is law on data protection and privacy for all individuals citizens in the European Union (EU) and European Economic Area (EEA). It also regulates the export of personal data outside the EU and EEA jurisdictions.

92 Protection of Personal Information Act 4 of 2013 (hereinafter referred to as the POPIA).

sector or industry can also develop their own code of conduct for consideration and approval by the regulator.⁹³

A code of conduct for AI was published by the UK government in 2018 for the NHS that seeks to assure patients and healthcare workers that technology driven by data is protected, efficient, and upholds their privacy.⁹⁴ The code also allows companies compensatory access to the NHS data-sets, aimed to develop critical life-saving AI technology.⁹⁵ A comparative analysis is therefore required on how the existing governance structures in South Africa can address the complexities that advanced AI systems bring.

Despite all the advances of science and technology offered by AI, it also raises challenges for healthcare where offsetting the human touch could be an unintended casualty of such technology. If we consider that African societies are community-orientated, the displacement of doctors and healthcare workers could lead to social isolation and issues of trust in the technology. Bringing about an artificially intelligent healthcare landscape will be significantly complex, considering the socio-cultural impact it may have amongst the African communities.⁹⁶

On the other hand, realising equitable health and generating efficient healthcare for helpless people are essential societal objectives. On account of the impoverished working conditions in low and middle income countries (LMICs), particularly in remote regions, it remains challenging to entice and keep qualified healthcare providers in these environments.⁹⁷ Reducing the burden of costs and improving on doctors' productivity and the standard of healthcare are not the only envisaged potential gains of AI, as the technology could also be used by nurses and paramedical health workers to compensate for the shortage of doctors, in particular in Africa and other LMICs.

93 Section 60(1) of the POPIA.

94 UK GOV <https://www.gov.uk/government/news/new-code-of-conduct-for-artificial-intelligence-ai-systems-used-by-the-nhs> (Date of use: 20 May 2019).

95 The code published in September 2018 has 10 principles prescribing rules of engagement between the AI industry and healthcare system.

96 Guo J and Li B "The application of medical artificial intelligence technology in rural areas of developing countries" 2018 *Health equity* 176.

97 Guo and Li 2018 *Health equity* 176.

China's rural areas have started to enjoy the benefits of medical AI technology. An all-in-one diagnostic portable device station was developed by a healthcare company in China with the backing of the national rural healthcare program. The device automatically uploads the medical data results to an online platform which analyses and produces a diagnosis for the rural healthcare staff to evaluate and treat the patients.⁹⁸ Other technology companies in China have followed suit and are also participating in AI based healthcare facilities for rural areas, for example AI "chat-bots" in order to communicate with patients and provide medical consultations, or to even provide online training for healthcare workers in rural areas.⁹⁹

While these advances are a welcome benefit in terms of resolving the disparity between urban and rural healthcare, thereby responding to the principle of distributive justice, caution must be exercised with regard to the risk of human interaction being replaced in its entirety. The December 2018 statement published by UNESCO, following international forums on AI in Africa, urges African governments to engage in dialogue with partners that will encourage AI development that is inclusive of all social groups, in vulnerable situations, and to encourage the implementation of AI to foster policies that will strengthen intercultural dialogue and transformations. UNESCO Director General, Audrey Azoulay, states that,

Artificial intelligence can be a great opportunity to accelerate the achievement of sustainable development goals. But any technological revolution leads to new imbalances that we must anticipate.¹⁰⁰

Many of the above arising from AI are considered to be gaps under the current law that could be covered by using regulatory mechanisms. Even if AI is not simply predisposed to explicit regulatory oversight by an institutional agency, AI could still react to incidental regulation spurred by the law. Governments and judicial systems provide structures that are able to assist to steer the progression of AI in social- and economic positive approaches.¹⁰¹ If a legislative framework is established in order

98 Guo and Li 2018 *Health equity* 178.

99 Guo and Li 2018 *Health equity* 178.

100 UNESCO "Artificial intelligence: Towards a humanistic approach"
<https://en.unesco.org/artificial-intelligence> (Date of use: 9 July 2019).

101 Scherer 2016 *Harvard journal of law and technology* 376.

to regulate AI, the identification of role-players that are responsible for the policing thereof will also have to be considered.

1.6. Research problem

The research question of this thesis is: how should AI be regulated in South Africa? This question is central to the analysis conducted in this research. This work intends to answer this question and address the lacunae on the subject, with the intention of laying a robust foundation for subsequent research directions on the regulation of AI for the healthcare sector.

1.7. Rationale

Advancements in AI are continuous and this technology is evolving to be more sophisticated and proficient, inclusive of those fields that are perceived to necessitate creativity. Already, AI is applied in various areas of research and development and is entrenched into a broad offering of goods and services in healthcare. As intelligent machines operate more like humans, the rationality of regarding them as humans will gain popularity.

It is only a matter of time before an intelligent computer is regarded as a lawful subject, capable of adjudication before the courts and is regarded as “persons” in the public interest. Although AI has many benefits for healthcare, it also has the potential for causing significant harm because of system- or programming errors and their opaque nature, giving rise to legal issues. As AI positions itself in healthcare and becomes more intertwined into our day-to-day lives, there is an imminent necessity for grasping the legal problems associated with the technology’s use and for considering a legal framework that is informed by ethical values and norms to mitigate risks and to diffuse public concerns.

1.8. Aim of the study

The aim of this thesis is to critically analyse from a legal perspective, aided by ethical values and norms, how AI should be regulated in South Africa for healthcare.

1.9. Objectives

In addressing the aim of this study, the following objectives are relevant:

- To describe AI from a historical perspective, its current impact and future potential in healthcare;
- To discuss the probable advantages and potential complexities of AI, specifically in the field of healthcare from an ethico-regulatory perspective;
- To examine the notion of regulation and the possible impact of regulation on AI.
- To critically analyse how AI should be regulated in South Africa;
- To analyse pertinent laws and regulations in South Africa to identify gaps that should be addressed in the context of AI in healthcare;
- To analyse international- and national ethical guidelines and policies, in order to identify gaps and guidance for AI in healthcare, and to ascertain which norms and standards from these documents could be useful when developing ethico-legal safeguards for South Africa; and
- To discuss if it is possible for a regulation on AI and to analyse the parameters required for an intervention on AI to be effective.

In addition, this thesis attempts to answer the research question: “how should AI be regulated in South Africa?” by responding to specific sub-questions, such as: What kind of regulatory approach should be implemented for AI? What are the possible obstacles and advantages in regulating AI in the healthcare sector? Should AI be regulated as a separate area for impacted sectors, or should recognised areas of regulation be analysed with due regard to the impact of AI?¹⁰² This results in a risk evaluation, assessing the different regulatory approaches pertaining to AI in South Africa for the healthcare sector.

In discussing the sub-questions, the legislative framework is analysed, as contained in the Constitution and other statutes, policies, and international instruments. This begins to address the central question as to how AI ought to be regulated in South Africa. With due regard to the foundation that serves as a cornerstone for this

102 Nuffield council on bioethics <http://nuffieldbioethics.org/wp-content/uploads/Artificial-Intelligence-AI-in-healthcare-and-research.pdf> (Date of use: 31 March 2019).

research, it evaluates the steps that are required for an intermediation on AI to be effectively regulated.

1.10. Methodology

This research involves legal and prescriptive research and assessment. The method entails desktop and library-founded research. No original data from study participants is collected or analysed and the research does not involve study participants. This thesis draws from the law and ethical-based literature, relevant to the research. The research involves the interpretation and critical analysis of relevant legislation, case law, journals, articles, textbooks, and other relevant literature.

My critical examination of the relevant literature entails the definition and elucidation of concepts, the identification and critical analysis of assumptions, the evaluation of ethico-legal frameworks, the advancement and defences of opinions, and an articulation of the most plausible interpretation regarding significant concepts found in ethico-legal sources. A qualitative research approach is used in this study on existing literature, in order to evaluate the most efficient laws and guidelines that can be adopted to regulate AI in South Africa, as to mitigate negative outcomes for society, with the aim of analysing how law, ethical guidelines, and protected values would influence the field of AI.

As South Africa tends to lag behind in the AI technology field, there is limited literature and authority relating to the research topic of this thesis under South African law. A comparative analysis relating to the practices adopted by other countries is necessary, so as to understand their ethico-legal approaches. For academic research, the internet engine Google Scholar is used to search for scholarly literature. Sources consulted included books, journals, and articles, from various academic publishers, professional bodies, and universities. For historical and industry perspectives on the subject matter of this research, search engines such as Google and Yahoo were used. Key search terms, for example, artificial intelligence (AI); AI healthcare; AI regulation; ML; Big Data (BD); and healthcare technology were applied.

1.11. Limitations

Due to the resurgence of AI technology in the last century, previous studies and literature review in this research area are limited, which may impact on the scope of discussions. The existing materials are adequate for purposes of this research, in order to establish the foundation for further research. Currently, the advancements in AI technology and its application in different sectors are taking place at a rapid pace.

The research accommodates the pace at which AI technology is developing. The research requires an intermediate understanding of the technical aspects relating to the previous studies on AI. The researcher has rudimentary knowledge of related concepts on computer science. Academic sources are consulted for clarification on advanced technical aspects.

1.12. Expected contributions / outcomes

The research will be presented at local and international conferences. It is envisaged that several articles will emanate from the thesis for publication in scholarly journals. A framework in respect of the regulation of AI for the healthcare industry is to be presented to the NDoH.

1.13. Chapter summary

The research is comprised of five chapters, which are summarised as follows:

Chapter one offers an overview of AI from a definitional and historical perspective and considers its attributes and contemporary impact for healthcare. A synopsis is presented on the current status of the legal framework in respect of the regulation of AI and the perceived gaps, depicting the research question on how should AI be regulated in South Africa and outlining the methodology and expected contributions of the research.

Chapter two examines the potential challenges that AI may pose. It begins by considering the specific domains in which contemporary AI technologies are applied in the improvement of healthcare, whilst emphasising how they could pose a threat to certain fundamental rights and general social values. Chapter two signifies the

need for AI systems to be cautiously integrated with regulation, by those who participate in the creation and application of AI, with due regard to the end-users and populations who are impacted by the use of AI in healthcare.

Chapter three considers the notion of regulation and the existence of the different meanings to regulation that may cause certain difficulties in understanding it, and indicates the way in which the term “regulation” is used within the framework of the thesis. After defining regulation, the thesis discusses the reasons behind regulatory intervention for AI. It also recognises the shortcomings associated with seeking to impute accountability for the implementation of highly sophisticated socio-technical machinery, through regulation and the possible impact that a regulation may have on AI.

Chapter four discusses the underlying types of civil liability that arises when AI systems are used. The established elements of delictual liability are outlined to identify liability when injury is sustained through the deployment of AI systems. A comparative law study is undertaken on the liability principles based in delict arising from the use of AI under the common-law legal systems of the United States (US) and United Kingdom (UK). The chapter further examines the basis for a good regulation of AI in South Africa, by analysing the existing laws and regulations that creates legal gaps. There is also an analysis of international and national ethical guidelines and policies to identify gaps and guidance for the regulation of AI in healthcare, to ascertain which norms and standards from these documents could be useful when developing ethico-legal safeguards for South Africa.

Chapter five provides a summary and conclusions of discussions on the previous chapters, as well as recommendations regarding putting in place an effective and efficient legal framework for regulation of AI in the healthcare sector.

1.14. Ethical clearance

Ethical clearance for this study was obtained and approved. Enclosed as Annexure 1 herein, is the ethical clearance certificate bearing the clearance number STF 134 OF 2019.

2. CHAPTER TWO: BENEFITS AND CHALLENGES OF AI

2.1. *Current and potential benefits*

This chapter considers the benefits and ethico-legal challenges arising from AI in the healthcare sector. Max Tegmark, the President of the Future of Life Institute, states:

Everything we love about civilization is a product of intelligence, so amplifying our human intelligence with artificial intelligence has the potential of helping civilization flourish like never before – as long as we manage to keep the technology beneficial.¹⁰³

The goal towards ensuring that AI's impact remains beneficial for society inspires research in various fields (i.e. commerce and the law) as well as pragmatic subjects (such as: authentication; legitimacy; risk; and control).¹⁰⁴ Notwithstanding, one of the long-term objectives of most scientists are to outperform narrow AI,¹⁰⁵ designed to complete a single and specific task by creating “strong” or “general” AI, comprehending and analysing its environment in a similar manner a human being thus outperforming humans at nearly every cognitive task.¹⁰⁶

The long-term goal is spurred by the fact that AI, even in its narrow form, has become a driving force for economic- and social development, through huge

103 Future of Life Institute “Benefits and risks of artificial intelligence”
<https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/?cn-reloaded=1>
(Date of use: 11 April 2020).

104 Future of Life Institute “Benefits and risks of artificial intelligence”
<https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/?cn-reloaded=1>
(Date of use: 11 April 2020).

105 Narrow AI also known as weak AI, finds application in many commercial services such as weather forecasts, speech translation, self-driving cars, and is finding important application in medical diagnosis, all of which have already had significant societal and economic benefits.

106 Future of Life Institute “Benefits and risks of artificial intelligence”
<https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/?cn-reloaded=1>
(Date of use: 11 April 2020).

efficiency improvements,¹⁰⁷ productivity,¹⁰⁸ and wealth creation.¹⁰⁹ This has caused living standards to rise for many.¹¹⁰ It may also impact on certain kinds of occupations in distinct ways, by lowering the need for those skills that are capable of being automated, whilst improving the demand for other skills that would enhance AI.¹¹¹

The economic gains from technological progress for both employees and consumers in LMIC countries are significant. In these LMIC countries today, 28 per cent of the population is said to have access to the internet at home. On average, 8 in 10 people own a mobile phone with exceptional processing capacity and unlimited access to information, increased by emergent technology in disciplines such as AI.¹¹²

Born in the internet era,¹¹³ young people have new opportunities, where advances in technology can create jobs in the technology sector, as well as in ancillary sectors

107 The technology has introduced novel goods and services which offers efficacy and convenience to people's daily lives. From booking a taxi or flight, buying e-commerce products online, downloading music or games, all of which activities can now be done on your laptop or computer.

108 The technological revolution is already leading to long-term supply-chain advantages in productivity. For example, logistics and trade costs will decrease and become more efficient, and will create opportunity for new entrants into the markets and spur economic development. Schwab <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/> (Date of use: 30 March 2019).

109 Graetz G and Michaels G "Robots at work" https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2589780## (Date of use: 11 May 2020). According to Accenture AI has the potential to markedly increase expected baseline profit levels in 2035, for healthcare by 55 per cent. See Accenture "Why artificial intelligence is the future of growth" <https://www.accenture.com/us-en/insight-artificial-intelligence-future-growth> (Date of use: 16 May 2020).

110 Schwab <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/> (Date of use: 30 March 2019).

111 LMIC countries like India and China, are ideal candidates for participation in the AI evolution as a majority of the population has a good command of IT knowledge and they have much more qualified professionals. If these developing countries can provide technological qualified staff that can be outsourced to countries such as the US and UK, the developing countries will in turn be able to profit from technological change.

112 World development report 2016 "Digital dividends" <http://documents.worldbank.org/curated/en/896971468194972881/pdf/102725-PUB-Replacement-PUBLIC.pdf> (Date of use: 11 May 2020).

113 The World Economic Forum Global Agenda Council undertook the "Technological Tipping Points" survey in March 2015 and based on this survey the tipping points for technology is expected to occur by 2025 whereby 90 per cent of the population will be connected to the internet and will be using smartphones as a supercomputer in their pocket. World Economic Forum 2015 Survey Report "Deep Shift technology tipping points and societal

that rely on Information and Communications Technology (ICT).¹¹⁴ It also enhances output by augmenting workers skills, thus increasing human productivity and earning potential, which is especially critical for connecting poor people to work and markets.¹¹⁵ An automation revolution would not only help offset recent reduced productivity progress, experienced in various countries, but will also assist with the descending pressure on development from ageing populations.¹¹⁶

For consumers, automated processes will lower production and distribution costs of companies and allow them to exploit economies of scale, which will ultimately benefit consumers through lowering prices and by expanding the variety of available goods and services.¹¹⁷ Globally, many governments and organisations have invested billions of dollars in fostering the development of AI based on the shared notion that, the technologies can deliver exponential advantages towards improved productivity and service delivery.¹¹⁸

-
- impact” http://www3.weforum.org/docs/WEF_GAC15_Technological_Tipping_Points_report_2015.pdf (Date of use: 24 May 2020).
- 114 <http://documents.worldbank.org/curated/en/896971468194972881/pdf/102725-PUB-Replacement-PUBLIC.pdf> (Date of use: 11 May 2020).
- 115 The impact of AI systems on business is estimated to spur labor productivity by 40 per cent and will allow people to be more productive in the use of the time afforded. See 2016 <http://documents.worldbank.org/curated/en/896971468194972881/pdf/102725-PUB-Replacement-PUBLIC.pdf> (Date of use: 11 May 2020). Accenture “Why artificial intelligence is the future of growth” <https://www.accenture.com/us-en/insight-artificial-intelligence-future-growth> (Date of use: 16 May 2020).
- 116 Fin24 “Digital technology could create, rather than destroy, jobs – expert” <https://www.fin24.com/Tech/digital-technology-could-create-rather-than-destroy-jobs-expert-20170730> (Date of use: 11 May 2020).
- 117 <http://documents.worldbank.org/curated/en/896971468194972881/pdf/102725-PUB-Replacement-PUBLIC.pdf> (Date of use: 11 May 2020).
- 118 According to the 2020 EU White Paper on AI, in 2016, €3.2 billion was invested towards AI in Europe and the EU pledged to solicit a further €20 billion investment per year towards AI. See “White paper on artificial intelligence – A European approach to excellence and trust” https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf (Date of use: 6 June 2020). The UK government’s white paper, “Industrial strategy (2017) of the Department for Business, Energy and Industrial Strategy” committed to £1bn in AI Investment. See https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/664563/industrial-strategy-white-paper-web-ready-version.pdf (Date of use: 15 June 2020). A 2019 study requested by Microsoft from prepared Ernest and Young on AI maturity in the Middle East & Africa indicates that the UAE investment of USD\$2.15 billion has been the second highest in AI in the past decade. See Microsoft “AI in Middle Eastern and African markets, South Africa outlook for 2019 and beyond” <https://info.microsoft.com/rs/157-GQE-382/images/MicrosoftSouthAfricanreportSRGCM1070.pdf> (Date of use: 15 June 2020).

In South Africa, the comparative expenditure relating to AI premised on the value of transactions for the period 2008 to 2018 was US\$ 1,658 million, with at least 46 per cent of South African entities that are actively testing AI systems in their organisations and healthcare firms. This indicates that they expect to benefit 100 per cent from AI optimising their operations such as automating processes, monitoring results, predicting trends, prescribing solutions.¹¹⁹ According to Lillian Barnard, managing director for Microsoft South Africa:

AI will bring immense opportunity for South Africa. Our research shows that AI has the potential to solve some of the most pressing challenges that impact the country, driving development in sectors crucial to social and economic growth such as agriculture, healthcare, public services, and education, unlocking the huge potential that already exists here.¹²⁰

We already see digital technologies interfacing with the natural world on an everyday basis, where computer engineers and technicians are combining computational ingenuity with everything that humans consume (or inhabit) in their daily lives.¹²¹ Healthcare is a vital sector for societies and markets globally and given the constant digitalisation of various health data, AI has therefore already moved beyond the promise stage and can counter challenges relating to accessibility, expenditure and quality for healthcare.

Despite the early successes emanating from the implementation of these sophisticated technologies, there has also been increased public concern relating to the potential harm these technologies may cause to individuals and for society more generally.¹²² In this chapter I consider the benefits and challenges of AI, and its ethico-legal impact in relation to healthcare. The following section investigates

119 Microsoft “AI in Middle Eastern and African markets, South Africa outlook for 2019 and beyond” <https://info.microsoft.com/rs/157-GQE-382/images/MicrosoftSouthAfricanreportSRGCM1070.pdf> (Date of use: 15 June 2020).

120 Microsoft <https://info.microsoft.com/rs/157-GQE-382/images/MicrosoftSouthAfricanreportSRGCM1070.pdf> (Date of use: 15 June 2020).

121 Schwab <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/> (Date of use: 30 March 2019).

122 Scherer 2016 *Harvard journal of law and technology* 355; White House “Guidance for regulation of artificial intelligence applications 2020” <https://www.whitehouse.gov/wp-content/uploads/2020/01/Draft-OMB-Memo-on-Regulation-of-AI-1-7-19.pdf> (Date of use: 8 June 2020).

some of the AI technologies initiatives that are currently available, indicating the high potential for scale and impact in the healthcare sector.

2.1.1. Big Data and Machine Learning

“Big data” (BD) depicts voluminous sets of data which are capable of being collated and analysed by algorithms more efficiently than humans, as to reveal patterns and links, especially concerning studies into human behaviour.¹²³ BD is categorised by: volume (substantial quantities of data); variety (diversity within data); and velocity (enhanced accessibility to the data).¹²⁴

A significant volume of data is collected and stored in virtual clouds assimilated from multiple sources that can be harnessed by organisations for crucial information, in order to explore and create. To train an AI model in the application of healthcare, AI developers rely on a vast amount of health data which are directly connected to a patient’s other biomedical information. This data can be sourced from electronic health data, scientific trials, insurance or pharmaceutical records, and also information uploaded by patients through health applications.

However, BD is of no use without the comprehensive analysis thereof. Therefore, organisations are using AI algorithms to “mine” relevant data,¹²⁵ and ML can analyse and establish trends that may be of relevance.¹²⁶ In doing so, companies could stumble across business prospects that might give them an effective edge over their competitors in niche markets, by simply taking advantage of the data analysed through AI algorithmics. Algorithms are capable of learning specific data sets relevant to a business and offering maximum return on profits by accessing information to develop predictive models.

123 Oxford Dictionary https://www.lexico.com/definition/big_data (Date of use: 15 May 2020)

124 Price WN II “Artificial Intelligence in Health Care: Applications and Legal Implications” 2017 *The SciTech lawyer* 10.

125 Matsuzaki K “Ethical issues of artificial intelligence in medicine” 2018 *California Western law review* 257. The objective of mining data is to establish undetected trends and relationships from volumes of data sets and to derive a commercial value from it.

126 Machine learning (ML) applies self-thought algorithms to enhance performance over time at any given task through experience.

This capability to study and acclimatise makes AI the ideal choice for continuous patient performance monitoring and treatment improvements in healthcare. Data mining and AI will also support a world where medical treatments can be individually designed for each person's unique genetic sequence or mutations, behavioural profile and specific circumstances, and eliminate trial-and-error inefficiencies that inflate healthcare costs.¹²⁷

AI is proving to be invaluable to the healthcare sector due to the increased availability of patient data which is collected and evaluated. The capacity of certain government agencies are being enhanced as they apply AI to execute their tasks more swiftly, effectively, and inexpensively.¹²⁸ Current methods for populating vast quantities of data-sets concerning healthcare (i.e. surveys conducted by medical practitioners, scientists or patients) are costly, protracted, and discriminatory towards patients who already happen to be involved in the medical structure.¹²⁹

Social media platforms are fast becoming an alternate method for collecting large scale data at minimal cost. AI is used to analyse social media data and build predictive models to suggest behavioural and environmental impacts on health. AI is also able to rapidly integrate and analyse large data sets, for instance electronic medical data, or other forms of information gathered within public- or private health systems, that contain information such as the genomic-, economic-, and social data of populations.¹³⁰

127 According to an Accenture analysis, by 2026 AI health platforms could likely produce \$150 billion towards yearly reserves in the US, available at Accenture "Artificial intelligence: Healthcare's new nervous system"

https://www.accenture.com/t20171215T032059Z_w_us-en_acnmedia/PDF-49/Accenture-Health-Artificial-Intelligence.pdf (Date of use: 16 May 2020).

128 Keeping abreast with the new health data could cost doctors invaluable hours a week which precludes them from analysing new insights into clinical trials that can give an advantage in patient treatment or diagnosis.

129 Hager GD *et al* "Artificial intelligence for social good" <https://cra.org/ccc/wp-content/uploads/sites/2/2016/04/AI-for-Social-Good-Workshop-Report.pdf> (Date of use: 14 May 2020).

130 Hager *et al* <https://cra.org/ccc/wp-content/uploads/sites/2/2016/04/AI-for-Social-Good-Workshop-Report.pdf> (Date of use: 14 May 2020).

This provides an opportunity to accelerate healthcare from personalised medicine to personalised health by preventing people from visiting hospitals to begin with.¹³¹ A positive example of the application of BD relating to universal healthcare can be found in India's national identification programme.¹³² From 2010, India's government has implemented the "Aadhaar" card, issued to around 1.2 billion of its citizens. The card contains a unique biometric identification number that offers the opportunity of creating a platform to electronically gather and access medical data, and provides healthcare insurance information for low-income communities.¹³³

The card system also allows for the gathering of health statistics on India's voluminous population, which could be used for improvements in planning and service delivery to the public. Such a system could also be successfully applied in many other LMICs, as AI tools are dependent on significant amounts of historical information (to teach algorithms enabling AI technology to provide precise and efficient outputs relevant to the population and environment) particularly as health data are generally not available in LMICs.¹³⁴

More recently, the healthcare provider American Hospital Dubai, launched one of Dubai's first AI research facility. The facility aims to capitalise on BD to attain an enhanced awareness concerning the healthcare requirements of the United Arab Emirates (UAE) population, with the objective of enabling it to develop personalised healthcare and early treatment interventions, and to predict disease using assessment models and electronic health records.¹³⁵

131 This is possible through predicting future health outcomes based on historic data. Predictive analytics can be applied through the various stages of patient care or treatment and thus improves service delivery.

132 Wyber R *et al* "Big data in global health: improving health in low- and middle-income countries" <https://www.who.int/bulletin/volumes/93/3/14-139022/en/> (Date of use: 15 May 2020).

133 Wyber *et al* <https://www.who.int/bulletin/volumes/93/3/14-139022/en/> (Date of use: 15 May 2020).

134 USAID "Artificial intelligence in global health defining a collective path forward" <https://www.usaid.gov/cii/ai-in-global-health> (Date of use: 22 May 2020).

135 Rowe J "Cerner partners with Dubai hospital in AI research center" <https://ai.healthcareitnews.com/ai-powered-healthcare/cerner-partners-dubai-hospital-ai-research-center-0> (Date of use: 16 June 2020).

To date, one of the greater successes in leveraging the power of BD has been towards developing a vaccine of COVID-19, which has imminently impacted the world. With the assistance of open-science, researchers are able to track all the open-data and documents that are made freely available. They are also aggregating data using BD and analytics, to better understand the coronavirus from various viewpoints.

However, between trying to analyse (and track-and-trace) the virus, find a cure and oversee healthcare resources, researchers are also having to deal with volumes of data, making it challenging and almost impossible for humans to analyse without the assistance of BD.

2.1.2. Outcome prediction

A fundamental benefit of AI is its capability to make cogent predictions about future events. In addition to extracting data, algorithms can calculate and forecast outcomes based on a specific data set. Predictive patterning can offer clinicians with added insight into the treatment of patients at various stages of care, improving patient outcomes with the ultimate objective of limiting end-stage illnesses.¹³⁶

Some organisations are already using AI as a predicative tool in the field of healthcare.¹³⁷ One example is where AI algorithms are used to control and slow-down the spread of outbreaks and diseases by viewing, analysing, and reacting to health data in the present moment, instead of analysing data that may be obsolete at a later stage.

This is particularly crucial in LMICs where predictive tools can assist public health organisations to timeously intervene and prevent outbreaks of diseases and viruses

136 Penn Medicine, a healthcare group in the US, began using predictive analytics in 2017. The program they use called “Palliative Connect” analysis data using patient electronic health information to develop a prognosis based on a scoring system premised upon 30 factors and can forecast a patient’s likely prognosis over 6 months.

137 Based on 2017 study undertaken by the Society of Actuaries, 93 percent of health organisations agree that predictive analytics is a vital part of their future business Society of Actuaries “Predictive analytics in healthcare trend forecast” <https://www.soa.org/globalassets/assets/Files/programs/predictive-analytics/2017-health-care-trend.pdf> (Date of use: 15 May 2020)

such as tuberculosis or Ebola, which are prevalent to these countries.¹³⁸ The Canadian firm, BlueDot, was able to issue a warning about COVID-19 days ahead of the official alerts from the World Health Organisation (WHO). This is because it was able to process different sources of data globally, through social media platforms or official statistics about the number of cases reported.¹³⁹ The company uses AI-driven algorithms to predict possible outbreaks of infectious diseases by searching for patterns in the data and then warning clients of traveling in at-risk areas.

2.1.3. Operational efficiency

AI not only finds its use in back-end operations, but it can also be applied at the front-end where users can interact in real time, yielding maximum efficiency. A large part of many businesses entail responding to clients' queries promptly. Increasing the workforce could address the problem, but it will increase expenditures exponentially.¹⁴⁰

A major contributor to the increase of costs for hospitals in South Africa is the demand for services and absorption of resources.¹⁴¹ AI has the ability to increase system efficiency and to reduce resource wastage. With the upsurge of healthcare expenses, many patients are not prepared to endure oversights and setbacks and are resorting to conventional doctors for any treatment. AI-based tools would deal with inevitable delays through the implementation of AI robots,¹⁴² AI diagnostic

138 South Africa identifies four historically entrenched epidemics in the country as causes of early mortality. See Valiani S "Public health care spending in South Africa and the impact on nurses: 25 years of democracy?" https://hasa.co.za/wp-content/uploads/2020/02/Public_Health_Care_Spending_in_South_Afr.pdf (Date of use: 18 May 2020).

139 Pressman A "How AI is aiding the Coronavirus fight" <https://fortune.com/2020/03/16/ai-coronavirus-health-technology-pandemic-prediction/> (Date of use: 13 May 2020).

140 Schwab <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/> ((Date of use: 30 March 2019).

141 Rural Health Advocacy Project https://www.groundup.org.za/media/uploads/documents/Austerity%20report_FINAL.pdf (Date of use: 18 May 2020).

142 According to Forbes, robots can study historic medical data on a patient and using this it is able to offer guidance to surgeons as a tool in performing operations, decreasing the patient's in-hospital stay by 21%. Marr B "How is AI used in healthcare" <https://www.forbes.com/sites/bernardmarr/2018/07/27/how-is-ai-used-in-healthcare-5-powerful-real-world-examples-that-show-the-latest-advances/#3756ee895dfb> (Date of use: 13 May 2020).

algorithms,¹⁴³ and AI being integrated in the technology to make it more “cognitive” to assist patients in real time.¹⁴⁴

These AI algorithms are also to perform tedious administrative duties,¹⁴⁵ thereby allowing limited resources such as doctors, to be available to cater to the needs of more important and complex tasks.¹⁴⁶ The NHS in the UK is establishing a research facility to develop AI systems with the goal of addressing some of the major challenges encountered by the NHS, such as advancing cancer diagnosis, detecting patients with high risk of cardiac disease or mental illness, and automating administrative functions, allowing healthcare workers to engage more often with patients.

At the forefront three of the applications that embody the highest near-term developments in AI are robotic-assisted surgery, virtual nursing care takers, and administrative assistants.¹⁴⁷ As AI applications like these become more progressive, their capacity to consume knowledge and perform will always develop into improvements in relation to productivity and cost reduction.¹⁴⁸ The diagnostic- and

143 Analytical image analysis assists in rural settings or remote areas where people have limited reach to healthcare. Patients with smart phones can remotely upload and share photos of ailments such as lesions or bruises to determine what treatment is required through AI image scanning on the cloud.

144 Virtual nursing or chat-bots application such as that of Care Angel's virtual nurse assistant, allow for cheaper and frequent interaction between patients and healthcare providers as an alternative to consultation room visits to prevent unnecessary hospital stays or visits. There is no down time with virtual assistance as it is available 24/7 and is able to answer to queries in a quicker time frame and at a reduced cost.

145 It is anticipated that this may generate a savings of USD\$18 billion for the healthcare sector as AI can alleviate certain administrative tasks that healthcare providers do not have to deal with. Marr <https://www.forbes.com/sites/bernardmarr/2018/07/27/how-is-ai-used-in-healthcare-5-powerful-real-world-examples-that-show-the-latest-advances/#3756ee895dfb> (Date of use: 13 May 2020).

146 New Scientist “The NHS is setting up a lab for medical artificial intelligence” <https://www.newscientist.com/article/2212928-the-nhs-is-setting-up-a-lab-for-medical-artificial-intelligence/> (Date of use: 20 May 2020).

147 Accenture https://www.accenture.com/t20171215T032059Z__w_/us-en/_acnmedia/PDF-49/Accenture-Health-Artificial-Intelligence.pdf (Date of use: 16 May 2020). According to the 2015 survey report issued by the World Economic Forum, the first robotic pharmacist in the US is expected by 2021, available at World Economic Forum 2015 Survey Report http://www3.weforum.org/docs/WEF_GAC15_Technological_Tipping_Points_report_2015.pdf (Date of use: 24 May 2020).

148 Accenture estimates that “AI-assisted technologies are expected to save the global healthcare industry approximately US\$150 billion a year by 2026. That the top costs savers will come from robot assisted surgery, virtual nursing assistants’ administrative workflow assistance, fraud detection, dosage error reduction”. See Accenture “Artificial intelligence: Healthcare’s new nervous system”

treatment abilities of physicians are greatly enhanced by mobile medical applications that rely on intelligent AI engines.

Instead of waiting weeks or even months to see an expensive or inaccessible specialist, patients can see a physician assistant trained to work in conjunction with a smart machine such as IBM's Watson.¹⁴⁹ For countries such as South Africa (with large rural populations and inadequate access to healthcare) AI abilities provide a novel method to counter the scarcity of healthcare professionals. This is more of a priority in these communities and add value in enhancing capabilities to deal with patients by providing cost effective health advice and eventually even early treatment and diagnoses, without patients having to travel to or avoiding unnecessary trips to health facilities.¹⁵⁰

In return, this addresses the issue of patient overload on already over capacitated facilities and allows healthcare providers to focus on their high-risk patients and to capitalise on their time and effort. For example, human immunodeficiency virus (HIV) in South Africa is widespread, with approximately 7 million of its inhabitants living with the disease.

The success of HIV treatments is highly dependent on regular visits to clinics and hospitals, in order to ensure that antiretroviral (ARV) treatment is uninterrupted. With the assistance of robotic pharmacists, drugs are now being readily dispensed on the streets to those that are infected with HIV in South Africa. By scanning in smart card ID's patients no longer need to travel long distances or que for long hours at clinics to collect their monthly medication.

The robots are also designed to not identify them for the purpose of HIV to avoid discrimination or social stigma linked with the disease.¹⁵¹ Similarly, a company in

-
- 149 https://www.accenture.com/t20171215T032059Z_w_us-en_acnmedia/PDF-49/Accenture-Health-Artificial-Intelligence.pdf (Date of use: 16 May 2020).
- 150 Khan F "Medicine in the age of smart machines: legal liability challenges" in Olleross FX and Zhegu M (eds) *Research Handbook On Digital Transformations* (Edward Elgar Publishing 2016) 68.
- 151 Guo and Li 2018 *Health equity* 176-177.
- 151 Basu M "Robot dispenses drugs to HIV patients in South Africa" <https://govinsider.asia/smart-gov/robot-dispenses-drugs-to-hiv-patients-in-south-africa/> (Date of use: 3 July 2020).

South Africa called Right ePharmacy is able to deliver pharmaceutical care to patients in remote locations where they do not have easy access to a pharmacist. Its cloud-based technology enables remote dispensing of medication by automated robotics, drug therapy monitoring and real-time patient counselling by linking the patient to a remote pharmacist via an audio-video link.¹⁵²

2.1.4. Precision diagnostics and error mitigation

Of the more significant contributions of AI is its ability to eradicate or minimise errors from processes with the likelihood of attaining a greater level of accuracy as opposed to humans. AI constructed solutions can help to reduce human and production related costs of any oversights. AI can play a vital role in well-timed detection and accurate diagnosing of life-threatening illnesses like cancer,¹⁵³ whose care and deterrence largely hinge on early detection of the symptoms.¹⁵⁴

According to a report published by the WHO, 2.6 million people are projected to die each year in LMICs, as a result of medical mistakes, linked to misdiagnosis and the maladministration of medical products.¹⁵⁵

Effective and timely diagnosis can cure or prevent these illnesses completely, whereas a late or incorrect diagnosis can have severe or even fatal consequences. AI algorithms can rapidly study and scrutinise volumes of samples and data sources in succession and formulate useful patterns,¹⁵⁶ and can thus be especially useful in hospitals that do not have the adequate healthcare infrastructure and resources.

More recently, AI applications have proven to be effective in diagnosing diseases. Lancet Digital Health conducted a study that compared the performance of AI deep

152 Right ePharmacy <https://rightepharma.co.za/solutions/tele-pharmacy/> (Date of use: 3 July 2020).

153 It has already been proven that IBM's Watson is capable of a far more precise diagnosis for lung cancers than that of humans.

154 Chatterjee A "Use of artificial intelligence to reduce medical errors"
<https://blog.myhealthvectors.com/index.php/use-of-artificial-intelligence-to-reduce-medical-errors/> (Date of use: 13 May 2020).

155 The Vaccine Reaction "Medical Errors Kill Five People Per Minute, 2.6 Million People Every Year" <https://thevaccinereaction.org/2019/11/medical-errors-kill-five-people-per-minute-2-6-million-people-every-year/> (Date of use: 13 May 2020).

156 Chatterjee <https://blog.myhealthvectors.com/index.php/use-of-artificial-intelligence-to-reduce-medical-errors/> (Date of use: 13 May 2020).

learning in identifying diseases from analytical medical imaging to that of healthcare professionals. The study concluded that, in the preceding years, AI has become more precise in recognising disease diagnosis by way of such images and found the diagnostic capability of the deep learning forms equals that of healthcare professionals.¹⁵⁷

It can be said that the healthcare system in South Africa, as is common in many other health systems universally, for most part operates on a reactive system of care that is focused on treating patients that are already ill instead of preventing illnesses before they occur. According to Erik Roos, Chief Executive of South African pharmaceutical firm Pharma Dynamics:

Curbing the onset of illness, especially diseases of lifestyle, such as high blood pressure, heart disease and diabetes, is the holy grail of healthcare transformation and sustainability. Proactive care solutions should include screening individuals based on known algorithms to ensure preventive action is taken long before the onset of symptoms or disease.¹⁵⁸

2.1.5. Application of AI for social good

AI can play a major role in contributing towards social good. This is premised, in part, on the manner the technology is developed.¹⁵⁹ Increased investments in AI-advanced research in the public- and private domain have already incited major advantages towards the public in critical fields such as healthcare.¹⁶⁰ To illustrate

157 Liu X *et al* "A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis" <https://www.thelancet.com/action/showPdf?pii=S2589-7500%2819%2930123-2> (Date of use: 16 June 2020).

158 IOL "What 2019 holds in store for pharmaceutical sector" <https://www.iol.co.za/business-report/economy/what-2019-holds-in-store-for-pharmaceutical-sector-19127833> (Date of use: 15 June 2020).

159 Hager *et al* <https://cra.org/ccc/wp-content/uploads/sites/2/2016/04/AI-for-Social-Good-Workshop-Report.pdf> (Date of use: 14 May 2020). The potential advantages of enhancing digital technologies access appear further in the "Digital Dividends Report" of the World Bank Group. See <http://documents.worldbank.org/curated/en/896971468194972881/pdf/102725-PUB-Replacement-PUBLIC.pdf> (Date of use: 11 May 2020).

160 Scherer 2016 *Harvard journal of law and technology* 354. According to Accenture, by 2021 the total investment in AI healthcare by public-private sector is expected to reach USD\$6.6 billion. See Forbes "AI and healthcare: a giant opportunity" <https://www.forbes.com/sites/insights-intelai/2019/02/11/ai-and-healthcare-a-giant-opportunity/#56d362cd4c68> (Date of use: 23 March 2019).

the range of fields where AI is able to promote social good, this section highlights a few ongoing advancements in AI that finds application.

2.1.5.1. Helping the visually impaired

Not being able to see is a major impediment for many across the globe. Globally, approximately 2.2 billion people suffer from vision impairment or loss of sight, and at least 1 billion of those people could have been spared from such vision impairment.¹⁶¹ It is said that a projected that 250 million people globally have minor to extreme visual deficiencies, of which 90 per cent reside in LMIC settings.¹⁶²

AI can alleviate some of the challenges for those that suffer from visual impairment through navigational assistance on smartphones or other devices.¹⁶³ AI's vision technology can detect objects and can translate handwritten or printed text to digital text, which can then be electronically voiced. Examples of the abovementioned technology are already in the market such as Microsoft's Seeing AI, globally accessible to users, at no cost. Other similar applications include the OrCam MyEye camera, which operates without the need for a smartphone and is mounted on standard spectacles which work to convert what is seen into spoken output. As technology such as these advances, it will offer the visually impaired insight into their environment as well as the opportunity to identify other people and determine colours.

2.1.5.2. Computer vision for skin cancer diagnosis

Skin cancer is a highly prevalent type of disease. If diagnosed early enough, survival rates for such cancer are said to be as high as 97 per cent, which drops to as little

161 WHO "Blindness and vision impairment" <https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment> (Date of use: 16 May 2020).

162 Mckinsey Global Institute "Applying AI for social good" <https://www.mckinsey.com/~/media/mckinsey/featured%20insights/artificial%20intelligence/applying%20artificial%20intelligence%20for%20social%20good/mgi-applying-ai-for-social-good-discussion-paper-dec-2018.ashx> (Date of use: 16 May 2020).

163 Mckinsey Global Institute <https://www.mckinsey.com/~/media/mckinsey/featured%20insights/artificial%20intelligence/applying%20artificial%20intelligence%20for%20social%20good/mgi-applying-ai-for-social-good-discussion-paper-dec-2018.ashx> (Date of use: 16 May 2020).

as 14 per cent due to late detection.¹⁶⁴ Today skin cancer is largely detected by dermatologists with a device referred to as a “dermatoscope”. AI is advancing how dermatologists diagnose skin conditions, including serious cancers like melanoma.

Korean researchers recently invented an AI deep learning algorithm that when used in conjunction with a smartphone, is capable of classifying skin disorders with precision, predicting malignancy, recommending treatment choices, and serves as a complementary aid to augment diagnostic accuracy using clinical imaging analysis. With this system, diagnostic precision of dermatologists has drastically improved.¹⁶⁵ This technology could also serve as a mobile application using image identification to make examination accessible to everyone, particularly rural populations globally that have limited access to skin doctors and are consequentially at an increased risk of late detection.

2.1.5.3. Detecting water service lines containing lead

A primary objective of the WHO is that,

all people, whatever their stage of development and their social and economic conditions, have the right to have access to an adequate supply of safe drinking water.¹⁶⁶

Based on studies conducted, a significant amount of people’s daily consumption of lead is obtained from their intake of water.¹⁶⁷ Lead is a cumulative general poison with adverse effects on the health of young children and pregnant women.¹⁶⁸ Ineffective management of municipal-, industrialised-, and agricultural waste water

164 Mckinsey Global Institute https://www.mckinsey.com/~media/mckinsey/featured_%20insights/artificial%20intelligence/applying%20artificial%20intelligence%20for%20social%20good/mgi-applying-ai-for-social-good-discussion-paper-dec-2018.ashx (Date of use: 16 May 2020).

165 Science Daily <https://www.sciencedaily.com/releases/2020/03/200331092704.htm> (Date of use: 16 May 2020).

166 WHO “Guidelines for drinking water quality” https://www.who.int/water_sanitation_health/dwq/chemicals/lead.pdf (Date of use: 16 May 2020).

167 WHO https://www.who.int/water_sanitation_health/dwq/chemicals/lead.pdf (Date of use: 16 May 2020).

168 WHO https://www.who.int/water_sanitation_health/dwq/chemicals/lead.pdf (Date of use: 16 May 2020).

has caused the drinking-water of millions of individuals being severely chemically polluted.¹⁶⁹

In the water crisis in Flint, Michigan, hazardous levels of lead seeped from old pipes whilst the council neglected to treat the water, creating a public health disaster that threatened the health of thousands of residents of the city.¹⁷⁰ Close to 9,000 children were subjected to the risk of impairment to brain development, compromised learning abilities, and behavioural conditions. Lead was transported through water lines which linked up the water supply to residents of the city.

However, locating and replacing water lines required excavation that was too expensive.¹⁷¹ Through advanced AI analytics, a model called Active Remediation, was developed by the University of Michigan, capable of predicting with 98 per cent precision whether a water line contains lead, using data analytics science that forecasts which residences bear lead pipes.¹⁷² Using this technology it was projected that three out of four residences in Flint contained lead in their water lines.

By implementing the predictive technology in Flint it drastically reduced the need for costly replacement excavations.¹⁷³ The model could be applied for other countries and would save money and ensure that clean drinking water is accessible, particularly for residents in impoverished countries that have limited access to safe sources of drinking-water.

169 WHO "Drinking water" <https://www.who.int/news-room/fact-sheets/detail/drinking-water> (Date of use: 16 May 2020).

170 National Geographic "Five years on, the Flint water crisis is nowhere near over" <https://www.nationalgeographic.com/environment/2019/04/flint-water-crisis-fifth-anniversary-flint-river-pollution/> (Date of use: 16 May 2020).

171 Mckinsey Global Institute https://www.mckinsey.com/~media/mckinsey/featured_sights/artificial%20intelligence/applying%20artificial%20intelligence%20for%20social%20good/mgi-applying-ai-for-social-good-discussion-paper-dec-2018.ashx (Date of use: 16 May 2020).

172 The AI technology uses a combination of data sources from the city and survey data to predict the homes most likely to have lead pipes

173 Abernethy J *et al* "Active remediation: The search for lead pipes in Flint, Michigan" <https://arxiv.org/pdf/1806.10692.pdf> (Date of use: 16 May 2020).

2.1.5.4. Fraud detection in healthcare

Healthcare fraud is often difficult to detect, monitor, and prevent. In South Africa the situation is exacerbated as fraud detection and mitigation across the medical schemes and administrators are fragmented and require consolidated efforts through an industry-wide central database to facilitate collective use and analysis of data.¹⁷⁴ Investigating claims is time consuming and costly and fraudulent claims translate to over-billing for services or supplies and out-of-pocket expenses for consumers, slower processing of valid claims, and higher healthcare premiums for patients.¹⁷⁵

According to Dr Sipho Kabane, the acting Executive and Registrar for the Council of Medical Schemes, as a result of fraud, wasteful expenditure and misuse of funds the healthcare industry is suffering losses of more than R 22 billion per annum.¹⁷⁶ The UK's NHS loses £ 1.27 billion every year due to fraud committed by patients, staff and contractors.¹⁷⁷ In the US the National Health Care Anti-Fraud Association, has estimated healthcare fraud to be over USD\$ 300 billion.¹⁷⁸

Apart from the financial losses, individual victims of healthcare fraud are also subjected to unwarranted or unsafe medical treatment, false patient diagnosis or

174 Council for Medical Schemes "Fraud in the medical schemes industry" <https://www.medicalschemes.com/files/CMS%20News/CMSNews1of2018.pdf> (Date of use: 22 June 2020).

175 NHCAA "The challenge of health care fraud" <https://www.nhcaa.org/resources/health-care-anti-fraud-resources/the-challenge-of-health-care-fraud/> (Date of use: 16 May 2020).

176 Medical Academic "ZAR R22 billion lost to fraud waste and abuse" <https://www.medicalacademic.co.za/news/r22-billion-lost-to-fraud-waste-and-abuse/> (Date of use: 22 June 2020).

177 NHS "Counter Fraud Authority Annual Report & Accounts 2018-19" https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/859362/nhscfa-annual-report-and-accounts-2018-2019.pdf (Date of use: 22 June 2020).

178 NHCAA "The challenge of health care fraud" <https://www.nhcaa.org/resources/health-care-anti-fraud-resources/the-challenge-of-health-care-fraud/> Date of use: 16 May 2020).

their insurance records are appropriated to submit fraudulent claims through medical identity theft,¹⁷⁹ which is becoming increasingly more sophisticated.¹⁸⁰

AI fraud analytical and detection tools are gaining popularity as they can be used to trace and detect behaviours associated with fraud, using sophisticated applications that scrutinise and learn from data over time.¹⁸¹ This would allow the sector to identify risks and protect applications and data from breach.¹⁸² South African healthcare providers such as Momentum Health, have already started to invest in BD analytics capabilities which uses behavioural analytics to identify healthcare providers and members exhibiting suspicious behaviours, indicative of fraud and/or abuse.¹⁸³

With the assistance of AI tools, early detection and resolution of issues concerning fraud and/or abuse can save the healthcare sector money, and at the same time ensuring that services are not compromised.

2.1.5.5. AI and mental health

The South African Depression and Anxiety Group (SADAG) states that, only 16 per cent of people that suffer with mental related illnesses obtain therapy and 85 per

179 According to Mimecast, 63 per cent of entities in South Africa noted a rise in identification fraud, 84% were exposed to email phishing and a 19 per cent year on year increase in ransomware attacks, aimed at South African companies https://info.mimecast.com/sa-the-state-of-email-security.html?utm_medium=SEMPPC&utm_source=GooglePPC&utm_campaign=7013100000TmCuAAK&utm_term=email%20security&qclid=EAlaIqobChMI_K_v5oW46QIVwu3tCh2jzqWiEAAAYASAAEgJL7PD_BwE (Date of use: 15 May 2020).

180 Council for Medical Schemes
<https://www.medicalschemes.com/files/CMS%20News/CMSNews1of2018.pdf> (Date of use: 22 June 2020).

181 ML models can assist with automating claims assessment and detecting existing fraud patterns. This model flags possible fraud claims required for further review. The technology can also automatically identify valid transactions which allows for fast-tracking of approval and settlement.

182 Bullock L “The top 6 ways that artificial intelligence will affect your business in the near future” <https://www.forbes.com/sites/lilachbullock/2019/02/25/the-top-6-ways-that-artificial-intelligence-will-affect-your-business-in-the-near-future/#4eb1811a1966> (Date of use: 15 May 2020).

183 Council for Medical Schemes
<https://www.medicalschemes.com/files/CMS%20News/CMSNews1of2018.pdf> (Date of use: 22 June 2020).

cent of patients are forced to rely on public health facilities.¹⁸⁴ Despite a dire need for mental health treatment, South Africa's resources are not adequately resourced to carry the burden of providing such services.

In the US over USD\$ 201 billion is spent on mental health services each year rendering it as the most costly component for its healthcare system.¹⁸⁵ The potential benefit of AI in the area of medicine has already been demonstrated by its promise in diagnosing disease, interpreting images and data for diagnosis, and focusing on treatment plans. Researchers are now looking beyond medicine and evaluating various ways in which AI can assist to identify, diagnose and heal mental illness.¹⁸⁶

Although mental health is mainly a distinctively subjective area, necessitating emotional aptitude and insight that current computers cannot mimic, researchers argue that the field could benefit from AI's aptitude to analyse data sets and detect patterns and subtle warnings that are unnoticeable to humans.¹⁸⁷ Mental health professionals would benefit for algorithms that could analyse data much faster for diagnosis, recommend treatments options, and screen a patient's development remotely and more frequently.¹⁸⁸

The result will be more individualised, cost effective management and treatment of many common mental health problems and improved outcomes. This will be particularly beneficial for those with no or limited access to mental health professionals, and those whom often avoid seeking assistance due to financial constraints.

184 SACAP "The shocking state of mental health in South Africa in 2019" <https://www.sacap.edu.za/blog/management-leadership/mental-health-south-africa/> (Date of use: 15 May 2020).

185 Marr B "The incredible ways artificial intelligence is now used in mental health" <https://www.forbes.com/sites/bernardmarr/2019/05/03/the-incredible-ways-artificial-intelligence-is-now-used-in-mental-health/#bf6b0b1d02e4> (Date of use: 16 May 2020).

186 Marr <https://www.forbes.com/sites/bernardmarr/2019/05/03/the-incredible-ways-artificial-intelligence-is-now-used-in-mental-health/#bf6b0b1d02e4> (Date of use: 16 May 2020).

187 Ducharme J "Artificial intelligence could help solve America's impending mental health crisis" <https://time.com/5727535/artificial-intelligence-psychiatry/> (Date of use: 16 May 2020).

188 Marr <https://www.forbes.com/sites/bernardmarr/2019/05/03/the-incredible-ways-artificial-intelligence-is-now-used-in-mental-health/#bf6b0b1d02e4> (Date of use: 16 May 2020).

2.1.6. The impact on business

The accelerated innovation and rate of disruption of AI renders it difficult to comprehend or anticipate the effect of its applications, even for the most connected and advanced businesses. However, across all industries, ML and automation offer significant opportunities for businesses, offering the ability to review their environment, to be prepared to rapidly, implement required changes, and to minimise disruption.

Managing cost containment while also improving access to healthcare can be achieved if health systems are innovative in supply and distribution. With AI becoming more proficient at analysing and interpreting volumes of data at a rapid pace, businesses will be able to develop more personalised and cost-effective solutions particularly in healthcare, by capitalising on the information interpreted from certain data sets. As AI applications produce trends and forecasting from data, numbers and imaging, businesses will be able to apportion the appropriate budget and resources premised on the predicative outcomes of AI, thus improving on decision making.¹⁸⁹

On the supply side, AI technologies is significantly disrupting industry value chains by creating novel ways of addressing existing demands. AI is enabling heightened dexterity and precision through the automation of labour-intensive processes.¹⁹⁰ Due to the universal digital landscape for research, marketing, sales and supply, businesses can outbid their competitors by refining the form, pace, or costs for delivery of commodities towards which value is supplied.¹⁹¹

189 Bullock <https://www.forbes.com/sites/lilachbullock/2019/02/25/the-top-6-ways-that-artificial-intelligence-will-affect-your-business-in-the-near-future/#4eb1811a1966> (Date of use: 15 May 2020).

190 From AI inventory-taking drones to driverless warehouse vehicles and data analytics, AI is able to make the entire supply chain and production line more efficient. According to Amazon, the robots used at its storage warehouses allows it to store 40 per cent more inventory, and in turn enables it to fulfil one- or two-day shipping orders even faster.

191 Schwab <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/> (Date of use: 30 March 2019).

Inventory management and distribution for medicines and treatments can be vastly improved through analytical predictions made by ML applications.¹⁹² As the need for new medications intensifies, pharmaceutical industries across the world are investing in automated machines and robotic systems, to increase productivity of pharmaceuticals and medical devices. These automated systems are capable of detecting cracks on tablets, colour checking, shape recognition, counting of tablets and inspecting for spills in packaging drugs resulting in reduced production downtime, increased health and safety, better waste management and lower operating costs.¹⁹³

In 2017, the Rashid hospital in Dubai introduced its first robotic dispensing chemist designed to store up to 35,000 medicines and distributing 12 prescriptions in a minute decreasing the consumer's waiting time. The robot also dispenses prescribed medication based on a barcode system, minimising any human error.¹⁹⁴

In South Africa, amongst the growing issues which the pharmaceutical sector encounters are conflicting interest between enhancing access to inexpensive medications and expanding the production of pharmaceuticals.¹⁹⁵ In view of these challenges, more health firms will need to consider combining AI devices and supply of medicines.

On the demand side, with increasing availability to sophisticated smartphones and data, AI is rendered easier to use by consumers. This places companies under pressure to acclimatise to the way they develop, market, and distribute goods and services.¹⁹⁶ Customer engagement continues to improve twofold, as using

192 For instance, Tanzania and Zambia maintain electronic immunization data registries which can be used by predictive ML models to forecast vaccine usage in their health facilities months ahead of time. These predictions close the gap between the supply-and-demand for vaccines and prevent stock shortages.

193 Markets and Markets “Pharmaceutical robots market” <https://www.marketsandmarkets.com/Market-Reports/pharmaceutical-robot-market-210985096.html> (Date of use: 15 June 2020).

194 Dubai Health Authority <https://dha.gov.ae/en/RashidHospital/Pages/Pharmacy.aspx> (Date of use: 16 June 2020).

195 IOL <https://www.iol.co.za/business-report/economy/what-2019-holds-in-store-for-pharmaceutical-sector-19127833> (Date of use: 15 June 2020).

196 Schwab <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/> (Date of use: 30 March 2019).

automated live-chat software assists with the delivery of more accurate and much quicker response times, and helps build stronger engagement and consumer trust.

Another advantage is that AI allows for predictive insights that can resolve the problem for healthcare providers that are trying to prevent unnecessary readmissions and improve patient care outcomes. For staff, product replenishment, for businesses such as pharmacies or dispensaries, can be made more efficient by using predictive analytics.¹⁹⁷ Ultimately, the technology platforms integrate people, assets, and data, and they aid in reducing the barriers for corporations and individuals to generate wealth.¹⁹⁸

According to Schwab,¹⁹⁹ the inevitable change from easy digitisation of the third industrial revolution to an evolution of technologies centred on amalgamations of the fourth industrial revolution, convinces businesses to rethink how they should approach the implementation of their enterprise. New technologies such as AI, make products and services more sustainable and robust, while ML data analytics are altering how they are preserved.²⁰⁰

Ultimately, AI will force business executives to begin to examine and analyse their fluctuating operations, test the assumptions of their operational staff and to consistently innovate.

2.1.7. The role of AI health services in rural communities

Deloitte's 2019 Global Healthcare Outlook states that, various public sector healthcare structures worldwide are under financial pressure relating to issues of:

Accessibility (imbalanced distribution, including a rural-urban divide), affordability (especially for patients with low economic status), awareness (of lifestyle diseases, risk

197 An Australian company Black.ai, has also developed AI technology that maps a store in real-time with 3D sensors in the ceiling to track what customers purchase and logs when the best times are to replenish stock according by using predictive modelling.

198 Schwab <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/> (Date of use: 30 March 2019).

199 Schwab <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/> (Date of use: 30 March 2019).

200 Schwab <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/> (Date of use: 30 March 2019).

factors, vaccinations), absent or inadequate infrastructure and skilled human resources.²⁰¹

The impediments relating to access to reliable, inexpensive and efficient healthcare are myriad, and include shortfalls in workforce, infrastructure and funding. These barriers are prevalent in rural areas.²⁰² The average life prospects and health related conditions of rural populations are normally poorer than urban populations on account of abject poverty and limited access to competent healthcare professionals and facilities.²⁰³

South Africa spends over 8 per cent (R 480 billion) of its gross domestic product (GDP) on health-related expenditure. Notwithstanding this, people in the rural areas still cannot access proper healthcare services with health spending applied on maintaining existing services and infrastructure, without escalating service delivery to address unmet needs.²⁰⁴ Due to poor working conditions in rural areas, it becomes challenging to attract or retain trained healthcare professionals in these environments.²⁰⁵ To counter the scarcity of doctors, LMIC-settings tend to only offer minimal training programs for doctors or they permit nurses to carry out certain doctor-tasks which they are not qualified to do.²⁰⁶

The implementation of AI systems in rural areas could alleviate the inequity between urban and rural healthcare delivery.²⁰⁷ China's rural areas has already experienced

-
- 201 Deloitte "2019 Global health care outlook"
<https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Life-Sciences-Health-Care/gx-lshc-hc-outlook-2019.pdf> (Date of use: 18 May 2020).
- 202 Based on the World Bank's collection of development indicators, South Africa's rural population was reported at 33.65 per cent in 2018. See Trading Economics
<https://tradingeconomics.com/south-africa/rural-population-percent-of-total-population-wb-data.html> (Date of use; 18 May 2020).
- 203 Deloitte <https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Life-Sciences-Health-Care/gx-lshc-hc-outlook-2019.pdf> (Date of use: 18 May 2020).
- 204 Rural Health Advocacy Project
https://www.groundup.org.za/media/uploads/documents/Austerity%20report_FINAL.pdf (Date of use: 18 May 2020). The report indicated that even the number of those people seeking medical assistance at local clinics has decreased because of poor services.
- 205 The NDoH in South Africa regards health labour planning as a failing issue, in relation to data processing and management. See Valiani https://hasa.co.za/wp-content/uploads/2020/02/Public_Health_Care_Spending_in_South_Afr.pdf (Date of use: 18 May 2020).
- 206 Guo and Li 2018 *Health equity* 177.
- 207 The inequality result in excessive morbidity and mortality rates in South Africa and flags the need for addressing inequalities in the various healthcare settings. Valiani (Date of use: 18 May 2020).

the benefits from AI healthcare technology. For example, village healthcare settings in China²⁰⁸ are using a transportable “all-in-one diagnostic station” which is capable of running various examinations, including blood pressure tests, an electrocardiogram (ECG) and regular urine- and blood assessments.²⁰⁹

Similarly in South Africa, the company Phulukisa Healthcare Solutions, has introduced a mobile based clinic that is countering the adversities of the rural healthcare settings, due to the initiatives ability to provide better patient care at more reasonable costs, making healthcare reachable to rural area patients who often commute long distances to access medical services. Phulukisa’s cloud empowered mobile backpack allows specialists to visit inaccessible areas and conduct tests for lethal diseases, for example HIV, diabetes, and tuberculosis.

This technology can, in real time, upload health data onto an online analytical platform and produce a diagnosis allowing healthcare providers to remotely manage patients. Technology corporations in China are opting for AI-enabled “smart clinics” for remote and rural districts, which include AI-operated chat-bots, which are able to support remote medical consultations and advice with patients, and provide online training for healthcare staff in remote and rural regions.²¹⁰

In African societies, where openly discussing health issues relating to sexual activity is usually a taboo, a chat-robot called Sophie-Bot offers a platform to deal with reproductive health issues under anonymity and a user-friendly conversational interface, which service is available on mobile messaging applications, including Facebook Messenger and Twitter.²¹¹

208 WeDoctor, a mobile clinic that is part of Tencent the Chinese technology conglomerate, signed an agreement with the Chinese government to provide healthcare services, insurance, pharmaceuticals and healthcare training in villages by using its AI technology.

209 Hawkins A “How elderly, sickly farmers are quenching China's thirst for data” <https://www.wired.co.uk/article/china-ai-healthcare> (Date of use: 18 May 2020)

210 A popular internet search engine in China called giant Baidu developed a medical conversational Chatbot in 2015 named Melody that allows for diagnosing illnesses more efficiently and enables users to make contact with local doctors, schedule appointments and ask medical questions.

211 For more information on Sophie Bot, see Yatich B “SophieBot AI wants to replace your doctor” <https://www.afritechmedia.com/sophiebot-ai-wants-to-replace-your-doctor/> (Date of use: 24 June 2020).

Thus, AI technology, such as the abovementioned, does not only reduce costs, improve doctors' productivity, and the efficacy of healthcare delivery, but other healthcare staff such as nurses could also be educated on how to work with the technology in order to counter the shortage of doctors.

To enhance the delivery of healthcare services in remote and rural areas of LMICs, AI technologies for healthcare should be specifically developed for rural settings. It would require many forms of support, for instance suitable electrical power and internet infrastructure, ongoing education, oversight, commercial and operational backing, and effective public healthcare policies. An elementary frontline AI system for healthcare should be used for rural healthcare, such as township or rural clinics.

However, this can only be realised with the intervention and support of government. While public sector health services in South Africa had been officially desegregated in 1988, the health system remains fragmented leading to disparities in resource allocation between geographic areas and to poor use and distribution of available healthcare resources, particularly for rural communities. Following 1994, the government in South Africa has had a fundamental obligation to sufficiently resource healthcare in recognition of the right to healthcare, particularly for the most vulnerable and to justify any limitations in allocating additional resources towards fulfilling this obligation.²¹²

It is therefore not sufficient for the government to simply state that resources are limited and that it cannot allocate more resources to healthcare. The current health crisis in South Africa, leading to inevitable trade-offs between core rights and a resource constrained environment, provides an opportunity to stimulate investments in AI healthcare technology as part of a broader social stimulus that would support healthcare reform.

212 Section 27 of the Constitution in South Africa states that people are entitled "to have access to healthcare services, including reproductive healthcare [...]".

2.2. Challenges

The approaches that are applied for BD and AI are also being implemented to healthcare services and structures globally.²¹³ However, despite the promise that these new technologies hold, it also gives rise to global ethical-, social- and legal issues concerning fair access, privacy, responsible use, accountability, and prejudice. Although some governments, non-profit agencies, and academic institutions are focussing on the ethical issues associated to the implementation of AI systems, international guidelines specific to AI's application in healthcare are lacking.²¹⁴ This section considers the challenges related to AI and the ethico-legal challenges that ought to be addressed when introducing AI to healthcare.

2.2.1. Data privacy and security

The promise of healthcare is a multifaceted effort at both individual- and resident levels. The compilation and assessment of reliable data are crucial towards the enhancement and effectiveness of healthcare supply as it is a determinant on recognising what treatment should be applied for specific patients in certain clinical situations.²¹⁵ Obtaining data from private individuals is a necessity in respect of any BD application and is beset with ethico-legal concerns.

Data protection are essential concerning the application of AI solutions in healthcare, for fostering public trust as well as regulatory compliance when it comes to AI. Even in the best of instances, a violation of personal medical information will always be a concern with the operation of technologies as diverse as mobile devices and national data processing centres, aimed for the collection and administering of even greater volumes of information.

Data privacy issues are particularly important for AI as health records is generally collected and owned by governments, and susceptible to private companies

213 Salathé M, Wenzel M and Kishnamurthy R “Focus group on artificial intelligence for health” https://www.itu.int/en/ITU-T/focusgroups/ai4h/Documents/FG-AI4H_Whitepaper.pdf (Date of use: 13 June 2020).

214 Mariano B “Towards a global strategy on digital health” 2020 *WHO* 231.

215 London School for Economics and Political Science “Electronic health privacy and security in developing countries and humanitarian operations” <http://personal.lse.ac.uk/martinak/eHealth.pdf> (Date of use: 19 May 2020).

acquiring access to and leveraging personal data for commercial gain.²¹⁶ This problem is augmented when the information of people in susceptible communities are exploited.²¹⁷

Even rudimentary health data relating to sexual reproductive health, HIV or other infections, genetic diseases and probability of exposures for such disease, can be misappropriated or hacked and may result in stigmatisation, prejudice, and endangerment to personal security. In *NM v Smith and others*,²¹⁸ the South African Constitutional Court held that a person's HIV status must be afforded protection against undiscerning disclosure on account of the nature and adverse social implications the disease carries, including the probable judgement and discrimination that would ensue from disclosure of such status.

South Africa is marred by historical systematic discrimination and ongoing societal marginalisation and due consideration needs to be given to the manner in which health data is shared. The danger of negligent or deliberate breaches of information security and mismanagement of health data, may be more prevalent in LMIC settings with greater levels of corruption and lack of education,²¹⁹ which are being exposed to a rapid technological evolution.²²⁰

Laws safeguarding privacy and data tend to be underdeveloped with poor enforcement in many LMIC settings.²²¹ The Life Healthcare Group announced in

216 Various technology companies freely offer their AI products or services to their users as a marketing initiative allowing such companies that own these technologies to sell the users' sensitive health data. USAID <https://www.usaid.gov/cii/ai-in-global-health> (Date of use: 22 May 2020).

217 London School for Economics <http://personal.lse.ac.uk/martinak/eHealth.pdf> (Date of use: 19 May 2020).

218 2007 5 SA 250 (CC).

219 Corruption leads to governments to introduce policies in scenarios where government intervention desperately needed such as health and safety oversight and social protections. The World Bank Group "Helping countries combat corruption: The role of the World Bank" <http://www1.worldbank.org/publicsector/anticorrupt/corruptn/cor02.htm> (Date of use: 19 May 2020).

220 Thomson Reuters Foundation and the Trust Law Connect "Patient privacy in a mobile world. A framework to address privacy laws in mobile health" <https://www.trust.org/contentAsset/raw-data/03172beb-0f11-438e-94be-e02978de3036/file> (Date of use: 19 May 2020).

221 Notably the African countries of South Africa, Mauritius, Morocco and Tunisia, have already implemented data protection legislation. However, the implementation of a suite of privacy laws is not common among African countries.

June 2020 that, its South Africa business was impacted by a cyber-attack in respect of its information technology systems and sensitive patient data was compromised, raising criticism regarding the level of protection over its systems.²²² Patient privacy in many African countries are not always adhered to by healthcare staff, mainly due to the fact that African cultures tend to be communal by nature, and families may demand that healthcare staff disclose records of a patient's condition and diagnosis, failing which medical attention to the patient would be withheld due to lack of financial support from the family.²²³ Many countries have also not introduced laws that specifically safeguard health and medical data.²²⁴

Various legal approaches exist relating to permission to share personal information. For instance, the privacy of personal information is safeguarded under South African law by the POPIA based on the right to privacy in the Constitution, which is commensurate with existing data protection laws around the world such as GDPR. When referring to personal information it encompasses information pertaining to:

[A]n identifiable, living, natural person, and where it is applicable, an identifiable, existing, juristic person.²²⁵

This information encompasses an individual's health records, ethnicity, sex, health status and biometric data. Although the POPIA was not intended in the context of health data, it will undoubtedly influence the sharing of health information and bring about amendments relating to management and oversight of such data.²²⁶ Health data qualify as special personal information for the purposes of the POPIA and safeguards that such information is collected, processed and retained by public and private bodies.

222 Life Healthcare "Life Healthcare announces cyber incident"
<https://www.lifehealthcare.co.za/news-and-info-hub/latest-news/life-healthcare-announces-cyber-incident/> (Date of use: 20 June 2020).

223 Thomson Reuters Foundation and the Trust Law Connect
<https://www.trust.org/contentAsset/raw-data/03172beb-0f11-438e-94be-e02978de3036/file>
(Date of use: 19 May 2020).

224 World Economic Forum 2015 Survey Report
http://www3.weforum.org/docs/WEF_GAC15_Technological_Tipping_Points_report_2015.pdf (Date of use: 24 May 2020).

225 Section 1 of POPI.

226 Staunton C *et al* "Protection of Personal Information Act 2013 and data protection for health research in South Africa" 2020 *Oxford University press* 1-10.

Whilst the POPIA makes provision for research, sector-specific guidelines dealing with the management and control of data distribution for research in healthcare is needed.²²⁷ The POPIA has tasked the information regulator and organisations to establish the codes of conduct²²⁸ for different sectors in which the responsible parties are operating, however nothing has been proposed for the healthcare sector as yet. Suitable measures in terms of POPI are to be implemented for safeguarding the interests of those who share data on unsolicited electronic communications and automated decisions. This ought to address the particular nuances that is essential in the management of health information.

Section 14 of the Constitution in South Africa similarly protects privacy rights, which by implication require patients to provide consent to release their medical information, as well as restricting the sharing of healthcare data in the absence of patient consent unless it is imposed by law or sanctioned by a court order or if reasonable and justified in society's interest.

Similarly, the National Health Act 61 of 2003 (NHA) renders it an offence to release patients' personal data in the absence of their consent, unless certain circumstances apply.²²⁹ The HPCSA's "confidentiality: protecting and providing information" contains ethical standards on ensuring confidentiality and disclosure of confidential information in different scenarios.²³⁰

Under the ethical guidelines, limited disclosures are permitted notwithstanding the general protection of the right to privacy; for instance, doctors may, release the HIV status of patients to their spouses or partners who may be exposed to any risk. Unfortunately, personal health information is often inadvertently shared or disclosed without the patients' consent. Such a breach often remains undetected such as through peer-to-peer computer or network file distribution applications that allows

227 Staunton et al 2020 Oxford University press 18.

228 Chapter 7, section 60(1) of the POPIA.

229 Sections 7, 14, 15 and 16 of the National Health Act 61 of 2003 (hereinafter referred to as the NHA).

230 HPCSA "Guidelines for good practice in the health care: Ethical guidelines for the practice of telemedicine – booklet 10"
https://www.hpcsa.co.za/Uploads/Professional_Practice/Ethics_Booklet.pdf (Date of use: 19 May 2020).

for various peer-to-peer users to browse and download data from different networks or via health risk assessments carried out by public and private organisations, employers, health insurers, public health departments, to name but a few.

The US Department of Health and Human Services (HHS) released the “Privacy Rule” standards with regulations prescribed by the Health Insurance Portability and Accountability Act of 1996 (HIPAA). The standards safeguard personal health data and cover who is protected, what medical and health data is safeguarded, and determines the manner in which health data that is protected should be processed and divulged.²³¹

HIPAA’s privacy rule governs “covered entities”, which are authorised to process or divulge protected health data relating to patient care, processing of fees and payments, and healthcare procedures, and may release such information to other “covered entities” for the same reasons.²³² In this instance, issues may arise when smart machines access electronic medical records. This is due to the fact that there is ambiguity as to whether the smart machine should be considered a “covered entity” under circumstances where although a smart machine’s access concerning a specific patient’s healthcare information may be authorised for medical treatment, access to other patients’ medical records for the purpose of predicting a specific patient’s outcome to treatment options may not.²³³

In terms of covered entities, it also means that HIPAA’s privacy rules does not direct that healthcare providers or agencies require the patient’s consent to release their medical or other personal information to third party agencies, to whom they have outsourced services, such as payment processing, administration and legal services. These outsourced agencies gain access to patient information as part of rendering their services.

231 The Privacy Rule can be found at 45 CFR Part 160 and sub-parts A and E of Part 164 available at <https://www.hhs.gov/hipaa/for-professionals/privacy/index.html> (Date of use: 19 May 2020).

232 See 45 CFR Part 160.103.

233 Khan *Medicine* 71.

Whilst companies who outsource the services do conclude written agreements with outsourced agencies to bind them to keep the patient information they receive classified, if there is an infringement or abuse of such information, the patient has limited recourse to legal action. HIPPA provides inadequate legal enforcement and traditional consent methods must be substituted or complemented by more accountable, open, and sensible methods for data governance, and to reinforce capabilities in data oversight, evaluation and policy use.²³⁴ Therefore, policymakers must strike a balance between data access and privacy, including the harmonising of the right to privacy with the constitutional right of access to healthcare. However, this may not be an easy task.

Private firms and public organisations are also involved in harvesting and exploiting personal patient data, to use it for purposes other than their specific offerings of healthcare services into complementary “health services” (such as health insurance and lifestyle programs). Under the South African Medical Schemes Act 131 of 1998, medical schemes are registered as non-profit trusts.²³⁵ However, in South Africa such medical schemes make a considerable profit and are even expanding their enterprise beyond their initial offerings to include banking, life insurance, investments, and car and home insurance.

An example of data exploitation is where London’s Royal Free Hospital did not secure suitable consent to use the data concerning 1.6 million patients in the creation of an AI program and transferred its personal patient data to one of Google’s subsidiary, Deepmind.²³⁶ South Africa has also not been spared in similar instances of data abuse. In South Africa, one of the biggest medical insurance company, Discovery Health, partnered with an American firm Human Longevity Inc.,

234 The WHO has observed that, “data security is a particularly important issue to address within the area of policy [...] Policy-makers and programme managers need to be made aware of security issues in the mHealth domain so appropriate policies and strategies can be developed and implemented.” MHealth refers to “medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants (PDAs), and other wireless devices”. See WHO “mHealth: New horizons for health through mobile technologies”
http://www.who.int/goe/publications/goe_mhealth_web.pdf (Date of use: 19 May 2020).

235 Medical Schemes Act 131 of 1998, sections 20, 24 and 26.

236 The Guardian “Royal free breached UK data law in 1.6m patient deal with Google’s DeepMind” <https://www.theguardian.com/technology/2017/jul/03/google-deepmind-16m-patient-royal-free-deal-data-protection-act> (Date of use: 13 June 2020).

to offer genetic testing at a cost to its members in the UK and South Africa, in exchange for giving their members data to the American firm.

As part of the deal the highly prized personal and health data, would be destined to be commercialised in other ways with numerous collaborators around the world. The firm needs to harvest a substantial amount of DNA samples in order to generate a database, which is substantial enough to commercially offer to researchers (like those based in pharmaceutical corporations).

There have been concerns raised on the exploitation of the data, such as that of Professor Michael Pepper, director at the Institute for Cellular and Molecular Medicine, who questioned the ethics of a US company positioning itself to exploit genomic data from South Africa and elsewhere, and he indicated that, “it will be Discovery’s responsibility to ensure that the interests of the country and its people are protected.”²³⁷ In 2019 France’s information watchdog, the Commission nationale de l’informatique et des libertés (CNIL) levied a EUR€50 million penalty on Google for neglecting to provide satisfactory information relating to its data consent policies.²³⁸ Its ruling relied on principles regarding clarity and accessibility enshrined in the EU’s strict GDPR.²³⁹

The obligation to ensure that explicit consent is obtained from patients to use their personal information in AI development is consequently a key factor to attain trust in AI’s implementation in the healthcare industry. Under the POPIA, consent of the data subject is one of the lawful basis on which data can be processed under POPIA.²⁴⁰ The POPIA does contain exemptions²⁴¹ to the general prohibition²⁴² of handling personal health data, which exemptions include:

-
- 237 Cyberstoep “Discovery breaks new ground with genetic testing”
<http://www.cyberstoep.co.za/discovery-breaks-new-ground-with-genetic-testing/> (Date of use: 11 June 2020).
- 238 Dispatch Live “Google submits to €50m fine by French authority”
<https://www.dispatchlive.co.za/news/2020-06-20-google-submits-to-50m-fine-by-french-authority/> (Date of use: 21 June 2020).
- 239 Articles 5-11.
- 240 Section 27(2)(a) of the POPIA.
- 241 Section 27 the POPIA.
- 242 Section 26 the POPIA.

[T]he processing by medical professionals, healthcare institutions or facilities or social services, if such processing is necessary for the proper treatment and care of the data subject, or for the administration of the institution or professional practice [concerned].

However, it is important to implement a clear governance framework which protects personal information and provides assurance that, such information is only disclosed if necessary and premised on requirements that serve to protect privacy as fundamental to public trust in accessing health data. The UK's implementation of a national "data opt-out" platform has allowed patients to control the access to their data. The platform enables patients to opt-out of their personal health information being accessed and applied for research and development.²⁴³ Similar services should be considered for South Africa, as more consumers are starting to realise the worth of their personal information.

Section 13 of the POPIA prescribes that personal information is to be collected and used for an intended purpose. Therefore, patients should have the option to revoke or withhold consent for any intended purpose. Privacy issues must also be resolved in more pioneering ways. Limiting access is not the only solution, and in the case of healthcare, sounder data-encryption-technology²⁴⁴ based alternatives can be considered. Cloud encryption services, such as advanced encryption standard (AES), may be used to encrypt electronic data before it is transferred to the cloud for storage. A virtual private network (VPN) that will allow sharing of data over an encrypted network without anyone intercepting could also be implemented. Other options are transport layer security (TLS) for emails which is the simplest of all the encryption options that protects an email in transit between two email servers; or

243 NHS Digital "National data opt-out" <https://digital.nhs.uk/services/national-data-opt-out> (Date of use: 13 June 2020).

244 Data encrypting involves the conversion of the original form of the information into encoded text, rendering the information unreadable unless someone has the necessary key or password to decrypt it.

the information rights management (IRM) platform enforcing the persistent protection of both files and email.²⁴⁵

Countries' national statistical services, such as Stats SA in South Africa, could also play an effective role, where their authority should be broadened to operate as custodians of personal data collected by third parties.

2.2.2. Data inequity

Governance is a major barrier when it comes to data, particularly concerning its ownership, the manner in which it is stored, how it is accessed, and who is entitled to access it, which are all important questions when it comes to AI. When it comes to data, the playing field is unequal and hinders data sharing worldwide. Scientists from high income countries (HICs) often undertake that they will study and publish research papers for universal public good by collecting data from LMICs that do not have analytical expertise or resources. International data sharing is, therefore, often achieved through a forced marriage and often LMICs are not the beneficiaries thereof.²⁴⁶

For global data sharing to be successful it would have to begin with endeavours at state level, or through multi-state research collaborations. The standards and interests of the intended societies should be converted into best practices, which offset the advantages and consequences of data sharing and usage.²⁴⁷

The COVID-19 pandemic has brought to fore the realisation that, data "ownership" strongly benefits the private sector as giant pharma companies are trying to win the

245 In the GDPR, article 32 recommends the use of encryption for personal data. GDPR excludes the reporting of a data breach if it involves data that was encrypted. Also see Business Tech "How encryption can help protect your sensitive data" <https://businesstech.co.za/news/industry-news/406191/how-encryption-can-help-protect-your-sensitive-data/> (Date of use: 16 June 2020).

246 Avian flu virus samples were shared by developing countries through the WHO which contributed to the production of avian flu vaccine at a significant price of US\$ 10–20 per dose being unaffordable in LMIC countries. This leaves only the wealthy countries' populations fully protected, without any ethical obligations to impoverished countries that shared their specimens.

247 Tangcharoensathien V, Boonperm J and Jongudomsuk P "Sharing health data: developing country perspectives" <https://www.who.int/bulletin/volumes/88/6/10-079129.pdf> (Date of use: 19 May 2020).

race for a vaccine. The most valuable data belongs to private sector business, such as: financial institutions; mobile operators; and technology titans like Facebook, Amazon, and Google.²⁴⁸

Data capturing and analysing is worth millions in profit and the few companies that acquire and mine such data acquire immense power to understand and predict events around the world.²⁴⁹ When various data sets are pooled, new information and value is generated. With the increasing use of AI, access to and sharing of data will significantly increase in value. However, if valuable data is only in the hands of a few private companies, then only a few can contribute to producing new ideas and knowledge. This decelerates innovation and progress. Individuals and governments often face barriers to data access, which may be exacerbated by reluctance to share data, including within organisations and across sectors.²⁵⁰

To address these challenges of significant amounts of data being at the disposal of the private sector, some private sector actors have already established some open source initiatives. Since 2010, British pharmaceutical company GlaxoSmithKline (GSK) has publicly made available a database to external researchers, scientists, and other companies, containing data for possible drugs against malaria and tuberculosis.²⁵¹ In the area of healthcare research, open science has made it possible to foster an environment where research can flourish.

By leveraging on technologies such as open source, research data and material is made freely available in the digital environment without any financial-, legal- or technical constraints. Open science, as a practice of science, also allows for others outside the research team to join forces and collaborate because of research data,

-
- 248 Google Alphabet, is leading the market for driverless vehicles, not because it has any expertise in building cars but because it of its access to greater geographical data than any other company. Similarly, some mobile network operators are making huge profits on banking and insurance because information they possess on their mobile phone consumers.
- 249 Nielsen M "Who owns big data?" <https://www.technologyreview.com/2015/01/05/169719/who-owns-big-data/> (Date of use: 20 May 2020).
- 250 OECD "Data governance: Enhancing access to and sharing of data" <https://www.oecd.org/sti/ieconomy/enhanced-data-access.htm> (Date of use: 20 May 2020).
- 251 GSK "Open innovation" <https://uk.gsk.com/en-gb/research/sharing-our-research/open-innovation/> (Date of use: 12 June 2020).

and related developments, being readily accessible (electronically) in conditions which allow for reprocessing, redeployment, and replication of the research.²⁵²

An early open science initiative was undertaken by the Academy Health's Electronic Data Methods (EDM) Forum, known as the Collaborative Informatics Environment for Learning on Health Outcomes (CIELO), in which the main objective in establishing CIELO was to establish end-user requirements for health research, in a field where data attribution and security were considered of importance.²⁵³ Since then, many open science platforms have been developed, such as the COVID-19 Technology Access Pool (C-TAP), an enterprise that seeks to increase access for all to data, vaccinations, tests and treatments to fight COVID-19.²⁵⁴

The Health Committee and Committee on Digital Economy Policy appointed by the OECD proposed the "recommendation on health data governance", introduced in 2016 by the OECD Council. This recommendation acknowledges that,

many OECD Members lack a coordinated public policy framework to guide health data use and sharing practices, so as to protect privacy, enable efficiencies, promote quality and foster innovative research.²⁵⁵

There remains a need for data governance frameworks to be coherent across economic sectors, public sector organisations, society, and countries.²⁵⁶ This requires governments to adopt extensive policies that do not merely seek to protect private information to data, but also to eradicate exclusive use rights to public sector

252 Payne P *et al* "Enabling open science for health research: Collaborative informatics environment for learning on health outcomes (CIELO)" 2017 *Journal of Medical Internet Research* 2.

253 Payne et al 2017 *Journal of Medical Internet Research* 2.

254 WHO "International community rallies to support open research and science to fight COVID-19" <https://www.who.int/news-room/detail/29-05-2020-international-community-rallies-to-support-open-research-and-science-to-fight-covid-19> (Date of use: 12 June 2020).

255 The OECD (2016) Recommendation of the Council on Health Data Governance recommends that, member states "should establish and implement a national health data governance framework to encourage the availability and use of personal health data to serve health-related public interest purposes while also promoting the protection of privacy, personal health data and data security", available at OECD "Recommendation of the council on health data governance" <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0433> (Date of use: 20 May 2020).

256 OECD <https://www.oecd.org/sti/ieconomy/enhanced-data-access.htm> (Date of use: 20 May 2020).

data, forcing private companies to make their data accessible to third parties. This can be achieved by issuing conditional licences analogous to patents of invention.

A powerful public data infrastructure could be created, which could be accessed by anyone on a global scale and which can be publicly funded and managed by non-profit organisations or NGO's, a government, or a network of private contributors.

2.2.3. Bias

AI technology can accelerate the advancement of global healthcare by facilitating equal access, and improved quality of care for all population classes. Matters relating to bias and quality impact directly on fairness, in relation to the assessment of possible decision patterns and outcomes. Formalising decisions made by a person and understanding how specific choices factor into those decisions is as crucial as the analysis involved in understanding those decisions.²⁵⁷ There are many illustrations where bias can unintentionally impact on the outputs generated by the application of ML methods, due to inherent bias embedded into the platform upon which the AI technologies are produced, data sets that inherently contain bias and are applied to the AI models to train them, or biases present when the AI technologies are applied in “real-world” scenarios.²⁵⁸

When designing AI systems with healthcare in mind, it is important that already existing human bias in healthcare settings is not programmed into the system.²⁵⁹ If the data is defective or biased from the outset, the outcome will also be compromised. There are apprehensions about the possibility of AI systems to cause bias in unidentified ways, which are inconsistent with legally protected attributes, including: sex; culture; disability; and age.²⁶⁰

257 Hager *et al* <https://cra.org/ccc/wp-content/uploads/sites/2/2016/04/AI-for-Social-Good-Workshop-Report.pdf> (Date of use: 14 May 2020).

258 Ho CLW, Ali J and Caals K “Ensuring trustworthy use of artificial intelligence and big dataanalytics in health insurance” 2020 *WHO* 265.

259 Pley C, Dhatt R and Keeling A “Gender bias in Health AI – prejudicing health outcomes (or getting it right!)” <https://www.womeningh.org/single-post/2019/09/11/Gender-bias-in-Health-AI---prejudicing-health-outcomes-or-getting-it-right> (Date of use: 13 June 2020).

260 Nuffield Council on Bioethics <http://nuffieldbioethics.org/wp-content/uploads/Artificial-Intelligence-AI-in-healthcare-and-research.pdf> (Date of use: 31 March 2019).

A major difficulty that arises in adapting AI models from health data bases is that, the data can be manipulated by many characteristics, including some not associated to the subject of interest, and influenced by the composition of the health ecosystem.²⁶¹ For instance, applying ML to human language, or text data for public health purposes, could present human based biases, including those that are discriminatory based on race or gender.²⁶²

In cases of cardiac arrests, symptoms in women manifest in a different way than in men, but present data sets on research tend to concentrate on symptoms suffered by men.²⁶³ Consequently, current data sets concerning heart attacks, is not necessarily as precise for females, and this results in misdiagnosis and mistreatment.²⁶⁴ Likewise, health related concerns could alter considerably across different cultures and countries and has been noted in hereditary illnesses or susceptibilities (e.g. diabetes tends to be more predominant in African and Indian populations than white populations in the South Africa,²⁶⁵ in urban versus rural settings, in the Africa region versus North American region),²⁶⁶ and inferior healthcare settings (e.g. the AI system envisages a sophisticated operating room as opposed to an undeveloped medical facility found in rural settings).²⁶⁷

261 Hager *et al* <https://cra.org/ccc/wp-content/uploads/sites/2/2016/04/AI-for-Social-Good-Workshop-Report.pdf> (Date of use: 14 May 2020).

262 Smith MJ *et al* "Four equity considerations for the use of artificial intelligence in public health" 2020 *WHO* 291.

263 Pley, Dhatt and Keeling <https://www.womeningh.org/single-post/2019/09/11/Gender-bias-in-Health-AI---prejudicing-health-outcomes-or-getting-it-right> (Date of use: 13 June 2020).

264 Unlike the typical chest pain symptoms experienced by men who suffer heart attacks, 40 per cent of women do not encounter any chest pain and instead experience symptoms such as nausea, vomiting, abdominal pain, dizziness or breathing problems. See Pley, Dhatt and Keeling <https://www.womeningh.org/single-post/2019/09/11/Gender-bias-in-Health-AI---prejudicing-health-outcomes-or-getting-it-right> (Date of use: 13 June 2020).

265 News 24 "Prevalence of diabetes in South Africa" <https://www.health24.com/Medical/Diabetes/About-diabetes/Diabetes-tsunami-hits-South-Africa-20130210> (Date of use: 13 June 2020).

266 International Diabetes Federation "IDF Diabetes Atlas" <https://www.diabetesatlas.org/en/sections/demographic-and-geographic-outline.html> (Date of use: 13 June 2020).

267 Sallstrom L, Morris O and Mehta H "Artificial intelligence in Africa's healthcare: Ethical considerations" https://www.orfonline.org/wp-content/uploads/2019/09/ORF_Issue_Brief_312_AI-Health-Africa.pdf (Date of use: 19 May 2020).

ML systems in healthcare may also be susceptible to algorithmic prejudice, perhaps forecasting greater possibility of illness or disease based on gender or race which may not necessarily be contributory factors.²⁶⁸

If bias is not factored into the application of AI, it can result in models that cause harm, as the healthcare provider practice patterns shift. This could have a negative bearing on individuals with erratic medical issues or those who are overlooked in clinical tests and research, such as for minority African people in the US.²⁶⁹ If health insurance coverage is required, it may also impact on differences in premiums as predictive AI algorithms detect characteristics that are associated with loss to the insurer, some of which can also be linked to race.

For example, the US healthcare system uses algorithms to steer its health assessments. In 2019, scientists discovered racial bias in one commonly used algorithm, which forecasts which patients will benefit from additional healthcare. The algorithm incorrectly determines that unwell African - American patients are in better health than similarly unwell Caucasian patients, as this algorithm takes in account health expenditure rather than sickness as a yardstick for health needs. A lesser amount of money is expended on black patients who are in need of the same level of healthcare but are allocated the same level of risk than Caucasian patients.²⁷⁰ It was determined that the racial bias of the algorithm diminishes by half the amount of black patients who are eligible for additional healthcare.²⁷¹

In terms of GDPR, persons have the “right not to be subject to a decision based solely on automated processing, including profiling which significantly affects him or her.”²⁷² As AI is only able to produce that, which it has been thought and learnt, human intervention is essential to safeguard against AI knowledge that is biased and not holistic and that algorithms are reformulated whenever predictive inaccuracy

268 Davenport and Kalakota 2019 *Future healthcare journal* 97.

269 Ho, Ali and Caals 2020 *WHO* 264.

270 Obermeyer Z *et al* “Dissecting racial bias in an algorithm used to manage the health of populations” <https://science.sciencemag.org/content/366/6464/447.full> (Date of use: 13 June 2020).

271 Obermeyer Z *et al* <https://science.sciencemag.org/content/366/6464/447.full> (Date of use: 13 June 2020).

272 Article 22.

or bias is detected. This can be achieved by carefully identifying, selecting, and changing the data that is fed to the algorithm and specifically, the labels that are assigned to data sets that predict factors such as consumer behaviour.²⁷³

Medical data also needs processing platforms that are “healthcare process aware”, to guarantee the compilation of data that is not biased or inaccurate.²⁷⁴ AI tools are dependent on voluminous historical health data areas, in order to teach their algorithms to render precise outcomes suitable to the geography and people involved. However, this sort of wider health data are generally missing in LMICs, which generally hinders the accuracy of the AI tools in the framework of LMICs, but also generates bias within such tools.

As AI algorithms naturally echo the predisposition of data sets that they are trained on, AI systems will reflect a bias in favour of the first-world countries where these systems are in most cases designed.²⁷⁵ There is, therefore, a need to acclimatise these biases in the implementation of AI tools, as to environments distinct from where their training data sets occurs and it is more imperative that the design is correct from the onset. There is already an advent of impartial designed robots (such the Tengai Robot) created to conduct automated interviews on behalf of human employment recruiters, ensuring that it is fair and impartial, and delivering objective candidate evaluations.²⁷⁶

The robot was designed in partnership between recruitment and staffing agency TNG, who provided the diversity and inclusion software program and Furhat Robotics, which provided the technology. If AI is to succeed in the healthcare industry in Africa, African scientists should participate in its design and advancement with African based data that informs such development.²⁷⁷

273 Obermeyer Z *et al* <https://science.sciencemag.org/content/366/6464/447.full> (Date of use: 13 June 2020).

274 Hager *et al* <https://cra.org/ccc/wp-content/uploads/sites/2/2016/04/AI-for-Social-Good-Workshop-Report.pdf> (Date of use: 14 May 2020).

275 USAID <https://www.usaid.gov/cii/ai-in-global-health> (Date of use: 22 May 2020).

276 Tengai “The interview robot – how it works” <https://www.tengai-unbiased.com/how-robot-job-interview/> (Date of use: 27 June 2020).

277 Sallstrom, Morris and Mehta https://www.orfonline.org/wp-content/uploads/2019/09/ORF_Issue_Brief_312_AI-Health-Africa.pdf (Date of use: 19 May 2020).

2.2.4. Social isolation and trust in AI

Humans are communal in nature and it is our personal association with others that allows us to endure and flourish. As we age, many of us do not have access to friends or family, exposing us to social solitude and related physical- and mental conditions, for example mental deterioration, anxiety, and heart disorders.²⁷⁸ Under these circumstances, visiting the doctor, especially for people who are isolated, is a social lifeline as much as it becomes a medical requirement.

More recently, COVID-19 has been a major force for social isolation across the world, which has been an abrupt change for many. In African countries, the effects may be more so because of the notion of “Ubuntu”, a social and political African philosophy.²⁷⁹ Ubuntu is premised on the concept that, “a person is a person through other persons,” founded on principles of humaneness, consideration, respect, empathy, as well as “ensuring a happy and qualitative human community life in a spirit of family.”²⁸⁰ The Ubuntu notion of trust in human relationships was echoed by South African liberation struggle icon Steve Biko, who wrote before his untimely death:

[W]e reject the power-based society of the Westerner that seems to be ever concerned with perfecting their technological know-how while losing out on their spiritual dimension. We believe that in the long run the special contribution to the world by Africa will be in this field of human relationship. The great powers of the world may have done wonders in giving the world an industrial and military look, but the great gift still has to come from Africa – giving the world a more human face.²⁸¹

AI challenges the perception of trust and benevolence in healthcare. The lack of human interaction and increased social disconnection are likely consequences if AI assistive technologies, such as robots, are intended to substitute healthcare workers in their interaction with patients.²⁸² However, given the challenges of under-

278 Australian Institute of Health and Welfare “Social isolation and loneliness” <https://www.aihw.gov.au/reports/australias-welfare/social-isolation-and-loneliness> (Date of use: 20 May 2020).

279 *S v Makwanyane* 1995 (6) BCLR 665 (CC).

280 Thomson Reuters Foundation and the Trust Law Connect <https://www.trust.org/contentAsset/raw-data/03172beb-0f11-438e-94be-e02978de3036/file> (Date of use: 19 May 2020).

281 Avis P, Pauw A and Van Der Spuy I *Psychological perspectives, an introductory workbook* 1st ed (Pearson Education South Africa 1999) 198.

282 Nuffield Council on Bioethics <http://nuffieldbioethics.org/wp-content/uploads/Artificial-Intelligence-AI-in-healthcare-and-research.pdf> (Date of use: 31 March 2019).

resourced healthcare settings, medical consultations may not be an ideal experience of benevolence or altruism.

The patient experience is often one that feels rushed through an impersonal appointment with a healthcare provider, where the patient is merely reduced to a number rather than a name. The benefit of AI in healthcare, and its potential for public good, far outweighs the notion that it will replace healthcare professionals entirely and requires the technology to be trusted.

To avoid a scenario where AI may be subordinated to communal interests, the technology must be designed with due consideration and compromise. Instead of AI being seen as a substitution for doctors, AI offers interventions aimed at solving the crises of overwhelmed care ecosystems, and offers the promise of teaching doctors to supervise and use the technology as a complementary tool, so as to augment their own analytical abilities and skills.²⁸³ To this end, healthcare chatbots can be used to socially interact with elderly or sick people as a therapeutic measure to reduce anxiety or loneliness.

With increasing advancements in AI assisted technology, one could also envision a future where ML could simulate not only cognitive human functions, but also social norms and other stereotypes, in order to standardise human-robot interaction for greater acceptability. Research in AI indicates that, the creation of benevolent machines is possible that accommodates value-multiplicity, recognising that patients with different values would have different concerns and needs when it comes to their healthcare, and which would also serve to alleviate the burden of the substantial emotional work required of healthcare professions.²⁸⁴

2.2.5. Effects on healthcare providers

Many healthcare professionals are concerned that AI technology could replace some of them and expose them to risk of skill erosion in certain areas, such as diagnostic capability and critical decision making, and consequently, it may reduce

283 Guo and Li 2018 *Health equity* 180.

284 Kerasidou M “Artificial intelligence and the ongoing need for empathy, compassion and trust in healthcare” 2020 *WHO* 247.

the healthcare labour market or lead to reduced earnings for healthcare workers. Whilst the introduction of AI into the healthcare sector will undeniably impact on the jobs of healthcare providers, it does not suggest that they are fated for redundancy. This is because the current AI technology in healthcare is only intended to serve as a decision- making tool for doctors and does not substitute their work.²⁸⁵

It seems likely that the healthcare jobs that are considered to become automated relate to administering of digital data, radiology, and pathology, as opposed to jobs that need personal patient contact.²⁸⁶ In rural areas, due to a shortage of doctors, it is unlikely that their jobs are threatened, or that their compensation will be reduced in the near future.²⁸⁷ It should also be borne in mind that, the duties of a healthcare provider include diagnosis consultations with patients, analysing patients' values and preferences, quality review and assurance, education and training, policy-making, intervention treatments and protocols, which thus far, cannot be executed by computer software alone.

It is increasingly apparent that AI will not take over from healthcare workers on a significant level but, will rather enhance the ability to treat patients by drawing on distinctively human skills like empathy and persuasion. Ultimately, those healthcare providers who may decline the opportunity to work alongside AI, may risk losing their jobs over time. Furthermore, research indicates that the collaboration between humans and intelligent machines yield far better outcomes than one or the other alone.²⁸⁸

285 Dr Curtis P Langlotz, a professor of radiology based at the University Medical Center, stated that, “[r]adiologists are being trained to recognize AI’s shortcomings and capitalize on its strengths [...]. Over the next 5 to 10 years, the most successful radiologists and pathologists will be those who are well-equipped and eager to participate in data management and integrated diagnoses.” He also added that, the work of radiologists goes beyond reading and interpreting images in contrast to radiology AI systems that can only execute single tasks. See Association of American Medical Colleges “Will artificial intelligence replace doctors?” <https://www.aamc.org/news-insights/will-artificial-intelligence-replace-doctors> (Date of use: 20 May 2020).

286 Davenport and Kalakota 2019 *Future healthcare journal* 97.

287 Guo and Li 2018 *Health equity* 180.

288 Patel BN *et al* “Human– machine partnership with artificial intelligence for chest radiograph Diagnosis” <https://www.nature.com/articles/s41746-019-0189-7> (Date of use: 16 June 2020).

A further concern is that healthcare professionals may become complacent, should they grow to become dependent on the technology.²⁸⁹ This may be detrimental if the technology is defective and fails, and healthcare workers cannot recognise or detect errors, nor execute essential functions without computer assistance, leading to job losses of less skilled staff.

On the other side, AI has the ability to promote and enhance required skills and earnings for undertrained healthcare workers.²⁹⁰ Skill is now recognised as a different form of investment, as a knowledge centred market transforms, and innovates individuals in the direction of higher-earning jobs that oversee the automated productivity created by AI modernisation.²⁹¹

2.2.6. Reliability and safety

An indispensable requirement for the pervasive deployment of AI applications is that it must be safe and reliable for humans. AI's slow rate of development has mainly been impacted by diminished trust because of doubts regarding its safety. The most obvious concern about AI systems is that it may commit errors, and that patients may be harmed or other medical problems may ensue. For example, should an AI system endorse incorrect medication or protocols, or if it overlooks a lump on a radiological image scan, the patient could suffer long term illness or death.²⁹² It may also not be possible that doctors will actually be able to assess the reliability of information derived from AI. This inability stems from the concept that an algorithm that cannot explain the route to its decision is referred to as a "black box."²⁹³

For example, MRI brain scans comprise of a neural network that, when given input data and trained on a substantial amount of data set, is able to locate and evaluate

289 Nuffield Council on Bioethics <http://nuffieldbioethics.org/wp-content/uploads/Artificial-Intelligence-AI-in-healthcare-and-research.pdf> (Date of use: 31 March 2019).

290 Guo and Li 2018 *Health equity* 180.

291 World Economic Forum 2015 Survey Report http://www3.weforum.org/docs/WEF_GAC15_Technological_Tipping_Points_report_2015.pdf (Date of use: 24 May 2020).

292 IBM Watson's developed its first AI application in 2013 for cancer treatment protocols ensuing in its partnership with several hospitals. Hospitals complained that the application rendered unsafe findings on cancer treatment protocols that could lead to severe or fatal consequences for patients.

293 Sullivan and Schweikart https://journalofethics.ama-assn.org/sites/journalofethics.ama-assn.org/files/2019-01/hlaw1-1902_1.pdf (Date of use: 4 June 2019).

a complex underlying pattern, and generate a diagnosis, but it cannot explain the rationale that generated its decision. Even if it were possible for a technically skilled doctor to inspect the process, many AI algorithms are protected proprietary information that cannot be accessed. Given the abstruse characteristic of AI as a “black box”, its implementation gives rise to complex legal issues.

This, however, provides a platform to contemplate possible consequences of AI for healthcare providers and product manufacturers, especially if they are not able to account for its outputs, and consider new solutions for addressing liability issues.²⁹⁴ Ironically, many injuries and fatalities arise on account of medical mistakes, experienced in healthcare settings today, even in the absence of AI.²⁹⁵ In HICs, academics have considered the quality of services in healthcare and have determined that despite the fact that it is human to make a mistake, most errors arising in the delivery of healthcare services may possibly be averted.²⁹⁶ This situation will further be complicated as doctors come to rely more on AI and it becomes more difficult to challenge an algorithm’s result.

As AI becomes even more ubiquitous in our everyday lives, traditional security perimeters will disappear, and prevailing standards of security should be re-established. New communication and operating structures will be required to institutionalise more effective security platforms and to infuse trust for individuals. Despite AI still facing various questions regarding trust and safety, recognising and understanding the risks will eventually enable stakeholders to align their interests to benefit from the technology the most.²⁹⁷

294 Sullivan and Schweikart https://journalofethics.ama-assn.org/sites/journalofethics.ama-assn.org/files/2019-01/hlaw1-1902_1.pdf (Date of use: 4 June 2019).

295 Worldwide primary healthcare experiences medical errors on a scale of between 5 and 80 times per 100 000 consultations with an average of 5% to 50% of all medical errors being attributable to administrative mistakes. WHO “Patient safety” <https://www.who.int/news-room/facts-in-pictures/detail/patient-safety> (Date of use: 20 May 2020).

296 Wang Z *et al* “Records of medical malpractice litigation: a potential indicator of health-care quality in China” 2017 *WHO* 430.

297 This sentiment was echoed by Nigel Duffy, EY’s Global AI Innovation Leader available at World Economic Forum “AI has started a financial revolution – here’s how” <https://www.weforum.org/agenda/2020/02/how-ai-is-shaping-financial-services/> (Date of use: 15 June 2020).

2.2.7. Transparency and accountability

In modest terms, trying to explain the outputs generated from complicated AI models continues to be a major obstacle to attaining consumer- and legal recognition. Referring to existing technology, Dr Daniel Bobrow of Xerox has submitted that,

[i]n any particular intellectual area, we can probably formalise the knowledge sufficiently so that the computer can do as well as or better than most people can do in that area. It turns out that consistency of judgment is at least as important as the knowledge.²⁹⁸

This is significant as AI algorithms consist of insurmountable parameters and even though these parameters are augmented by automatic computations referred to as the “hidden layer”, it is almost impossible to establish the manner in which these parameters operate and connect with one other.²⁹⁹ This inability to establish the way in which the parameters operate is termed as the “black box”, because it can be challenging to establish the core reasoning behind the outcomes generated by AI.³⁰⁰

This leaves it open to be employed for malevolent objectives. For instance, like the Russians in the 2016 US presidential election,³⁰¹ AI can be used autonomously in order to exploit vulnerabilities in systems, by improving target selection and prioritisation, evading detection, and creatively respond to changes in the target’s behaviour.³⁰²

According to Andrew Tutt,³⁰³ ML algorithms pose three main challenges when they tend to harm people: “(1) algorithmic responsibility that is difficult to quantify; (2) algorithmic responsibility that is difficult to trace; and (3) human responsibility that difficult to assign.” ML systems can be especially abstruse because of their innate ability to constantly alter parameters and rules whilst they are learning making it

298 Willick 1983 *The AI Magazine* 7.

299 Matsuzaki 2018 *California Western law review* 267.

300 Mckinsey Global Institute <https://www.mckinsey.com/~media/mckinsey/featured%20insights/artificial%20intelligence/applying%20artificial%20intelligence%20for%20social%20good/mgi-applying-ai-for-social-good-discussion-paper-dec-2018.ashx> (Date of use: 16 May 2020).

301 The Russian government unlawfully interfered in the US presidential election of 2016 by creating thousands of social media accounts through autonomous AI that falsely appeared to be Americans in support of President Trump.

302 Brundage *et al* “The malicious use of artificial intelligence: Forecasting, prevention, and mitigation” 2018 *Computer Science* 25.

303 Tutt A “An FDA for algorithms” 2017 *Administrative law review* 105.

difficult to authenticate their outputs or detect inaccuracies or biases in the data sets used.³⁰⁴

Even if algorithms were designed with due regard to characterised legal standards, it may be rather challenging to establish if the algorithm behaved in accordance with the legal standard in any given scenario.³⁰⁵ This is especially difficult when different stakeholders (such as government) need a basic degree of transparency in relation to use, and with the expectation of providing others with unambiguous explanations. The issue is that AI outcomes often lacks transparency as doctors and even patients cannot verify how the technology came to its decision.³⁰⁶ The EU's GDPR provides for access to "meaningful information about the logic involved" relating to automated decisions.³⁰⁷ When using AI deep learning, it will be difficult to comply with GDPR and this restrictive regulation may inhibit the ongoing development of AI. It would be desirable if the GDPR provides for separate regulation, or exemption of AI that recognises the nuances of technology.

Errors will likely be made due to the application of AI systems in healthcare. Various theorists outline two conditions to impute moral accountability to an individual for his conduct: (1) the "control condition" (where loss of control over a decision is not on account of negligence); and (2) the "epistemic condition" (appreciation of the decision and its ensuing consequences, where negligence is not the cause of ignorance).³⁰⁸ It may be difficult to determine accountability, not only in those situations where physicians apply the suggestions of AI, but also when they elect to overrule or disregard the system's suggestions, and the physician makes the final decision. In the latter instance, both the "control condition" and "epistemic condition" as outlined above, would be met.³⁰⁹

304 Nuffield Council on Bioethics <http://nuffieldbioethics.org/wp-content/uploads/Artificial-Intelligence-AI-in-healthcare-and-research.pdf> (Date of use: 31 March 2019).

305 Tutt 2017 *Administrative law review* 105.

306 Matsuzaki 2018 *California Western law review* 269.

307 Article 15.

308 Habli I, Lawton T and Porter Z "Artificial intelligence in health care: accountability and safety" 2020 *WHO* 252.

309 Thompson CL and Morgan HM "Ethical barriers to artificial intelligence in the national health service, UK of Great Britain and Northern Ireland" 2020 *WHO* 293.

However, even in terms of the final choice, the physician cannot directly modify or interfere in the system's decision-making, once in progress, and has no knowledge that the system is arriving at decisions that reflect the physician's clinical aims or desires. The decision as to whether to apply the system's recommendations will also be motivated by broader structural and organisational procedures and policies, such as a physician's workload and the degree of dependence on automation.³¹⁰

Therefore, this raises a question with regards to who or what should be held accountable when injury or harm arises due to AI, and whether the blame should lie with the developer, producer, data supplier, the health institution that procured and implemented the AI system, or the healthcare professional who used and relied on it?

Currently there is no classification of adequate performance standards or accuracy levels against which to measure AI, and there are differing perspectives on the performance standards that should apply to measure AI tools against, in order to be safely applied in healthcare, and for both doctors and patients to trust the technology.³¹¹ Some in the medical community argue that that AI tools, particularly those on patient diagnosis, should always generate accuracy levels equal to-, or above those delivered by highly skilled doctors. Others hold the view that, as some of these AI tools are aiding patients who in daily circumstances have no or little access to skilled doctors, lower accuracy levels would be adequate.³¹²

However, the issue is that, if intelligent machines are introduced into healthcare because they are considered to outperform humans at making certain logical judgments: in what way is it possible for humans to rationally argue against the decisions they make?³¹³

310 Habli, Lawton and Porter 2020 *WHO* 252.

311 USAID <https://www.usaid.gov/cii/ai-in-global-health> (Date of use: 22 May 2020). There is also dissenting opinions regarding the accuracy standards amongst the various disciplines for AI tools, such as, epidemic predictive models that may require a lesser degree of precision than clinical decision support models.

312 USAID <https://www.usaid.gov/cii/ai-in-global-health> (Date of use: 22 May 2020).

313 Kerasidou 2020 *WHO* 247.

Not all AI models can be said to be complex and it is possible to design simple AI models that tend to be more explicable.³¹⁴ This is particularly crucial in cases relating to any decision-making impacting on individuals. Over recent years, AI advanced through developing systems from the bottom-up, by using a vast amount of data sets to teach them. Moving forward, however, scientists aim to develop top-down systems, not requiring the huge volumes of data sets (which are quicker, more adaptable, and more intrinsically intelligent), which strongly bear a resemblance to the way humans deal with problems-solving and tasks.³¹⁵

According to Accenture's AI Maturity report published for the Middle East and Africa, 94 per cent of companies support a top-down involvement in AI.³¹⁶ Many organisations are working towards designing machines that will direct the world through applying common sense, in order to grasp daily objects and behaviours, converse naturally, manage unanticipated scenarios, and learn from certain experiences.³¹⁷ Scientists have already started to design AI neural networks capable of performing complex reasoning tasks, in a transparently-interpretable manner.³¹⁸

The scientists and the organisations that employ them, therefore, have the opportunity to design the security landscape of the AI-enabled world in endeavouring to achieve user and regulatory acceptance. The "Smart Dubai" government office developed the "Ethical AI Toolkit", that proposes making AI models explicable and to identify and mitigate any significant risks inherent in the AI designed, in an

314 An example of simple AI models are digital assistants for nurses and doctors that are able to perform repetitive administrative tasks that do not require adjusted parameters on outcomes.

315 Wilson HJ, Daugherty PR and Davenport C "The future of AI will be about less data, not more" 2019 *Harvard Business Review* <https://hbr.org/2019/01/the-future-of-ai-will-be-about-less-data-not-more> (Date of use: 20 May 2020).

316 Microsoft <https://info.microsoft.com/rs/157-GQE-382/images/MicrosoftSouthAfricanreportSRGCM1070.pdf> (Date of use: 15 June 2020).

317 A company known as AI2 is developing a set of tasks against which progress can be measured with the goal of establishing what it means for machines to possess common sense. Researchers from Microsoft and McGill University have collaborated on a system that is aimed to resolve ambiguities in language processing, a problem that requires various forms of inference and knowledge base. Wilson, Daugherty and Davenport 2019 *Harvard Business Review*.

318 Mascharka D *et al* "Transparency by design: closing the gap between performance and interpretability in visual reasoning" 2018 *Cornell University* 4942-4943.

endeavour to eradicate the “black box” problem inherent in AI.³¹⁹ To assist users, security mechanisms (to mitigate the risk of users blindly placing their trust in AI) can be implemented. Educating patients and prescribing that a disclaimer be read and understood (on each occasion when an AI solution is used to provide a result or recommendation) could be of assistance.

2.2.8. Challenges for governance

AI embedded machines is already making financial-, health related-, legal-, and other decisions, which will have a significant influence on individuals that are the targets or subjects of those decisions. It will undoubtedly challenge established ethical and regulatory frameworks, and with the realisation that we should not give machines authority without responsibility, a key issue for consideration is if AI ought to be regulated as a separate field, or if the various fields of regulation ought to be considered with due regard to AI’s likely effect on individuals?

Additional obstacles involve the requirement for assurance that AI is created and implemented in an open, accountable, and harmonious manner, where public interest is offset against the need to prevent stifling innovation.³²⁰ It is important that, society is educated and prepared for these new found platforms of knowledge which AI offers and that there are safeguards in place in order to protect those who are disadvantaged by it.

This newfound knowledge will place society in a better position, but only if we jointly carry the obligations that these discoveries will impose on individuals.³²¹ The challenge for stakeholders and legislators is to be educated in order to gain the relevant knowledge so as to evaluate, and limit, any negative consequences of AI, and to establish structures to monitor key issues.

The governance of AI enabled systems is further challenged by divergent and weak regulatory- and policy ecosystems, which exist not only amongst countries, but also

319 Smart Dubai Artificial intelligence principles & ethics” <https://www.smartdubai.ae/initiatives/ai-principles-ethics> (Date of use: 16 June 2020).
320 Powles J and Hodson H “Google DeepMind and healthcare in an age of algorithms” 2017 *Health Technology* 351-367.
321 Powles and Hodson 2017 *Health Technology* 351-367.

within territories of countries, because AI is an emerging field.³²² For example, many LMIC regimes do not have the financial resources and technological knowledge to generate coherent strategies in relation to population health and analysis, nor monitoring and treatment protocols for use on a national level. This establishes a hurdle for adoption of AI technology when it comes to the delivery of healthcare services. Various LMIC countries have also not yet adopted effective regulations for the implementation of AI technology by different sectors and AI suppliers, with the effect that AI operability is generally dependent on the inherent discretion of local officials.³²³

To ensure a sustained advancement and recognition of AI driven tools it becomes crucial that, clear direction and unanimity comes from multilateral policy makers and regimes regarding when and where its regulation is needed. Awareness and education initiatives must be a critical obligation for public- and private participants that, design, deploy, and incentivise the technology, as society will only support and have trust in solutions they can understand and appreciate. This requires government and technologists to find innovative approaches to link-up users to innovative technology.

For instance, new methods to data usage and security policies will be useful. Protracted and complex policies that embody inscrutable technical terms and legal jargon, which the ordinary user will not understand, must be avoided.

2.3. Conclusion

AI has an imperative function in the healthcare service delivery of the future. It has the capability to augment healthcare access and efficiency, particularly for rural settings of LMICs.³²⁴ The technology is promising, and it guarantees that the most critical part of healthcare is accomplished for the future, i.e. a congenial patient experience.

322 USAID <https://www.usaid.gov/cii/ai-in-global-health> (Date of use: 22 May 2020).

323 USAID <https://www.usaid.gov/cii/ai-in-global-health> (Date of use: 22 May 2020).

324 Guo and Li 2018 *Health Equity* 180.

Perhaps the most prominent characteristic of AI is that, it is so flexible that it can be applied to various situations and these opportunities continue to expand in various areas of healthcare. We are witnessing the technology being applied in areas as diverse as data and diagnostic analytics, distributing of medicine, assisted robotic healthcare, and the detection and treatment of illnesses more swiftly and efficiently. AI ML, it is the driving force behind the progression of predictive healthcare, widely acknowledged as a critical advancement in healthcare.

Earlier attempts at offering diagnostic and treatment approaches have turned out to be challenging, but it is anticipated that AI will in future, overcome the challenges and offer a resolution in this area as well.³²⁵ Medical imagery or scanning is widely employed in clinical practice for analysis and therapy. Due to the already swift improvements in AI for imaging assessment, some of the work of radiology and pathology images are already being examined by machines. This is particularly useful for inexperienced physicians, especially those working in rural areas with relatively low healthcare quality, and where writing medical imaging reports is challenging and requires complex skills they might not yet possess for experience. It is also beneficial for physicians in countries with large populations where writing imaging reports is time consuming.³²⁶

In the near future, AI systems are likely to become more cutting-edge and will be able to execute a broader scope of complex tasks, independently of human command or involvement. Should this materialise, it will become necessary for AI systems to also learn to produce ethical judgments, having due regard to the ethical apprehensions that arise when these systems are used.³²⁷ Matters of accountability, security of data, consent, transparency, and public trust are essential when deploying AI in healthcare.

Therefore, it will be unacceptable for governments or nations to simply allow the implementation of AI systems. They should individually create their own ethical

325 Davenport and Kalakota 2019 *Future healthcare journal* 97.

326 Jing B, Xie P & Xing EP "On the automatic generation of medical imaging reports" 2018 Carnegie Mellon University 1.

327 Nuffield Council on Bioethics <http://nuffieldbioethics.org/wp-content/uploads/Artificial-Intelligence-AI-in-healthcare-and-research.pdf> (Date of use: 31 March 2019).

structures to manage the design and application of AI systems and towards identifying, measuring and responding to the risks of AI. The framework must be contextual and should promote ethical and accountable use. Explicit regulation and policies must be implemented, built on the solid foundation of existing governance and control structures, in order to provide direction on when, where, and in what manner AI technologies must be implemented, to control the way that it is used in healthcare settings, analogous to those that are already deployed or in production by the European Union (EU) and the WHO.³²⁸ This will offer the legal support and framework, necessary to deal with litigation that may emanate on account of the abuse or misuse of AI. This can also lay the groundwork towards the acceptance of AI for the healthcare industry, mainly for countries such as South Africa which would benefit from a progressive healthcare system.

The “Pandora’s box” with AI is already unlocked, and in order to leverage the capabilities of these digital technologies towards realising national- and international healthcare objectives, AI technologies must be certified and officially accepted by regulators, standardised to a reasonable degree, and cautiously integrated into healthcare systems, taught to healthcare professionals, funded by government (or private stakeholders) and updated in the field as the technologies evolve.³²⁹

This will require all stakeholders to develop a culture of collaborative and accountable innovation, building of trust, and a governance architecture in order to realise the values and principles of ethical, fair, and safe AI.³³⁰ Brad Smith, President of Microsoft expresses that,

[a]s this technology continues to grow, we will work to deploy AI around the world ethically, inclusively, and with transparency to ensure that it works for everyone.³³¹

-
- 328 The EU’s High-Level Expert Group on Artificial Intelligence published “Ethics Guidelines for Trustworthy AI” in 2019. Also late in 2019, The International Telecommunication Union (ITU) and the WHO collaborated to establish a Focus Group on AI seeking to promote worldwide regulation, standard and guidelines for AI in the digital health environment.
- 329 Davenport and Kalakota 2019 *Future healthcare journal* 97.
- 330 The Alan Turing Institute “Understanding artificial intelligence ethics and safety”; Mariano 2020 *WHO* 231. See https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf (Date of use: 21 May 2020).
- 331 Microsoft <https://info.microsoft.com/rs/157-GQE-382/images/MicrosoftSouthAfricanreportSRGCM1070.pdf> (Date of use: 15 June 2020).

In the following Chapter after an analysis of the challenges arising from AI, I will discuss why AI requires regulation, the risks and difficulties of regulating AI and its possible impact.

3. CHAPTER THREE: REGULATION AND ITS IMPACT ON AI

This Chapter commences by examining the different connotations to the term “regulation”, which often causes difficulties and confusion in understanding the concept. An explanation is offered on how regulation is employed in this thesis, to achieve a working definition for regulation towards the aims and objectives of this thesis. After defining regulation, this Chapter discusses the reasons behind regulatory intervention for AI, as the impetus for regulation. It also recognises that there are challenges related to assigning of accountability by way of regulation when sophisticated socio-technical technologies (such as AI) are applied, and the possible impact that a regulation may have on AI.

3.1. *A definition for regulation*

In the Oxford dictionary, the term “regulation” is assigned different connotations. In a general sense it means “the action or process of regulating or being regulated.”³³² In a governmental sense, it can take the form of “a rule or directive made and maintained by an authority which controls an activity or process.”³³³

The different meanings assigned to “regulation” are extensively canvassed in scholarly literature. While “regulation” encompasses a wider definition, being the “intentional influencing of someone’s or something’s behaviour (comprising of wider social or commercial influence) according to standards or goals”,³³⁴ the term is commonly understood as government intervention relating to private rights and

332 Oxford Dictionary <https://www.lexico.com/en/definition/regulation> (Date of use: 30 May 2020).

333 Cliff Dekker Hofmeyer “AI Regulation in South Africa and the global regulatory trends” <https://www.cliffedekkerhofmeyr.com/export/sites/cdh/en/news/publications/2019/technology/downloads/Technology-Media-and-Telecommunications-Alert-1-July-2019.pdf> (Date of use: 30 May 2020).

334 Koops BJ “Should ICT regulation be technology-neutral?” in Koops BJ *et al* (eds) *Starting points for ICT regulation. Deconstructing prevalent policy one-liners* (IT & Law Series, T.M.C. Asser Press The Hague 2006) 4; Koops BJ “Ten dimensions of technology regulation: Finding your bearings in the research space of emerging technologies” in Goodwin MEA, Koops BJ and Leenes RE (eds) *Dimensions of technology regulation* (Wolf Legal Publishers Netherlands 2010) 309-310; Black J “Critical reflections on regulation” 2002 *Australian Journal of Legal Philosophy* 1; Lambert TA *How to regulate: A guide for policymakers* (Cambridge University Press New York 2017) 1-6.

preferences by way of the promulgation of statutory rules or decrees that prescribe the existing legal options.³³⁵

The latter understanding accords with the legal definition of regulation which is “a rule or order issued by a government agency and often having the force of law.”³³⁶ Thus, in the traditional and narrow sense, regulation can be seen as reserved to measures that are statutory, when government and legislation is involved.³³⁷

However, the existence of different ideologies about political affairs and government involvement leads to some challenges in grasping what government intervention means and whether it is needed. There is a common perception of “regulation” equating it with laws that serve industries and that operate primarily for its benefit, whilst others are of the view that regulation is for protecting and facilitating transactions based on general principles that are applied to supply and demand.

Lawyers often use the term regulation in relation to binding directives of governmental agencies.³³⁸ Thus, society is increasingly polarised along ideological lines, as people appear to reject notions that collide with their principles. Consequently, these difficulties contribute to the confusion in understanding the definition of regulation, leading to the creation and application of informal definitions for each scenario.³³⁹ Given the elusive nature of the term “regulation”, it is necessary to delineate the way in which the term is applied within the framework of this thesis.

So, what does regulation mean for this discussion? I return to the earlier reference to the common interpretation of the term “regulation” (i.e. government “intervention in the private domain”) by way of statutory rules or decrees.³⁴⁰ The enforceable directive is “a binding legal norm created by a state organ that intends to shape the conduct of individuals and firms.”³⁴¹ A “state organ” could refer to a government

335 Orbach B “What is regulation?” 2012 *Yale journal on regulation* online 3.

336 Merriam-Webster <https://www.merriam-webster.com/dictionary/regulation#legalDictionary> (Date of use 30 May 2020).

337 Brownsword R Rights, regulation and the technological revolution (Oxford University Press 2008) 7. Lambert How to regulate 1-6.

338 Orbach 2012 *Yale journal on regulation* online 5.

339 Orbach 2012 *Yale journal on regulation* online 3.

340 Orbach 2012 *Yale journal on regulation* online 3.

341 Orbach 2012 *Yale journal on regulation* online 6.

agency or any other regulatory agency which carries lawful authority to make a “binding legal norm”.³⁴² The definition denotes “intervention in the private domain”, instead of “intervention in choices”, and addresses the confusion brought about as a result of the opacity of the latter arising from different perceptions of ethics, religious beliefs, politics, and societal relations.

The notion of regulation, characterised as a state intervention of designed legal norms to intervene in matters concerning the economy or commonly as a means of social oversight,³⁴³ has been a field of research in public policy and governance for some time. This notion of regulation stems from the nineteenth century, at a time when philosopher John Stuart Mill, informally referred to the term “regulation” in reference to “governmental intervention in the affairs of society” and “laws that implement intervention.”³⁴⁴

Based on the various schools of thought concerning how regulation can be defined today, the concurrence is that regulation is a set of instructions, in the form of binding rules, created by government or another regulatory body; the act of the government intended to shape or steer certain behaviour; to direct the market, and other approaches to prescribing rules, structures, and stimuluses, like economic regulation, professional regulatory bodies and voluntary associations.³⁴⁵

With the onset of what is known as the fourth industrial revolution encompassing digitalisation, ML, BD, IoT, autonomous machines, and robotics, today’s regulatory states have already started to move towards e-regulation.³⁴⁶ E-regulation

342 Baldwin R, Cave M and Lodge M *Understanding regulation: Theory, strategy, and practice* 2nd ed (Oxford University Press New York 2012) 2-3.

343 According to Lambert, to increase total social welfare, regulation will constantly involve “mixed-bag” of human conduct that consists of both good and bad, which requires policymakers to “craft legal directives so as to prevent the bad aspects of mixed-bag behaviour without simultaneously forbidding or discouraging the good aspects”; Lambert *How to regulate* 8.

344 Mill JS *Principles of Political Economy with some of their applications to social philosophy* 1st ed (John W. Parker London 1848) 525-71.

345 Baldwin, Cave and Lodge *Understanding* 2-3.

346 For example, in this age of e-regulation, administrative agencies use social media, internet platforms and mobile applications for regulatory purposes. In the US machine-readable regulatory data published on Data.gov facilitates the private-sector development of software applications which provide users with detailed information on restaurant and market health violations issued in certain sectors based on city inspection See DATA.GOV <https://www.data.gov/opendata/get-local-government-data-gov/> (Date of use: 2 July 2020).

constitutes a paradigm shift in government regulatory strategies requiring governments to monitor and control behaviour through digitised governmental regulation, public responsiveness, and corporate reputation. Therefore, the e-regulation era does not, at all, suggest a withdrawal by the regulatory state but rather an evolution. This is because it sets rules and responds to changes in technology, markets, and public expectations.³⁴⁷ Furthermore, is more focused on how regulators should approach novel technological disciplines.³⁴⁸

For the purpose of this thesis, regulation consists of a top-down approach where the government and its agencies sets binding rules to regulate AI. The research focuses on how they can harness and influence the advancement of AI. The approach I am advocating for is contemplating amendments to laws or purposeful regulation as an ongoing approach in regulating behaviour in AI technological opportunities.

3.2. *The reasons to regulate AI*

3.2.1. To prevent public harm and advance social goals

Criticisms and concerns with tough regulatory conventions gained prominence during the US banking and financial crisis (2007 and 2009), which impacted institutions and many individuals,³⁴⁹ when demands for deregulation or malleable regulation were superseded by ongoing calls for stricter regulation relating to financial trades. This has dramatically changed attitudes concerning the appropriate role of government in controlling markets. Baldwin *et al* states that, the result of the catastrophe and calls for intervention was the era of the “regulatory state” which led to scholars developing novel approaches when contemplating regulation and the establishment of a global “regulatory community” (for exchanging of common ideas, challenges, and concerns).³⁵⁰

347 Yadin S “E-regulation” 2019 *Cardozo arts & entertainment law journal* 152.

348 Moses LB “How to think about law, regulation and technology: Problems with ‘technology’ as a regulatory target” 2013 *Law, innovation and technology* 5.

349 Such problems came to the fore in the aftermath of scandals in the US involving financial firms such as Enron, where the incentives of the firms making the lending decisions were not aligned with the consumers’ interest or even that of their shareholders’.

350 Baldwin, Cave and Lodge *Understanding* 1-2.

The rapid growth and implementation of AI is an important technological evolution of today's civilisation.³⁵¹ This AI evolution can bring immense opportunities to promote the sustainable and natural progression of today's society and the ecosystem. However, in the absence of effective regulation and governance, AI's effects could be unparalleled and harmful.³⁵² For these transformations to become beneficial (prior to them becoming entirely entrenched into the structure of society), we must create a concrete and viable governance framework for AI (according to the ethics and values of humanity), in order to guide the behaviour of people and corporations whilst, achieving goals considered important for society.

The objective of AI must be value creation for the greater society, and the measure of its success lies in the value it can generate for human lives.³⁵³ Progressive technologies such as AI can positively contribute towards healthcare offerings of the future. It also requires new interactions between man and machine to be regulated as this relationship becomes ever more nuanced, agile, and personalised. The notion of capitalism and value creation prompted by AI, should not undermine the interests of patient- and physician integrity, and unintentional or unforeseen harm must not be overlooked.

According to Susan Brenner, there is the propensity to consider problems with new technology in relation to its "misuse", which encompasses a wilful endeavour to disrupt basic principles of social order.³⁵⁴ If we were to consider how we should guard against threats linked to the technology abuse, the possible conclusion one arrives at is that, there is a necessity to introduce laws intended to control the risks linked to the technology, and to discourage the infliction of harms that threaten the social order.³⁵⁵ The purpose and justification of imposed directives through a

351 See the definition for machine intelligence in the chapter one as "the study and design of intelligent agents".

352 Wu W, Huang T and Gong K "Ethical principles and governance technology development of AI in China" 2020 *Science direct* 308.

353 Saxenian A *et al* "Artificial intelligence and life in 2030: One-hundred year study on artificial intelligence" 2015-2016 *Study Panel, Stanford University* 42.

354 Brenner SW *Law in an era of smart technology* (Oxford University Press New York 2007) 21-22

355 Brenner *Law in an era* 21-22.

regulatory system should therefore be to create accountable AI and to promote society's trust in such technology.

In South Africa, disparities in wealth, education and healthcare are amongst the highest in the world. Access to basic healthcare is an essential goal which can be realised with the assistance of AI-enabled technologies that can expand the capacity of healthcare providers and reduce the cost of healthcare, in particular for the South Africa that has a large rural population. However, despite the benefits that are already seen globally from AI in healthcare sectors, LMICs such as South Africa will remain vulnerable to the technology due to unstructured medical data and mismanagement of sensitive health data in these countries. Therefore, to successfully integrate the technology into the South African healthcare market, the government must have laws in place to control the risks linked to the technology. It is important that these laws address issues of ethics, trust, fairness and data protection.

3.2.2. To prevent market failures

Market forces on its own can be the subject matter when it comes to government regulation. One of the reasons behind regulatory intervention relates to the issue of market failure. Many common-law doctrines address market failures, which is recognised as a state of affairs where the unregulated marketplace fails to generate behaviour or outcomes based on public interests, and where market forces fail to allocate value effectively.³⁵⁶

Government will intervene in an attempt to remedy the failure where such intervention is attained through regulation.³⁵⁷ In the case of a monopoly and/or abuse being an example of market failure, actors overlook the socially preferred standard of an enterprise by invoking responsibility when the ultimate production costs of the enterprise are significantly in excess of ultimate rewards, leading to the

356 Moss D & Cisternino J, *New perspectives on regulation* 1st ed (The Tobin Project 2009) 11; Baldwin, Cave and Lodge *Understanding* 15-22.

357 Terry NP "Of regulating healthcare AI and robots" 2019 *Yale journal of law and technology* 26; Iszaid I, Hafizan A and Muhamad HJ "Market failure in health care: A review" 2018 *International journal of public health and clinical sciences* 21.

obliteration of a robust competitive environment.³⁵⁸ Large scale AI projects in healthcare technology are currently being subsidised by other businesses (owned by technology giants Amazon and Google).³⁵⁹

For instance, Google's parent company, DeepMind, posted losses of USD\$ 341 million in 2017 and USD\$ 572 million in 2018.³⁶⁰ Amazon's growing footprint in healthcare is congruent with its commercial objective (of extracting revenue from all operations coupled with its own healthcare objectives).³⁶¹ Smaller tech players in the market are therefore not able to compete with corporations that are backed by "big" money.

Healthcare is regarded as a "market failure" because it obstinately defies market solutions, and failure in competition profoundly impacts everybody as it results in healthcare costing being far more than it should, relative to what it delivers.³⁶² A part of such failure has to do with "price discovery", where the markets cannot function optimally unless the buyer is aware of the price of something before accepting the order.

In most instances, the real cost for your medical services only become evident once services are executed, or following the insurance company's review and approval of the healthcare provider's account.³⁶³ A patient's incapability to calculate the cost of medical treatment (or to measure the quality of healthcare received) inhibits the

358 Hylton KN "Nuisance" 2014 Encyclopedia of law and economics 1.

359 Terry 2019 *Yale journal of law and technology* 25.

360 Marcus G "DeepMind's losses and the future of artificial intelligence" <https://www.wired.com/story/deepminds-losses-future-artificial-intelligence/> (Date of use: 5 July 2020).

361 Becker's Health IT "15 things to know about Amazon's healthcare strategy heading into 2020" <https://www.beckershospitalreview.com/healthcare-information-technology/15-things-to-know-about-amazon-s-healthcare-strategy-heading-into-2020.html> (Date of use: 5 July 2020).

362 Brodwin D "Healthcare is a market failure" <https://www.usnews.com/opinion/economic-intelligence/articles/2017-06-23/senate-obamacare-repeal-plan-ignores-market-failure-of-us-health-care> (Date of use: 5 June 2020).

363 Brodwin <https://www.usnews.com/opinion/economic-intelligence/articles/2017-06-23/senate-obamacare-repeal-plan-ignores-market-failure-of-us-health-care> (Date of use: 5 June 2020).

ability of the healthcare market to deliver optimal conditions to the patient.³⁶⁴ In an open market, transactions do not ensue unless there is consensus on the value.

In South Africa, improvements in access to healthcare are obstructed by increasing economic barriers and corruption leading to a failure or delay in implementing many advances in healthcare. This has resulted in healthcare costing being more than it should, relative to what it delivers. The only way to resolve this issue is to create a balanced healthcare system that offers more innovative opportunities. The proposal for a national health insurance (NHI) in South Africa is part of a welcome resurgence that will allow the patient to measure the cost and quality of healthcare received.

Another reason for market failure in healthcare relates to the immense lack of real competition, or barrier to entry in the industry that inflates the price of goods or services in healthcare.³⁶⁵ In many countries, including South Africa, all healthcare facilities and health insurance schemes are owned by two or three major chains due to obstacles or hindrances,³⁶⁶ which deter new entrants from entering the healthcare market, leaving no room for negotiating prices as the consumer is forced to transact with a particular supplier.³⁶⁷

There is also a substantial reason for opposing new technologies due to the element of unfair competition, because of the concern that they generally advantage already privileged and wealthy people over others.³⁶⁸ For example, in today's era of smart technology, information is readily available and makes wealthy patients desperate enough to travel worldwide to obtain medical treatment at their own costs, creating

364 Iszaid, Hafizan and Muhamad 2018 *International journal of public health and clinical sciences* 21-22.

365 Iszaid, Hafizan and Muhamad 2018 *International journal of public health and clinical sciences* 22.

366 Obstacles or hindrances may comprise of technology treatment barriers, government regulations, patents, start-up costs, licensing requirements, etc.

367 Brodwin <https://www.usnews.com/opinion/economic-intelligence/articles/2017-06-23/senate-obamacare-repeal-plan-ignores-market-failure-of-us-health-care> Date of use: 5 June 2020).

368 Pasquale F "Technology, competition, and values" 2007 *Minnesota journal of law science and technology* 607.

a corrosive environment that can give rise to supplier-induced demand, which only favours a certain sector of society.³⁶⁹ According to Friedrich Hayek:

[W]here it is impossible to create the conditions necessary to make competition effective, we should resort to other methods of guiding economic activity.³⁷⁰

A commitment to free markets and vigorous competition can fundamentally be achieved through regulatory intervention, in order to remedy the market failures as well as enhance social welfare, affected by private ordering defects. In order for individuals acting on their own to obtain ideal outcomes, which do not amount to a trade-off of one individual interest for another, they must be well-informed and operate in competitive marketplaces.

In the absence of these ideal circumstances, governments impose regulations to prevent economic exploitation and to potentially increase societal efficiency and/or equity.³⁷¹ Some of these major contributors of government interventions are widely accepted. For example, antitrust (competition) laws serve as a deterrence for monopoly-power and/or abusive practices, whilst data and consumer protection laws are intended to tackle issues of exploitation due to information misuse.

Similarly, government intervention is essential for ensuring that AI's uses are protected against commercial exploitation and market failures.³⁷² A governance framework will promote a democratic and fair distribution of AI's benefits in order to reduce existing inequalities in healthcare and to avoid the concentration of control and advantages for the privileged few.

3.2.3. Protection of fundamental human rights

Further reasons to regulate arise from governments' duty to protect fundamental human rights and social solidarity.³⁷³ The basic right of access to healthcare is

369 Iszaid, Hafizan and Muhamad 2018 *International journal of public health and clinical sciences* 22.

370 Hayek FA *The Road to Serfdom* (University of Chicago Press Chicago 1944) 37.

371 Moss & Cisternino *New perspectives* 11-12.

372 AI-based applications could improve access to healthcare simply by its inherent ability to correct market failures such as inflated healthcare costs stemming from monopolies.

373 Baldwin, Cave and Lodge *Understanding* 22; Baweja S and Singh S "Beginning of Artificial Intelligence, End of Human Rights"

determined by an intricately connected set of human rights, that entail rights to fairness, dignity, privacy, and physical- and mental integrity.³⁷⁴ For instance, the government's authority to limit any basic rights during an epidemic (by collecting personal information for purposes of contact tracing) must be measured against its duty to uphold its constitutional duties, by obeying the right to privacy. For example, in South Africa, Section 14 of the Constitution protects the right to privacy and protection of an individual's personal information. In the case of *Bernstein v Bester* 1996 (2) SA 751 (CC), the court discussed the right to privacy and stated that:

A very high level of protection is given to the individual's intimate personal sphere of life and the maintenance of its basic preconditions and there is a final untouchable sphere of human freedom that is beyond interference from any public authority. So much so that, in regard to this most intimate core of privacy, no justifiable limitation thereof can take place.

Regulation through protective laws serve as lawful and useful approaches to deter and anticipate violations to protect these rights.³⁷⁵ It is based on serving legitimate public interests of life, health, safety, and freedom. When these basic rights are threatened or disregarded, they may have socially undesirable outcomes. Although AI is viewed as an enhancement of society's present-day healthcare, the conflict between this technological advancement and basic human rights is simultaneously being exposed.³⁷⁶

AI is viewed as a threat to equal security, economic rights, and fundamental liberties.³⁷⁷ These threats are discussed in Chapter two, which includes the right to privacy, fairness, accountability and transparency, and job security. In the US case of *KW v Armstrong*,³⁷⁸ it was held that the unreliability of the "budget tool" software program, used by the Medicaid administrator of the Idaho Department of Health and

<https://blogs.lse.ac.uk/humanrights/2020/07/16/beginning-of-artificial-intelligence-end-of-human-rights/> (Date of use: 20 July 2020).

374 These rights are provided for in the Constitution of South Africa.

375 An example of the regulation of human rights is that of the European Convention on Human Rights (ECHR), which allows any individual whose rights have been allegedly violated under the Convention by a member nation to lodge a case with the court.

376 Baweja and Singh <https://blogs.lse.ac.uk/humanrights/2020/07/16/beginning-of-artificial-intelligence-end-of-human-rights/> (Date of use: 20 July 2020).

377 Baweja and Singh <https://blogs.lse.ac.uk/humanrights/2020/07/16/beginning-of-artificial-intelligence-end-of-human-rights/> (Date of use: 20 July 2020).

378 180 F. Supp. 3d 703 (D. Idaho 2016).

Welfare, involving AI automatic spreadsheet calculations used to prepare home health aide budgets, was unreliable as it improperly reduced assistance for some recipients and arbitrarily deprived participants of the property interest in their benefits when their budgets were decreased and thus, violated due process.

To reverse these negative trends, suitable legal standards are required towards achieving a form of “good AI societies”, in which human dignity may thrive.³⁷⁹ To this end, the CPA in South Africa was passed to safeguard consumer interests and safety, in recognition of the reality that emerging technological changes, and trading activities would continually offer new benefits, prospects, and raise concerns.³⁸⁰

Many legal issues result from the new technological front, such as the privacy implications of social media, or access to public health data, ethical concerns linked to medical technologies, or developments in technology from a point of bias, accountability and transparency. Thus, regulation can be used not only as an instrument for promoting an equitable economic environment, but also as a means to preserve the values that are considered to be important by society. It can also ensure that responsibility for undesirable consequences linked to the advancement and deployment of pioneering technologies is appropriately assigned.

By its very nature, a regulation restricts an individual or firm from doing what it otherwise would have done, for the avoidance of loss of profits and threatened barriers to innovation. However, the purpose of government intervention is to deal with potential negative outcomes towards the protection of the general public interest, in instances where private profit is not a good measure of social impact.³⁸¹

3.2.4. Industry advancement

Regulation can also serve towards the control and advancement of a specific industry sector. Through self-monitoring, industries could manage their own compliance to ethical-, legal-, and safety requirements, in lieu of having an external

379 Cath C *et al* “Artificial Intelligence and the ‘Good Society’: the US, EU and UK approach” 2018 *Science and engineering ethics* 506-509.

380 Preamble to South African CPA.

381 Moss & Cisternino *New perspectives* 13.

agency having to oversee them. However, history is riddled with instances of organisations that have prioritised revenue, investor returns, and competitive gain over social responsibility. In furtherance of these objectives, there are corporations that participate in socially- and morally irresponsible practices, and regulation has to be forced upon them by governments as a rule of law.³⁸²

Whilst regulation is instituted primarily for the safety and benefit of society at large, it can also be designed and operated for a sector's benefit by providing the sector with legal certainty, and through offering various kinds of incentives and financial aid. Hence, enhancing its development in a way that it would not have without the intervention. Industrial technologies are context specific with distinct functions and each are implemented in a precise societal context, requiring the assistance of new specialists who are trained to carry out complex industrial processes.³⁸³ In the healthcare industry, if medical experts and doctors can significantly contribute within society's goals, regulators will be encouraged to participate in the sectors' development and growth, by assessing who is likely to obtain the gains or obligations of regulation, how it could find application, and its impact on the distribution of resources.³⁸⁴ The significance of AI's regulation for industry has also been expressed by leading technology innovators. Elon Musk states that,

there should be some regulatory oversight, maybe at the national and international level, just to make sure that we don't do something very foolish.³⁸⁵

Academics and renowned individuals in the computer science and technology sector have also pledged their support for regulation by investing towards keeping AI safe whilst preserving its innovation.³⁸⁶

382 Stigler JG "The Theory of Economic Regulation" 1971 *The Bell journal of economics and management science* 3 <https://www.sjsu.edu/faculty/watkins/stigler.htm> (Date of use: 5 June 2020).

383 Brenner *Law in an era* 29.

384 Stigler 1971 *The Bell journal* 3.

385 The Guardian "Elon Musk: artificial intelligence is our biggest existential threat" <https://www.theguardian.com/technology/2014/oct/27/elon-musk-artificial-intelligence-ai-biggest-existential-threat> (Date of use: 5 June 2021)

386 Elon Musk invested USD\$10 million of his own funds to finance an endeavor to keep AI user friendly, see Mack E "Why Elon Musk put down \$10million to keep artificial intelligence friendly" <https://www.forbes.com/sites/ericmack/2015/01/15/elon-musk-puts-down-10-million-to-fight-skynet/?sh=1fe213452e5b> (Date of use: 6 June 2020). From Max Tegmark's Future of Life Institute to the Harvard Kennedy School of Government's Future

In South Africa, it would be important for industries to adapt AI technology to the African market as most of the technology is currently being developed in Western or Asian market. The technology should be adapted to factors such as the local languages and speech, and the different healthcare needs and management structures in the South African market. In this way industries can define what the AI systems should contain and can structure its development processes that allows for ethical and safety considerations of the local population.

3.2.5. Legal necessity to regulate

The judiciary and legal practitioners are grappling with new, but extremely significant, challenges regarding AI and its vast applications. This may have been a motivating reason behind the American Bar Association's promulgation of Resolution 112 (12 August 2019), regarding AI. It states that,

RESOLVED, That the American Bar Association urges courts and lawyers to address the emerging ethical and legal issues related to the usage of artificial intelligence ("AI") in the practice of law including: (1) bias, explainability, and transparency of automated decisions made by AI; (2) ethical and beneficial usage of AI; and (3) controls and oversight of AI and the vendors that provide AI.³⁸⁷

In 1942, "the three laws of robotics" was introduced by Isaac Asimov under his works called "I, Robot".³⁸⁸ Asimov's "three laws of robotics" states that,

A robot may not injure a human being or, through inaction, allow a human being to come to harm (the "first law");
A robot must obey orders given to it by human beings except where such orders would conflict with the first law (the "second law"); and
A robot must protect its own existence as long as such protection does not conflict with the first or second law (the "third law").³⁸⁹

Notwithstanding these fictional laws, Asimov persuasively offer rules that could be considered for AI, particularly, with reference to healthcare settings. The question arises as to what the existing legal framework (governing AI in South Africa) is. Due

Society, they are globally recognised experts that are collaborating to address AI's disruptive developments and risks, see Creighton J "OpenAI wants to make safe AI, but that may be an impossible task" <https://futurism.com/openai-safe-ai-michael-page> (Date of use: 6 June 2020).

387 Arkfeld M "Litigating and judging artificial intelligence cases" 2020 *Judges' Journal* 6.

388 Asimov I I, *Robot "Runaround"* The Isaac Asimov Collection ed (New York City: Doubleday 1950) 40.

389 Asimov I I, *Robot "Runaround"* The Isaac Asimov Collection ed (New York City: Doubleday 1950) 40.

to technology evolving at a rapid pace, legislators and regulations have battled to keep up. This view of the law “limping” or trailing behind new technology is shared by the public who regards innovators always as appearing to be one step ahead of policymakers.³⁹⁰

John Hennessy is the Chairman of Alphabet, a Google owned company, who states,

the technology industry moves too quickly for the government to effectively regulate it.³⁹¹

Advancements in AI present a distinctive phenomenon as the prospect of “general AI” that, equates to or outpaces human across various spheres, generates a competitive environment that encourages the hasty expansion of AI over and above its regulation. As a result, AI technology has already started to challenge existing legal regulatory frameworks. Moreover, the legal need to consider a regulation in this context is motivated as a result of the fact that conventional regulatory frameworks appear to be ill-equipped to deal with minimising risk to the public, while continuing to support innovation.

As computers are increasingly beginning to bear a resemblance to their human creators, in terms of their functions, this growing similarity might ultimately necessitate that, computers be assigned legal recognition as “persons”, a status that has been limited to humans thus far.³⁹² It seems obvious that the granting of some kind of legal status would be the first step toward acknowledging the legal rights vested in autonomous technologies.

The allocation of some form of legal status or personhood to autonomous machines or systems, either through specific legislation or through currently existing law, will be a significant event that warrants careful attention.³⁹³ Current laws are developed on legal principles centred on human conduct, which, if used in relation to AI, may

390 Calo R “Artificial intelligence policy: A primer and roadmap” 2017 *University of California* 406-410.

391 Musariri D “John L Hennessy: Profiling the man at the helm of Google’s parent company alphabet” <https://www.ns-businesshub.com/business/john-l-hennessy-alphabet/> (Date of use: 2 July 2020).

392 Willick 1983 *The AI Magazine* 5.

393 Chinen M *Law and autonomous machines: The co-evolution of legal responsibility and technology* (Edward Elgar Publishing Cheltenham UK and Massachusetts USA 2019) 234.

not be effective on account of the challenges in understanding how an AI program that, is trained on certain data sets reaches or forecasts its decisions or outcomes.³⁹⁴

Consequently, the doctrines of intent and causation within the contexts of criminal- and delict law are challenged, and are two of the most universal in South African law and many other jurisdictions.³⁹⁵

The application of these existing legal frameworks to AI technologies presents tough legal questions. Regarding the dilemma of new or changing technology and the law, Roger Brownsword references the “challenge of regulatory connection” concerned with the misalignment between regulatory frameworks and laws, which are outdated as they were intended for technologies that are surpassed by time and the regulatory disconnection with current laws, where new technologies enter into a “regulatory void”.³⁹⁶ Even some of the current regulations that deal with AI, such as US district laws regulating autonomous cars,³⁹⁷ fall short as these laws do not deal with the broader assertions about AI technology and instead focus on particular risks caused by specific AI applications.³⁹⁸

Principles and directives published in South Africa pursuant to healthcare laws envisage that, healthcare delivery is predominantly undertaken by humans and not intelligent machines. Some automated phases of healthcare service delivery, such as drug dispensing machines, have already been recognised by the authorities and

394 Neethling, Potgieter and Visser *Law of delict, H v Fetal Assessment Centre* 2015 (2) SA 193 (CC); *Oppelt v Department of Health, Western Cape* 2016 (1) SA 325; *Goliath v The MEC for Health, Eastern Cape* 2015 (2) SA 97 (SCA); *Links v MEC Department of Health, Northern Province* 2016 (4) SA 414 (CC)); Bathae 2018 *Harvard journal of law and technology* 891.

395 *H v Fetal Assessment Centre* 2015 (2) SA 193 (CC); *Van Schalkwyk v S* 2016 (2) SACR 334 (SCA); *S v Humphreys* 2015 (1) SA 491 (SCA); *S v Ngubane* 1985 (2) All SA 340 (A). Delict law requires that injury or harm ought to be reasonably foreseeable but, criminal law also requires that injury or harm is to be intended. US law places great significance to the principle of *mens rea*, i.e. the intention, to comprise the elements of guilt in criminal law.

396 Brownsword Rights, regulation and the technological revolution Chapter 6.

397 Several US states have enacted law concerning driverless vehicles, see National Conference of State Legislatures “Autonomous Vehicles - Self Driving Vehicles Enacted Legislation” <https://www.ncsl.org/research/transportation/autonomous-vehicles-self-driving-vehicles-enacted-legislation.aspx> (Date of use 9 June 2020).

398 The Future of AI Act 2017 passed by the US is another example of enactments specifically for AI that only creates a special committee to consider and renders guidance on AI related issues but steers away from general recommendations concerning AI.

are regulated under the guidelines of the South African Pharmacy Council.³⁹⁹ The HPCSA has also issued guidelines concerning telemedicine and the delivery of electronic diagnoses in medicine and associated information, as exchanged between healthcare practitioners and their patients.⁴⁰⁰

However, it becomes necessary to further research the regulation of AI in the provision of healthcare services, instead of responding reactively when things go wrong. With reference to the lessons that cyber-law can teach us about regulating robotics, Ryan Calo notes that, transformative technologies require new legal and regulatory frameworks because these technologies tend to distort the purpose of the existing laws and regulations.⁴⁰¹

There are also certain primary questions which should be examined, particularly with due regard to whether or not AI, in the widest form, when employed for healthcare settings would be regarded as an already regulated device or apparatus, particularly bare software algorithms. Novel and emergent medical devices are usually regulated for effectiveness and safety by health authorities of a country. In this regard, the South African government effected amendments in 2017 to the Medicines and Related Substances Act.⁴⁰² This legislation includes the introduction of a definition of “medical device”, now regulated in terms of law. This definition is broadly covers “any instrument, appliance, material, machine, apparatus, implant or diagnostic reagent.....”⁴⁰³

Premised on Asimov’s expectation regarding the degree at which intelligent robotics would operate in human civilisation, the issue is whether the present definition of “medical device” envisions AI in the delivery of healthcare that, would under usual circumstances be rendered by humans that are trained in healthcare. Having regard

399 Rule 1.9 of the Good Pharmacy Practice Manual and Associated SAPC Rules, containing minimum standards concerning the automated dispensing machines used to distribute medicines and devices, see South African Pharmacy Council “Good pharmacy practice manual and associated SAPC rules” https://www.pharmcouncil.co.za/Legislation_Rules (Date of use: 3 July 2020).
400 HPCSA https://www.hpcsa.co.za/Uploads/Professional_Practice/Ethics_Booklet.pdf (Date of use: 19 May 2020).
401 Calo R “Robotics and the lessons of cyberlaw” 2015 *California law review* 549-558.
402 101 of 1965 as amended (“Medicines Act”).
403 Section 1.

to the extent to which AI has developed and is developing in terms of its intellectual capabilities and autonomy, the present definition of “medical device” may not sufficiently cover instances where AI is predominantly perceptive and is capable of applying various scenarios of intelligence.

The healthcare regulations were not drafted with AI in mind as it contemplates that healthcare services are, for the most part, provided by natural persons as opposed to intelligent machines. The definition was intended to apply distinctively from scenarios of a human using the application of intelligent parts of AI devices for delivering healthcare in healthcare settings.

In 2017, IBM lobbied the US Congress to exempt its Watson Health, which applies data-driven analytics and advanced AI technologies in healthcare solutions, from device regulation, by arguing that it is not a medical device as its software is designed to merely assist physicians in diagnosis and treatment.⁴⁰⁴ This contributed to the partial deregulation of medical software, where lawmakers imposed that, software which analyses medical data and offers recommendations to healthcare specialists in areas of treatment or diagnosis is to be legally exempted, but only in instances where these specialists can evaluate and scrutinise the source of the data.

However, the tipping point is where AI becomes capable of independent decision-making and where it no longer allows the healthcare provider to understand and independently evaluate the source for its decisions, as the techniques used by the software cannot explain why or how they reach the conclusion they do, called “black box medicine”.⁴⁰⁵ Historically, there have been many milestones concerning medical device legislation in the US.⁴⁰⁶ For example, since the introduction of the Medical

404 Becker’s Health IT “IBM lobbies Congress to ease FDA regulations on Watson” <https://www.beckershospitalreview.com/healthcare-information-technology/ibm-lobbies-congress-to-ease-fda-regulation-on-watson.html> (Date of use: 5 July 2020).

405 Terry 2019 *Yale journal of law and technology* 17; Price 2017 *The SciTech lawyer* 10-11.

406 US FDA “A history of medical device regulation & oversight in the United States” <https://www.fda.gov/medical-devices/overview-device-regulation/history-medical-device-regulation-oversight-united-states> (Date of use: 5 July 2020).

Device Amendments Act of 1976. To date, the Food and Drug Administration (FDA) has struggled to keep up with rapid innovations in AI healthcare.⁴⁰⁷

As AI systems continue to develop the capability to conjecture additional knowledge from an already existing knowledge-base and theoretically to adapt and be self-thought, it is at the very least debatable as to whether non-customised software systems that results in harm is subject to strict product liability. In regulating medical devices, the lines also become blurred when healthcare applications, devices, or robots, primarily target a consumer rather than the medical market. Certain medical devices, such as radiography devices or robotic physician-assisting extenders, are to be regarded as regulated medical devices and must require premarket approval.

However, the same would not apply for caretaker robots, marketed for in-home application, that have autonomous or semi-autonomous characteristics or mobile applications, which are intended for only professional handling.⁴⁰⁸ Essentially, in “black-box” healthcare the conventional techniques of assessing new and emerging healthcare technologies for safety and efficacy will likely not to work for all instances.

So how can the issue of regulating AI as a medical device be tackled? The most effective path will likely be a more agile approach, as opposed to a rigid application when it comes to AI devices. This could involve softer pre-market inspection that, concentrates only on procedural safety such as the quality of data entered as well as the design and authentication procedures, followed up by vigorous post-market supervision as these technologies are deployed into healthcare.⁴⁰⁹

Authors Jason Chung and Amanda Zink⁴¹⁰ present four thought provoking questions regarding the execution of law to AI in relation to healthcare:

As technologies mature, laws to regulate AI must mature organically to fit the technology. As guiding questions, we propose that lawmakers adopt the following four-part test before beginning to regulate and/or restrict AI: (1) to what degree does the machine enjoy autonomy?; (2) to what degree does the machine interact with

407 Terry 2019 *Yale journal of law and technology* 15-19.

408 Terry 2018 *Journal of health care law & policy* 147.

409 Price 2017 *The SciTech lawyer* 11.

410 Chung J and Zink A “Hey Watson – Can I sue you for malpractice? Examining the liability of artificial intelligence in medicine” 2018 *Asia Pacific journal of health law & ethics* 78.

users/patients?; (3) to what degree does the machine provide reliable options?; and (4) to what degree does the machine implement such options?

These questions may provide a start for other questions to follow, on the issue of regulation. It is clear that the law should be flexible to cater for the range of very diverse tasks and dimensions of AI much similar to the way that one would ascribe varying standards to humans. As AI's technology is becoming more significant and intrusive at a level that is at this time, not complemented by legal developments, it will also have its influence on people, their communities, and society, permeating various branches of law.⁴¹¹

Hence, it is essential to modify the current legal norms so as to address the inevitable shifts introduced by emerging technologies. Precisely how the law will adjust to these technological developments (when it comes to AI) and how it will adjust to the ideals manifested in law, will depend on a multiplicity of social, commercial, and other considerations, which may differ from jurisdictions.⁴¹² The extent to which existing legal regimes and the frameworks (within which they are being developed) are said to be stifling AI innovations, means that it is time for regulators to consider whether those existing frameworks may or may not be working.

3.2.6. Regulation of AI is already underway

Whilst the broader debates about AI continue, developments in AI regulation are already underway in the shape of international guidelines, such as the: GDPR of the EU;⁴¹³ Ethics Guidelines for Trustworthy AI introduced by the European

411 See examples in the introductory to Chapter one. Also see the One Hundred Year Study on Artificial Intelligence's Study Panel report highlighting a number of regulatory issues which include: agency; privacy; labor; civil- and criminal accountability. Saxenian *et al* Artificial Intelligence and Life in 2030 44-48.

412 Saxenian *et al* Artificial Intelligence and Life in 2030 45.

413 GDPR (EU) 2016/679 <https://gdpr.eu/article-1-subject-matter-and-objectives-overview/> (Date of use: 20 June 2020).

Commission;⁴¹⁴ UK's Independent Review on AI and Public Standards;⁴¹⁵ and OECD Principles on AI.⁴¹⁶

To date, South Africa has not implemented any policies or legislation dealing with AI regulation. However, in order to aid the South African government in leveraging the promise of the fourth industrial revolution, in April 2019, the presidential commission on the fourth industrial revolution (4IR commission) was formed by the President. The commission is composed of members of various industries, such as technology companies, universities, cyber-security professionals, researchers, social experts, trade unions. Terms of reference for the 4IR commission were also gazetted, which includes the tasks of the 4IR commission to consider and implement appropriate policies and strategic planning, which will allow South Africa to participate as an international competitor for technological advances.⁴¹⁷

In the context of healthcare, the International Telecommunication Union (ITU) and the WHO have joined forces in order to establish a "Focus Group" relating to AI, with the objective of encouraging international "collaboration and regulation in the digital health ecosystem" and to create a standardised framework for the assessment of AI-developed approaches for diagnosis and treatment in the healthcare sector.⁴¹⁸

During May of 2018, WHO member states adopted a resolution concerning digital health with a global digital health strategy. There are four strategic goals of the global strategy, namely: (1) to encourage international participation and enhance transfer of information and knowledge relating to "digital health"; (2) enhance the

414 "European Commission's ethics guidelines for trustworthy AI" <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

415 UK's independent review on artificial intelligence and public standards https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/868284/Web_Version_AI_and_Public_Standards.PDF (Date of use: 20 June 2020).

416 OECD Principles on AI <http://www.oecd.org/going-digital/ai/principles/> (Date of use: 9 July 2019).

417 Notice 209 *Government Gazette* 42388 of 4 April 2019: Terms of reference for the presidential commission on the fourth industrial revolution <https://www.gov.za/documents/presidential-commission-fourth-industrial-revolution-members-and-terms-reference-9-apr> (Date of use: 3 July 2020).

418 International Telecommunication Union "Focus group on artificial intelligence for health" <https://www.itu.int/en/ITU-T/focusgroups/ai4h/Pages/default.aspx> (Date of use: 3 July 2020).

application of state policies on “digital health”; (3) toughen governance on digital health on international, provincial and domestic levels; and (4) promote people-focused health systems through “digital health”.⁴¹⁹

Politicians from various nations (such as the UK,⁴²⁰ US,⁴²¹ and the EU⁴²²) have already begun discussing the issue of regulation, and are organising their efforts through national policies and strategies. In a 2019 study conducted by the Law Library of Congress, the majority of surveyed jurisdictions, including South Africa, regarded AI to be promising and expressed their intention to take the lead on AI.

A number of these nations have since then established, or have already started forming, domestic AI or digital policies and action planning. Similarly, public support is increasing for such regulation. A 2019 survey of the Center for the Governance of AI revealed an overwhelming majority of 82 per cent of the US public believe that, AI technologies ought to be meticulously managed and recommends that, US citizens should prioritise most governance challenges relating to AI.⁴²³

Views over regulating AI technology have also been expressed by specialists and experts that are engaged in the field.⁴²⁴ A support for regulation can also be seen in the rise of voluntary projects within which technology corporations have

-
- 419 WHO “Data and Innovation: draft global strategy on digital health” https://apps.who.int/gb/ebwha/pdf_files/EB146/B146_26-en.pdf (Date of use: 3 July 2020).
- 420 UK Industrial Strategy White Paper <https://www.gov.uk/government/publications/industrial-strategy-building-a-britain-fit-for-the-future> (Date of use: 6 June 2020). The UK Industrial Strategy sets out AI and BD as areas where it can drive the international technological revolution.
- 421 White House Report 2016 “Preparing for the future of intelligence” https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf (Date of use: 11 May 2020). In 2020, the US introduced guidance for the regulation of AI to serve as a basis for future legislation, see <https://www.whitehouse.gov/wp-content/uploads/2020/01/Draft-OMB-Memo-on-Regulation-of-AI-1-7-19.pdf> (Date of use: 8 June 2020).
- 422 White Paper on Artificial Intelligence “A European approach to excellence and trust” https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf (Date of use: 6 June 2020).
- 423 Zhang B & Dafoe A “Artificial intelligence: American attitudes and trends (January 2019)” https://governanceai.github.io/US-Public-Opinion-Report-Jan-2019/us_public_opinion_report_jan_2019.pdf (Date of use: 6 June 2020).
- 424 Future of Life Institute “Benefits and risks of artificial intelligence” <https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/?cn-reloaded=1> (Date of use: 11 April 2020); Russell S “Provably beneficial artificial intelligence” <https://people.eecs.berkeley.edu/~russell/papers/russell-bbvabook17-pbai.pdf> (Date of use: 6 June 2020).

promulgated codes of good practice. OpenAI, founded by Elon Musk, has released a charter to guide the company in advancing the interests of humankind during the course of AI's development and to endeavour building safe and beneficial AI.⁴²⁵

There is also growing sense among legal scholars and practitioners on the importance for discussion over juxtaposition between technology on the one end, and regulation on the other end. The various journals published on the connection between regulation and technology bear proof to this.⁴²⁶ In most discussions on the regulation of technology, the problem is associated with ensuing ecological-, physical- or social problems, which emanate from technology related applications and practices.⁴²⁷ As this thesis accepts that regulation will entail a top-down approach (where the regulatory government or agency prescribes mandatory rules for the regulation of AI) then "regulation" would be seen as the method used by government to lessen or prevent such problems.

There appears to be an explicit attempt at self-regulation through private partnership initiatives such as "Partnership on AI" formed by a number of NGOs, technology industry giants and academic institutions, in order to research and create best practices for AI. However, calls for a modest regulation of AI (which sees the technology industry playing a principal role on the matter) are part of a failed regulatory philosophy as they fall short of institutional policies for impactful public engagement and involvement towards the development of appropriate guidelines.

The public has long suspected that some companies may act with no transparency or accountability and may be primarily interested in promoting their own objectives, without dealing with harmful consequences.⁴²⁸ As has been demonstrated in Chapter two, AI offers important socio-economic advances. Therefore, it is too

425 "OpenAI Charter" <https://openai.com/charter/> (Date of use: 6 June 2020).

426 Moses 2013 *Law, innovation and technology* 1; Scherer 2016 *Harvard journal of law and technology*; Willick 1983 *The AI Magazine*; Brownsword Rights, regulation and the technological revolution; Russell SJ and Norvig P *Artificial intelligence: A modern approach* (Pearson Prentice Hall New Jersey 1995).

427 Moses 2013 *Law, innovation and technology* 5.

428 Partnership on AI <https://www.partnershiponai.org/#> (Date of use: 6 June 2020).

important to be regulated by way of a reactive approach, waiting on obstacles to emerge and then looking for solutions after the fact.

Regulation is inevitable and government involvement would balance both individual and societal interests, offering a more satisfactory development of AI. Whilst there is a general unanimity on the necessity to consider regulation, it still begs the question whether good regulation for AI is possible and how we would ensure that it will be successfully enforced?

3.2.7. Conclusion

Regulation has proven to be an effective means to constitute and sustain large-scale societies. It is used to advance certain social goals, such as affordable and quality healthcare, to prevent public harm, and to ensure legal certainty. The concept of policing power to protect and promote public welfare has continued to expand to incorporate a large variety of regulatory purposes, from the improvement of individuals' lives, the safeguarding of basic values (such as equality and social justice), and the general progression of humanity.

The success of digital transformation for healthcare calls upon all stakeholders in healthcare to reach a common ground in working together, guided by a coherent regulatory framework. This, will protect social norms and mitigate harm in relation to a developing socio-technical environment, as opposed to merely questioning how AI should be regulated. Without regulatory measures in place to protect providers and patients, neither will ever benefit from the increased accuracy and effectiveness that, medical smart machines can offer in their diagnosis and treatments.

3.3. *Risks and difficulties of regulating AI*

3.3.1. The definition issue

AI lacks a concise and universally accepted definition which complicates efforts to develop an appropriate governance infrastructure. For example, Stuart Russell and Peter Norvig have identified eight separate meanings of AI, organised in relation to four categories, benchmarking them against the internal processes of human

intelligence.⁴²⁹ However, Alan Turing has focused on a machine's external manifestation of intelligence, or analytical ability, where through a method of inquiry, a computer's achievement at "thinking" could be determined, as to establish if it can mimic human responses under specific conditions and if there is a likelihood of it being mistaken as the human.⁴³⁰

Definitions of "intelligence" differ, often relating it to human characteristics, which are inherently complicated on their own, notably including an understanding of awareness, linguistics, learning, an ability to conceptualise, an ability to acclimate, and an ability to comprehend.⁴³¹

Whilst these elements may render human intelligence a logical preference for benchmarking AI's progression today, human ability is no contest when it comes to the organic and artificial domain, as technologies which surpass human aptitude and intelligence are already in existence, insofar as versatility and speed is concerned.⁴³² The problem is further exacerbated when stringent definitions of intelligence give rise to philosophical debates about what it means to be human, since in the western tradition, so much of what constitutes the person turns on our cognitive abilities.⁴³³ AI therefore becomes difficult to define because human intelligence, on a multi-dimensional spectrum, is hard to define in the first instance.

Another major problem in defining AI is that, as the technology continues to develop the goalposts in pursuit of a workable definition keep moving. As machines become capable of resolving a problem which was only possible within the realm of human intelligence, there is a propensity to redefine what intelligence means in relation to AI.⁴³⁴

Thus far, the assertion has been that, machines can be designed to "act as if" they are smart, known as the "weak" or "narrow" hypothesis for AI. We are now looking towards a "strong" hypothesis, where technologies are thinking and are not simply

429 Russell and Norvig *Artificial intelligence* 3rd ed 2.

430 Turing AM "Computing machinery and intelligence" 1950 *Mind* 433.

431 Scherer 2016 *Harvard journal of law and technology* 360.

432 Saxenian *et al* *Artificial intelligence and life in 2030* 13.

433 Chinen *Autonomous* 4.

434 Smith C *et al* "The history of artificial intelligence" 2006 *University of Washington* 4-5;

“simulating” thinking or “imitating” humans.⁴³⁵ Consequently, many of the rules and logic-based systems which were previously considered AI are no longer classified as such, since these systems are designed in a manner which permits them to execute autonomous decisions, without any human intervention.

Most AI researchers do not focus on the distinction between “weak” and “strong” AI hypotheses, because as long as their applications work, they are not concerned with labelling it as a “simulation of intelligence” or “actual intelligence”, nor in establishing how human reasoning works.⁴³⁶

As submitted in the introduction to Chapter one, in order to overcome the difficulties in defining AI for the purposes of this thesis, I approach the problem by adopting the scientific definition of machine intelligence as “the study and design of intelligent agents”,⁴³⁷ that can adapt to various circumstances previously unknown and learning through familiarity, reaching a goal and objective not conceivable to conventional computers.⁴³⁸ This is an expansive definition as it encompasses new technological improvements and accommodates the strong AI hypothesis, allowing for the definition of AI to which is flexible, and can shift according to the goals that the AI system is designed to reach.

Regulators are likely to consider a flexible definition of AI, as it reaches more programs and devices as they are likely to have a broad legislative mandate, either intentionally or as interpreted by the regulators themselves.⁴³⁹ This also allows for the notion “form follows function” attributed to product design. This was coined by

-
- 436 Russell and Norvig *Artificial intelligence* 3rd ed 1020.
Russell and Norvig *Artificial intelligence* 3rd ed 1020. Marr B “The key definitions of artificial intelligence (AI) that explain its importance”
<https://www.forbes.com/sites/bernardmarr/2018/02/14/the-key-definitions-of-artificial-intelligence-ai-that-explain-its-importance/#fe0a3744f5d8> (Date of use: 10 June 2020).
- 437 Poole, Mackworth and Goebel *Computational intelligence* 1; Kok *et al Artificial intelligence* 1095-1096.
- 438 Scherer 2016 *Harvard journal of law and technology* 361. According to Scherer, “it appears that the most widely-used current approaches to defining AI focus on the concept of machines that work to achieve goals.” Also see Hutter and Legg 2007 *Minds and Machines* 391, 405-423, where the authors submit “intelligence measures an agent’s ability to achieve goals in a wide range of environments.”
- 439 Weaver JF “Regulation of artificial intelligence in the United States” in Woodrow Barfield W and Pagallo U *Research handbook on the law of artificial intelligence* (Edward Elgar Publishing 2019) 155-213.

Louis Henry Sullivan,⁴⁴⁰ where AI should primarily relate to its intended function or purpose assisted by its form. It would make more sense to regulate AI's function instead of the AI devices (form) themselves, because if the architecture dictates how the devices operate then the technology is likely to be stifled by inflexibility, unable to acclimate to change.

3.3.2. Understanding AI and its impacts

Apart from the fact that AI is known to be challenging to define, regulators are still required to have a general understanding as to how the technology is created, operates, and its impacts on society, as there is a risk that the language used and assumptions built into the regulation, may render the regulation premature and ineffective.⁴⁴¹

Policy and technology law experts are of the view that, developing new regulation for emerging technologies is an art form, and that an understanding the law and its reasoning is vital, but not enough.⁴⁴² Healthcare regulators often struggle to keep-up with the pace in healthcare technology innovation.⁴⁴³ Attempts to regulate in the absence of appropriate insight as to the way AI functions in real-world contexts, or with absolutely no understanding of the technology, may lead to disasters.⁴⁴⁴

440 Stinson L "Remembering the legend behind 'form follows function'" <https://www.wired.com/2015/09/man-coined-form-follows-function-born-today/> (Date of use: 20 July 2020).

441 Yudkowsky *Artificial intelligence as a positive and negative factor in global risk* 1-28; Scherer 2016 *Harvard journal of law and technology* 355-363; Etzioni A and Etzioni O "Why regulating AI is a mistake" <https://www.forbes.com/sites/ciocentral/2017/01/09/why-regulating-ai-is-a-mistake/#201e8c9f2be3> (Date of use: 24 June 2020); Strous L "Should artificial intelligence be more regulated?" in Strous L and Cerf VG (eds) *Internet of Things. Information processing in an increasingly connected world* (Springer International Publishing Cham 2019) 28.

442 Lawyer Monthly "Where artificial intelligence will take legislation" <https://www.lawyer-monthly.com/2017/04/why-regulators-should-take-the-drivers-seat-of-autonomous-vehicles/> (Date of access: 27 June 2020).

443 Raths D "Digital health dilemma: Regulators struggle to keep pace with health-care technology innovation" <https://www.govtech.com/health/Digital-Health-Dilemma-Regulators-Struggle-to-Keep-Pace-with-Health-Care-Technology-Innovation.html> (Date of use: 29 June 2020).

444 Smialowski B "Should artificial Intelligence be regulated?" <https://www.forbes.com/sites/quora/2017/08/31/should-artificial-intelligence-be-regulated/#26674c3d331d> (Date of use: 24 June 2020).

A misunderstanding about what AI is (and is not), could hinder its capabilities if regulated and deny society the benefits of these technologies. Since the time Peter Szolovits predicted the benefits of using AI in medicines (in 1982), and five decades since healthcare facilities first utilised computers, impediments to the adoption of health technologies still remain due to barriers (such as the health practitioners' lack of understanding the technology and the significant implementation costs).⁴⁴⁵

As demonstrated in Chapter two, AI is a technology which is already being used and commercialised. It offers immense societal benefits in various areas, especially in healthcare settings. It can no longer be thought of solely as a “rudimental field” of research and development, nor as a scientific discipline (such as quantum mechanics, nanotechnology, biochemistry or nuclear energy). There is already research undertaken regarding the implications of AI for healthcare and its outcomes are developing to a stage where it can be more understood.⁴⁴⁶

These studies have also cautioned that, the application of AI systems gives rise to various ethical- and legal challenges of which the consequences must be understood. Some AI applications are in fields which have been, and would already be, regulated – healthcare being an example.⁴⁴⁷ However, existing laws and regulations in South Africa may not be sufficient to deal with AI in its full extent. The existing legal framework, which relates to AI, needs to be carefully calibrated and enhanced, in order to prevent unfair practices and harm to patients, whenever AI tools are deployed by healthcare practitioners.

As with other areas outside of the AI world, any regulation ought to encompass a risk-based approach. Regulation is not required in the event where the risk is

445 Khan *Medicine* 61-62.

446 Saxenian *et al* Artificial intelligence and life in 2030 25-30; Peek N *et al* “Thirty years of artificial intelligence in medicine (AIME) conferences: A review of research themes” 2015 *Artificial intelligence in medicine* 61; Buch VH, Ahmed I and Maruthappu M “Artificial intelligence in medicine: current trends and future possibilities” 2018 *British journal of general practice* 143-144.

447 The relevant regulatory authorities in South Africa include the NDoH, the HPCSA, the South African Nursing Council, the Allied Health Professions Council, the South African Pharmacy Council and the South African Dental Technicians Council. The Health Professions Act 56 of 1974 (HPA), National Health Act 61 of 2003 and Medical Schemes Act 131 of 1998 are the main legislation governing healthcare in South Africa.

minimal and if the risk is considered to be high, regulation will be necessary.⁴⁴⁸ Taking into account the current knowledge on some of the effects of AI and the known ethical and legal challenges posed by AI's implementation, this thesis argues that, it is currently possible to enforce the regulation of AI in a sensible and impactful manner, rather than adopting the reactive approach to regulation (where the consequence of waiting may be severe, rendering *Ex post facto* regulation inconsequential and ineffective).

3.3.3. AI and the existing regulations

Whilst there may in principle be an argument for the regulation of some areas of AI, it could prove to be challenging if we cannot exactly identify what to regulate and how to regulate it. The difficulty is that, the construction process for AI is already complex and any new regulation would require justification for its existence, if it is to be successful. In the development of algorithms for diagnostic support in healthcare, dimensions; recommendations; and authentication must be meticulously defined in cooperation with medical professionals.⁴⁴⁹

Individuals as well as entities (ranging from individual designers, corporations, health care professionals, contractors, and all their employees) are involved in the development and production of the firmware, software program, and the computer hardware that render autonomous machines and systems possible.⁴⁵⁰

AI cannot be regarded as a uniform system as it is integrated with additional systems as is the case in mobile health (which constitutes the use of wireless devices, i.e., smartphones, tablets, watches and bracelets) in order to improve health. AI is designed upon systems which already exist, complicating the management of the advancement of these systems and the control of its use.⁴⁵¹

AI systems are multifaceted in their functional reasoning, producing outcomes which are difficult to anticipate. Only some AI types of systems allow for its underlying

448 Strous Should artificial intelligence 28.

449 Felländer-Tsai L "AI ethics, accountability, and sustainability: revisiting the Hippocratic Oath" 2020 *Acta Orthopaedica* 1.

450 Chinen *Autonomous* 11.

451 Strous Should artificial intelligence 31.

reasoning to be tracked and comprehended. On the other hand, there are some which do not, on account of the functional aspects of complicated and interactive socio-technological systems (especially those that operate on neural networks). A commonly recognised example of this is that of computerised chess systems. In reviewing a computerised chess system, mathematician Nate Silver notes:

We should probably not describe the computer as “creative” for finding the moves; instead, it did so more through the brute force of its calculation speed. But it also had another advantage: it did not let its hang-ups about the right way to play chess get in the way of identifying the right move in those particular circumstances. For a human player, this would have required the creativity and confidence to see beyond the conventional thinking.⁴⁵²

According to Scherer, this demonstrates an important distinction between the reasoning methods of humans and that of contemporary AI, where these differences can result in AI technology that can produce outcomes which a human would not foresee.⁴⁵³ Both the complexity and opacity of AI systems give rise to issues of transparency if the end-user is unable to comprehend or decipher its description as a product or its instructions. This could lead to unintended consequences, especially if they function in unforeseen ways that cause injury or harm, or violate basic rights.⁴⁵⁴

It is precisely due to the most obvious feature of AI (its autonomous reasoning and problem solving) where most legal problems emanate due to the assigning of blame where the system causes injury or harm, separating it from earlier technologies. Law in its authoritative and expressive facets is called upon to foresee and address these technological shifts.

The development relating to hardware, software, and infrastructure (which makes AI possible) are already in place and the complex ecosystem of individuals and entities which participate in the design and production of the sophisticated technology is part of the broader question: who must be held accountable due to harms generated by AI?⁴⁵⁵ In the treatment of patients, the answer will be contingent

452 Silver N “Rage against the machines” <https://fivethirtyeight.com/features/rage-against-the-machines/> (Date of use: 25 June 2020).

453 Scherer 2016 *Harvard journal of law and technology* 364.

454 Felländer-Tsai 2020 *Acta Orthopaedica* 1.

455 Chinen *Autonomous* 21.

upon the legal relationships in place (the persons involved in the treatment of patients personally, and the relationships in place between them and the harmed patient) as well as the differences in output generated by AI system.

Where the AI systems impacts on a functional outcome, a causal connection could generally be established between a defective system and a resultant injury or harm. However, in instances where the system renders an intellectual outcome alone, the involvement of a human end-user can result in breaking the causal connection between a faulty system and a resultant injury or harm. Therefore, this necessitates changes to pre-existing legal concepts and doctrines, particularly the existing theories of liability.

Whilst the objective of AI scientists is to create and deploy strong or general AI (able to execute assignments and provide solutions, ordinarily only undertaken by humans),⁴⁵⁶ present-day AI systems have constraints and shortcomings and is not capable of entirely substituting doctors in the discovery and healing of health issues. The type of care potentially recommended by an AI system is restricted to the individual and personal scenarios of medical cases. Hence, due regard should be paid to the fact that, at present AI systems can only apply the principles and information as programmed by the software- and hardware developers, and scientists.

Even though AI's creators will not be in a position to anticipate how the technology will behave post-design, the capability to produce unforeseen outcomes is an intentional construct by AI's designers despite the fact that a certain outcome was not anticipated.⁴⁵⁷ This poses a challenge as to how the law must change in order to allocate rules concerning legal responsibility where AI is involved. It is required to protect victims of AI-caused harm, as well as to assess how future machines will need to be designed. Effective regulation in a midst of technological change is also about seeking new solutions, in order to protect settled principles.

456 Future of Life Institute "Benefits and risks of artificial intelligence"
<https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/?cn-reloaded=1>
(Date of use: 11 April 2020).

457 Scherer 2016 *Harvard journal of law and technology* 366.

3.3.4. Disruptive sector landscape

As products and services evolve due to technological changes, they also tend to shift from (or cut across) one regulatory landscape to another.⁴⁵⁸ For example, Uber (a former exclusive passenger services company) has recently ventured into the food delivery business, which now brings their activities within the ambit of health regulators. If Uber were to use self-driving vehicles to transport its passengers, it may be subject to the scrutiny of the telecommunications regulations and watchdogs.

The challenge for regulators with evolving and interconnected technologies is in keeping regulations consistent. This could cut across different sectors, making it particularly difficult to allocate legal responsibility for consumer injury or harm. In this case, the ubiquitous properties which render AI technologies as appealing may also lead to a disruptive regulatory landscape. Therefore, coordination with regulators across the different sectors becomes important.

3.3.5. Technological neutrality

There is an idea that, when it comes to the regulation of technology it should be technology-neutral. This is articulated in various policy papers and legal instruments.⁴⁵⁹ For example, in its general policy document on ICT regulation, the Dutch policy memorandum “Legislation for the electronic highways of 1998” formulates it as follows:

Technology-independent legislation is to be preferred. This usually establishes an equality between the ‘off-line world’ and the ‘on-line world’. Also, technology-independent legislation can better withstand technological turbulence. However, sometimes technology-dependency will be called for instead. For instance, the need for legal certainty could be a reason for technology-dependent legislation.⁴⁶⁰

458 Eggers WD and Turley M “The future of regulation: Principles for regulating emerging technologies” <https://www2.deloitte.com/content/dam/Deloitte/lu/Documents/risk/lu-future-of-regulation.pdf> (Date of use: 29 June 2020).

459 Koops *Should ICT* 6.

460 Koops *Should ICT* 1.

Whereas the Dutch policies focused on the formulation of the rules, the UK government's e-Principles (at the time) stressed the effects of regulation:

Regulation should be technology neutral in its effects. The effects of the offline and on-line regulatory environments, including the criminal and civil law, should be as similar as possible. There may be occasions when different treatment is necessary to realise an equivalent result.⁴⁶¹

Scholars, notably Maxwell and Bourreau,⁴⁶² later described the concept of “technological neutrality” as having more than a few meanings:

“Meaning one: technology neutrality means that technical standards designed to limit negative externalities (e.g. radio interference, pollution, safety) should describe the result to be achieved, but should leave companies free to adopt whatever technology is most appropriate to achieve the result.

Meaning two: technology neutrality means that the same regulatory principles should apply regardless of the technology used. Regulations should not be drafted in technological silos;

Meaning three: technology neutrality means that regulators should refrain from using regulations as a means to push the market toward a particular structure that the regulators consider optimal. In a highly dynamic market, regulators should not try to pick technological winners”.⁴⁶³

In terms of both the earlier and later meanings ascribed to the notion of “technological neutrality”, it is the idea that regulation should be technologically neutral, and the regulatory standards should be consistent, irrespective of the technology applied, save for some exceptions. Emerging technologies are to be regulated at the onset by broader regulatory frameworks.⁴⁶⁴ Any technology-specific legislation will not accommodate emerging technology and this will require the legislation to be modified sooner.⁴⁶⁵

The traditional doctrine of legislation is that, it should be transparent. Thus, the notion that regulation should be technology-neutral may be a means to achieve the goal of transparency and legal certainty.⁴⁶⁶ Certain scholars who defend the theory of technological-neutrality believe that, market related standards are desirable where the benefits associated to the technology are exceedingly uncertain and that,

461 Koops *Should ICT* 2.

462 Maxwell WJ and Bourreau M “Technology neutrality in internet, telecoms and data protection regulation” 2014 *Computer and telecommunications law review*.

463 Maxwell and Bourreau 2014 *Computer and telecommunications law review* 1.

464 Moses 2013 *Law, innovation and technology* 8.

465 Koops *Should ICT* 10.

466 Koops *Should ICT* 11.

a comparable inference could be applied in relation to technological-neutrality where if there is an augmented degree of uncertainty associated to technological development, the more it calls for the necessity to render regulation technologically-neutral.⁴⁶⁷

When it comes to the regulation of technology, two different ideas generally arise. One idea stresses that, regulation in its broad meaning should be technology-neutral. The emphasis on this idea seems to focus on the “effect” of regulation. The other idea emphasises that, legislation or legal rules should be technology-neutral and tends to focus on the “formulation” or “wording” of regulation.⁴⁶⁸ Regulations considered to be technologically-neutral allow regulators with the room to focus on the performance standards and impacts of such technology (instead of the technology itself), and at the same time, as the technology advances the rules can be easily adapted to same.⁴⁶⁹

On the other hand, the principle of neutrality may also have certain challenges. Certain areas of law (focussing on procedural issues) may carry the risk that, over the years the interpretation of the law will differ for different sectors and will lead to unintended technology specificity.⁴⁷⁰ Where the purposes and ensuing impacts of specific technology are distinctive (and tend to be distinguishable from systems) technology-specific norms would be ideal. This is evident, for example, in the healthcare industry (where the equipment and the technology the industry uses is of immense value and consequence). Although healthcare law is technology-neutral in theory, it is technology-specific in its application.

A more important risk with eminently neutral laws is that, in order to achieve sustainability it must be uninformed of how the existing technology works, or it must operate as if such technology does not exist at all. Therefore, the formulated laws are so technology-neutral that they become meaningless and redundant.⁴⁷¹ The notion of technological neutrality creates a contradictory element. In turn, this calls

467 Maxwell and Bourreau 2014 *Computer and telecommunications law review* 4.

468 Koops *Should ICT* 4.

469 Maxwell and Bourreau 2014 *Computer and telecommunications law review* 5

470 Koops *Should ICT* 10.

471 Koops *Should ICT* 10-11.

for producing technology-neutral laws whilst also advocating that laws must be altered as and when technologies progress.

Advancements in today's society is much greater than what it was. This creates a necessity to acclimatise our needs of sustainability according to the pace at which the technological landscape is moving.⁴⁷² For the first time, we are dealing with computers which possess super-intelligence, with the ability to learn and adapt through their own design – surpassing the abilities of their designers. Therefore, when it comes to the regulation of AI, the notion that regulation should be technology-neutral should be superseded with the traditional law-making principle that, legislation should be sustainable so that it does not lag behind new technological developments.

The extent of the sustainability called for may be considered “less” in terms of years than is expected for non-technological areas of law, due to the technology behind AI being better defined than its consequences.⁴⁷³

3.4. *Impact of regulation on AI*

As indicated earlier, the medical field as a whole has been slow in its adoption of basic technological advances due to a lack of understanding of the technology. Some contend that, AI technology is not at the point of maturity and is still in its infancy, and will continue to encounter many changes and improvements over the years.⁴⁷⁴ Therefore, attempts to regulate AI could be misguided as lawmakers would likely encounter challenges as to create a regulatory framework based on technology, which may still precipitously change, in both construction and utilisation.

Collingridge notes a concern that regulators face in dealing with new technology, known as the “double-bind” dilemma.⁴⁷⁵ In the beginning stages of a technology's design, regulation was difficult as there was limited information about the technology's likely effect (which could not be projected or quantified) until extensive

472 Koops *Should ICT* 11.

473 Koops *Should ICT* 11.

474 Saxenian *et al* *Artificial intelligence and life in 2030* 48.

475 Collingridge D *The social control of technology* (New York: St. Martin's Press 1980).

development and wider use of the technology occurs. At a subsequent stage, by the time the technology is implemented, regulation proves to be challenging due to the technology becoming more ingrained, rendering any modifications required by regulators as complex, time-consuming, and expensive to implement.

Collingridge's concerns are intended to raise the awareness that, as new technological systems progress further and become more complicated, they would become more divisive to regulation. This implies that, lawmakers (in attempting to control the technological landscape) should not make decisions in ignorance. It is necessary to act early whilst there is still minimal disruption, and to take minor steps and have an exit strategy for new technologies.

Committing entirely to the AI technological revolution is not a decision which can easily be undone. In adopting a balanced AI regulatory framework and usage portfolio (as some countries are attempting to do) it provides an exit strategy from any component of the technology.

There are also fears that regulation may create an unfair competitive advantage for those nations that do not impose regulation, allowing them to overtake those nations that do impose it, just as AI is starting to make progress in various fields.⁴⁷⁶ Jurisdictions proposing regulation which is considered to be too hostile may delay or restrict advancements in AI and will likely be disadvantaged from an economic perspective.⁴⁷⁷ For example, the GDPR covers specific aspects which could be argued to restrict the growth and application of AI for the market in Europe.

This, in turn, is expected to impact the EU's fiscus for the worse, as international participants will be deterred from marketing their AI inventions or related services in the EU territory. This prevents individuals or businesses in the EU from accessing the beneficial products or services which are accessible to other markets, ultimately rendering the AI-market in the EU far less competitive and pioneering.

476 Etzioni and Etzioni <https://www.forbes.com/sites/ciocentral/2017/01/09/why-regulating-ai-is-a-mistake/#201e8c9f2be3> (Date of use: 24 June 2020).

477 Health Capital Consultants "Artificial Intelligence in Healthcare – Competition" https://www.healthcapital.com/hcc/newsletter/07_17/HTML/AI/10.7_formatted_esp_hc_topics_ai_competition_draft_cite_check_7.28.php (Date of use: 25 June 2020).

Article 22 of the GDPR potentially restricts the use of AI as it places an obligation on industries to have human oversight for certain algorithmic decision-making. Such a restriction not only intensifies labour costs but also deters the use of AI, because a major motivation for developing AI is to automate tasks, which are normally slower, more expensive, and tougher to accomplish when they are undertaken by humans.

There is argument for the fact that, a delayed regulation has promoted competition in the case of the healthcare industry, as rigorous regulation of treatments emanating from the new technology can likely disrupt conventional research and development and the AI market, which may be a key driver of quality enhancement and the fundamental motivator for an organisation's aptitude to be competitive.⁴⁷⁸

Some authors argue that, restrictive and harsher regulations for AI technologies could sway the growth of AI from lawful to unlawful actors, spurring what is known as a black market, thus increasing the risks inflicted by such technologies and the costs during the development and deployment of AI.⁴⁷⁹ Illegitimate actors would ultimately promote their private agendas, to the detriment of society.

One only has to consider the case of the tobacco ban imposed by the South African government in regulations imposed under the Disaster Management Act 57 of 2002, prohibiting the sale of tobacco and tobacco products during the COVID-19 lockdown. This led to a flooding of the market with a number of syndicates and illicit sales of cigarettes, at exorbitant costs for the consumer and the fiscus.⁴⁸⁰

In essence, the introduction of control mechanisms to the application of AI systems can be catastrophic for researchers, industries, entrepreneurs, users, and for the public at large. A regulation could delay advancements in the area of healthcare,

478 Health Capital Consultants "Artificial Intelligence in Healthcare – Competition" https://www.healthcapital.com/hcc/newsletter/07_17/HTML/AI/10.7_formatted_esp_hc_topics_ai_competition_draft_cite_check_7.28.php (Date of use: 25 June 2020).

479 Miron JA and Zwiebel J "The economic case against drug prohibition" 1995 *The Journal of economic perspectives* 189.

480 Haffajee F "Dlamini Zuma turns cigarettes into illicit drugs as the underground economy takes over" <https://www.dailymaverick.co.za/article/2020-06-12-dlamini-zuma-turns-cigarettes-into-illicit-drugs-as-the-underground-economy-takes-over/> (Date of use: 29 June 2020). Also see *Fair-Trade Independent Tobacco Association v President of the Republic of South Africa and Another* 2020 (6) SA 513 (GP).

which is globally considered as an important social development goal.⁴⁸¹ It could thwart the deployment of medical results relying on AI, which can be life threatening, decrease the pace of healthcare delivery, hinder competition and lead to number of financial disruptions in the sector.

From the other perspective, AI technology stands to benefit from decisive and well-timed regulation. According to Daniel Carpenter, some early regulatory interventions might even become “market-constituting”, by facilitating a vigorous and dynamic market that, would otherwise not be attainable, particularly with regard to “credence goods”, which are not easy for consumers to assess. Carpenter explains:

[D]rugs are types of credence goods, whose quality consumers can assess neither through inspection (as for “inspection goods” like a tomato) nor experience (as for experience goods like a job). Such goods, social scientists have demonstrated both theoretically and empirically, create “lemons problems.” On account of information inadequacies, consumers will unknowingly procure or use substandard products whilst better options are accessible.⁴⁸²

Thus, for consumers, the establishment of more specific regulations for AI would enhance understanding and promote acceptance of the technology, which would foster a robust market. A regulatory system would educate people and make them more acquainted with AI technology, augment the interest for other products, which complement the technology, and generates more demand and support for the technology. Similarly, the enactment of safety standards, clear accountability laws, and greater collaborative research can generate safer products and services, can enhance the quality of service delivery, and augment consumer belief in the technology. Defining the new technology in terms of its effects and structuring rules to deal with those effects, allows consumers to overcome the fear that historically defines new technology.

For scientists and companies, a regulation would foster a legitimate exchange network and engagement amongst these role-players, one that can lead to the

481 Sallstrom, Morris and Mehta https://www.orfonline.org/wp-content/uploads/2019/09/ORF_Issue_Brief_312_AI-Health-Africa.pdf (Date of use: 19 May 2020).

482 Carpenter D “Confidence games: How does regulation constitute markets?” in Balleisen EJ and Moss DA *Government and markets: toward a new theory of regulation* (Cambridge University Press New York 2009) 174.

creation of different levels of regulation representing the interests of these stakeholders, such as legislation that prescribes the basic guidelines and rules, and a technical and scientific grounded oversight agency to handle more multifaceted and capricious nuances. Since organisations have not yet found a way on how to deal with the undercurrents of “bad” AI, they could instead become a participant to the dialogue on how it should be controlled.⁴⁸³

Corporations are already collaborating to join the discussions, which includes, *inter alia*, “The partnership on AI to benefit people and society”, established by AI scientists from well-known technology companies, namely IBM, Amazon, Apple, Microsoft, Google and Facebook. This serves as an open channel to engage and to create best practices for the use of AI technologies. Regulation would also encourage a culture amongst corporations to manage AI from the viewpoint that, they are individually responsible for it, encouraging investment in appropriate methods and applications in respect of the technology, including structures for fair, understandable, and responsible AI.⁴⁸⁴ For scientists, a proper regulation by countries could encourage greater investment confidence in research for the scientific field, steering investments and funding towards accountable AI development.

For society, regulation is essential for rendering AI not only to be highly beneficial, but to be safer for deployment and use. The public concern about AI is real, as it has already shown itself to be capable of bias and a threat to data privacy. Regulation can call for mandatory transparency and accountability, in order to empower civil society with more credible sources in courts.⁴⁸⁵ Given that AI is still in its early development, but is rapidly being assimilated into every aspect that influences our daily lives, it is important to have unanimity on the standards that should be developed. Regulation can enhance prevailing risk evaluation and

483 Howard A “The regulation of AI — should organisations be worried?”
<https://sloanreview.mit.edu/article/the-regulation-of-ai-should-organizations-be-worried/>
(Date of use: 29 June 2020).

484 Saxenian *et al* Artificial intelligence and life in 2030 48-49.

485 Saxenian *et al* Artificial intelligence and life in 2030 49.

mitigation on AI, and circumvent avoidable harm on account of functional or ethical calamities.⁴⁸⁶

Commercial and academic rivalry and competition in the AI sector is increasing and there are governments that are convinced that commercially driven innovation in its own will not be enough to sustain the competitiveness in AI.⁴⁸⁷ For this reason, many countries are beginning to focus on domestic strategies which will steer AI development, and simultaneously endeavouring to deal with regulatory issues on a primary-, theoretical-, and direct level. The government's role in every sector can be useful and important, even in the context of a market-focused financial system, such as the US and China, where competition between the states can sometimes lead to a race to the top, inclined toward optimal legal regulation.

For these abovementioned reasons, state regulation might prove flexible, responsive, and effective.⁴⁸⁸ This is especially valid for those sectors that are predisposed to modernisation and transformation, using AI technologies and further strictly regulated, as in the case of the healthcare sector. Consequently, the healthcare sector in particular would benefit greatly in directing special awareness and interest to government proclamations on laws and rules dealing with AI.⁴⁸⁹

The gap between innovations in the technology field and the methods employed to regulate such innovations are growing wider, known to be the "pacing problem."⁴⁹⁰ It means that, there is a separation between the velocity, advancements, and pervasive attributes of AI, and the traditional governing frameworks are not equipped to support the pace of the modern innovations. For the regulation of AI, the pacing problem has acquired greater urgency given that, the entire policy adoption cycle usually requires between five to twenty years, unlike a technology

486 Yudkowsky Artificial intelligence as a positive and negative factor in global risk 13-17.

487 Van Demark DC "The future of AI regulation: The government as regulator and research & development participant" <https://mcdermott-will-emery-2793.docs.contently.com/v/the-future-of-ai-regulation-the-government-as-regulator-and-research-development-participant> (Date of use: 29 June 2020).

488 Tutt 2017 *Administrative law review* 112.

489 Van Demark <https://mcdermott-will-emery-2793.docs.contently.com/v/the-future-of-ai-regulation-the-government-as-regulator-and-research-development-participant> (Date of use: 29 June 2020).

490 Eggers and Turley <https://www2.deloitte.com/content/dam/Deloitte/lu/Documents/risk/lu-future-of-regulation.pdf> (Date of use: 29 June 2020).

start-up that can grow into an enterprise with international reach in just a few months.⁴⁹¹

To deal with the challenges of emerging technologies, the regulators globally are re-evaluating their methods and are opting for models of regulation that tend to be flexible and collective, by relying on research and testing and a collaborative design of regulation through alignment and engagement with a wider constituent of stakeholders across the AI network.⁴⁹² The regulator's position is not exclusively that of a regulator: it has become more of a partnership with private companies, scientists and other organisations in providing reliable and innovative AI-driven technologies to people who need to have confidence in these technologies.

In conclusion, on the aspect of the regulation of AI, there are good reasons for it to be regulated. The shortcomings in the current legislation, the importance of protecting social norms and values, the balancing of individual and public interest, the need to manage advancements and the commercially driven market for AI in a responsible manner, are some of the motivators that call for its regulation. In the following Chapter, I discuss the basis for a good regulation within South Africa towards the application of AI technologies for the healthcare sector by considering existing laws and regulations that may create legal gaps. I also consider the international and national ethical guidelines and policies in order to identify gaps and guidance for the regulation of AI in healthcare, and to ascertain which norms and standards from these instruments could be useful when developing ethico-legal safeguards for South Africa.

491 Eggers and Turley <https://www2.deloitte.com/content/dam/Deloitte/lu/Documents/risk/lu-future-of-regulation.pdf> (Date of use: 29 June 2020).

492 Eggers and Turley <https://www2.deloitte.com/content/dam/Deloitte/lu/Documents/risk/lu-future-of-regulation.pdf> (Date of use: 29 June 2020).

4. CHAPTER FOUR: TOWARDS A GOOD REGULATION OF AI IN HEALTHCARE

4.1. *Criteria for determining if a good regulation is conceivable*

As demonstrated in Chapter three, compelling reasons exist to advocate for the regulation of AI. However, recognising when to regulate and how to do it successfully is just as important. Despite the absence of specific legislation in South Africa regulating AI, there are existing legal doctrines that can apply to its regulation. This Chapter thus considers how AI should be regulated in South Africa and analyses pertinent laws and regulations in South Africa, in order to determine if there are any gaps that should be examined when it comes to AI. It also analyses international and national ethical guidelines and policies, in order to identify gaps and guidance for AI in healthcare, and to ascertain which norms and standards from these documents could be useful when developing ethico-legal safeguards for South Africa.

At the heart of it is that, regulation should work, and appear to work, with due regard to the interest of the public.⁴⁹³ It is an instrument that can be applied by government towards achieving broad social goals, such as improvements towards a sustainable and safe environment, public health and safety, equitable and just business practices, racial and gender equality – all important national goals that societies entrust government with.⁴⁹⁴ Healthcare is an important social goal and a basic human need, but it is also a scarce and expensive resource and requires institutions to organise its allocation through controlled mechanisms.⁴⁹⁵

Regulation also prevents bad actors from damaging the reputation of their industries. In a perfect world, doctors would pledge to do no harm, as their objective would be to profit from activities that are consistent with the public good. However,

-
- 493 Thomadakis SB “What makes good regulation?” (IFAC Council Seminar, Mexico City, 14 November 2007)
https://www.ifac.org/system/files/downloads/30th_anniversary_Thomadakis_Pres_Nov_07.pdf (Date of use: 15 July 2020).
- 494 Beales H *et al* “Government regulation: The good, the bad, & the ugly”
<https://regproject.org/wp-content/uploads/RTP-Regulatory-Process-Working-Group-Paper.pdf> (Date of use: 14 July 2020).
- 495 Iszaid, Hafizan and Muhamad 2018 *International journal of public health and clinical sciences* 17.

there are many industries that fail to adhere to the “corporate” hippocratic version for their industries, acting to capitalise in furthering their own interests, often to the prejudice of the broader public, and thus they cannot be entrusted to regulate themselves.⁴⁹⁶ Considering the increasingly shorter innovation cycles in digital technologies in general, there is a need to ensure that essential regulation is adopted at the earliest stage, in order to prevent emerging technologies from exposing anyone to irreversible damage.⁴⁹⁷

In order for regulatory actions to succeed, the structure of the regulatory process must be sound. Inadequate or ineffective regulations can result in greater harm than its anticipated benefits, such as stifling innovation and economic growth.⁴⁹⁸ Whilst AI offers many benefits for the healthcare industry, a good regulation will support more innovative products and augment trust in the reliability of healthcare services in general. Therefore, it can be helpful to encourage policy makers and implementers to adhere, as far as possible, to certain criteria⁴⁹⁹ outlined below, as the basic test to assess whether regulation is good and fit for purpose.⁵⁰⁰

Necessity

Regulators should only intervene where necessary and justified and to realise certain goals. Regulation must be appropriate to the needs, perceived risks, and costs identified.⁵⁰¹ The gains that are anticipated from regulation should offset the

-
- 496 Beales *et al* <https://regproject.org/wp-content/uploads/RTP-Regulatory-Process-Working-Group-Paper.pdf> (Date of use: 14 July 2020).
- 497 Waltz A and Firth-Butterfield K “Implementing ethics into artificial intelligence: A contribution, from a legal perspective, to the development of an AI governance regime” 2019 *Duke law & technology review* 229.
- 498 UK Department for Business, Energy and Industrial Strategy “Better regulation framework guidance August 2018” https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/872342/better-regulation-guidance.pdf (Date of use: 14 July 2020); Beales *et al* <https://regproject.org/wp-content/uploads/RTP-Regulatory-Process-Working-Group-Paper.pdf> (Date of use: 14 July 2020).
- 499 Baldwin, Cave and Lodge *Understanding* 26-39.
- 500 Business Advocacy Network “Principles of good regulation: OECD and UK better regulation task force” <http://www.businessadvocacy.net/downloads/fsPrinciplesGoodRegulation.pdf> (Date of use: 14 July 2020); Thomadakis https://www.ifac.org/system/files/downloads/30th_anniversary_Thomadakis_Pres_Nov_07.pdf (Date of use: 15 July 2020); Baldwin, Cave and Lodge *Understanding* 26-39.
- 501 The BMJ “What does successful regulation look like?” <https://www.bmj.com/content/350/bmj.h1641> (Date of use: 14 July 2020); Beales *et al*

drawbacks of such regulation.⁵⁰² The healthcare sector has grown on the back of innovation and through AI, it is once more embarking on another major and crucial innovation step towards offering improved and inexpensive healthcare services.⁵⁰³ As demonstrated in Chapter two and three, there are substantial benefits derived from AI in healthcare and persuasive reasons for considering a regulation when it comes to AI.

For instance, in determining whether AI stands to be regulated, governments ought to assess if a substantive marketplace failure in healthcare is evident.⁵⁰⁴ The reason for this is that free and competitive markets not only promote the allocation of scarce resources to their most suitable purposes, but encourages entrepreneurial endeavours and innovation.⁵⁰⁵ An understanding of public concerns is also not exclusively limited to the demands of key investment markets, but it also calls for an appreciation of the concerns and challenges of a wider array of citizens.⁵⁰⁶

Lawfulness

There must be legitimate rules and principles as a legal basis, in order to balance democratic legitimacy and scientific innovation, so that all regulatory decisions rigorously respect the rule of law. Given the new and advanced technology that AI brings, it is important that regulators possess or consult experts in the field, in order to establish legitimacy in the decisions they make.

The regulators for AI should base their regulatory decisions on the most reliable technological and scientific data available. Regulation can offset the safeguarding

<https://regproject.org/wp-content/uploads/RTP-Regulatory-Process-Working-Group-Paper.pdf> (Date of use: 14 July 2020).

502 Beales *et al* <https://regproject.org/wp-content/uploads/RTP-Regulatory-Process-Working-Group-Paper.pdf> (Date of use: 14 July 2020).

503 Erixon F, Ferracane MF and Van der Marel E “The health of nations: A transatlantic trade and investment agenda for better healthcare” 2015 *European Centre for international political economy* 5.

504 The issue of market failure in healthcare was discussed in Chapter 3, as a reason for regulation.

505 Beales *et al* <https://regproject.org/wp-content/uploads/RTP-Regulatory-Process-Working-Group-Paper.pdf> (Date of use: 14 July 2020).

506 Thomadakis
https://www.ifac.org/system/files/downloads/30th_anniversary_Thomadakis_Pres_Nov_07.pdf (Date of use: 15 July 2020).

of ethical norms in a manner that is relative to the objective to be achieved. The objective of regulation is to augment, and not inhibit, public interests. Therefore, government must have due regard to key trade-offs and develop regulatory frameworks based on this. The healthcare industry is investing towards improving the rate of innovation for healthcare delivery, but regulations must facilitate a platform for pioneering technologies that will interact with patients in a protected and inexpensive way.⁵⁰⁷

Legal certainty

Regulation should be predictable and implemented fairly in order to give certainty to those being regulated. Placing reliance on a robust foundation, such as an acceptable definition of AI and established rules and criteria on AI's development and use, can render the regulation consistent for advancing legal certainty. New regulations should have due regard to other existing or proposed regulations, whether domestic or international in origin.

Transparency and accountability

There should be openness and transparency to the public in the development of the regulation to support evaluations, impact assessments, and decisions.⁵⁰⁸ For instance, in order to improve healthcare services through the application of AI systems, governments must be more open with regards to policies and goals for healthcare delivery, including proposals on products or services that will be procured.⁵⁰⁹ In this regard, it can at an early stage publicly disclose the essential data, prototypes, specifications and any other knowledge applied towards decision-making, and provide a platform for public participation. Such information can be easily accessible and inclusive, hence allowing public scrutiny.

507 Erixon, Ferracane and Van der Marel 2015 *European Centre for international political economy* 4.

508 EU Better Regulation Guidelines "Guidelines on Stakeholder Consultation" <https://ec.europa.eu/info/sites/info/files/better-regulation-guidelines-stakeholder-consultation.pdf> (Date of use: 14 July 2020).

509 Erixon, Ferracane and Van der Marel 2015 *European Centre for international political economy* 3.

Effectiveness or flexibility

Regulation can constantly be evaluated as the technology develops so as to guarantee that it is still suitably targeted, effective and necessary, or whether they should be modified or eliminated, and that it is grounded on a robust socio-political framework and regulatory standard.⁵¹⁰ Compliance and monitoring can involve the assistance of both public and private industries so as to counter the challenges of limited resources. Regulation must be fluid in order to acclimate to factual transformations that allows for AI innovation and to remain relevant and effective to AI's potential future development.

Targeting

The issues at hand must be countered by regulation and it should aim to diminish consequences, and regulators must adopt a goal-orientated approach to regulation. The design of a good regulation is possible,⁵¹¹ if all these fundamental elements are achieved in real-life settings or, if when impossible, there is an attempt to focus on some criteria that considers diverse and usually conflicting goals and purposes.⁵¹² The question that follows is: what would the required steps be to achieve good regulation of AI?

An inclusive framework regulating consumer affairs is required that will protect the interests of consumers when AI is applied in the healthcare industry in South Africa, as it presents critical practical issues and gives rise to uncertainties. This thesis focuses on the top-down approach of government regulation towards the healthy development of AI for healthcare. It is recommended that the government intervenes by issuing rules and standards through its legal and political authority, in order to assist in establishing the foundation towards an acceptable AI-driven culture.

510 Cath *et al* 2017 *Science and engineering ethics* 508.

511 Thomadakis
https://www.ifac.org/system/files/downloads/30th_anniversary_Thomadakis_Pres_Nov_07.pdf (Date of use: 15 July 2020).

512 Due to the rapid pace at which AI is developing, it may not be feasible to achieve transparency or accountability in all instances through accommodating a public participation for every decision because of the time it consumes, which will consequently slow down regulation which targets the innovations. The other criteria are considered as trade-offs such effectiveness and lawfulness.

The goal of rapid developments in AI should be socio-politically driven with regard to the direction it intends to go, as opposed to how fast it would reach such goal. There is a danger that, inadequate intervention may encourage private industry, academia, and researchers to seal the regulatory vacuum by unilaterally establishing the standards based on what they regard as good AI for the public, as lawmakers are currently not able or willing to deal with same. This remains less than ideal due to a lack of public and political concern and accountability on the part of private actors.⁵¹³

Effective regulation through government should embody a convergence of both horizontal and vertical regulatory approaches. A vertical regulatory approach occurs when government or regulatory agencies intervene by enacting rules and guidelines, which are binding and enforceable on other stakeholders. In the case of a horizontal regulatory approach, the government can organise its endeavours to regulate a sector through collaboration with other stakeholders.⁵¹⁴

This convergent of both the vertical and horizontal regulatory approach facilitates effective management and administration of investment and resources, as it accounts for the intricacies of different government structures. It also shadows investments and incentives in responsible AI research and innovation, up-skilling of a participative labour force, protection of basic human rights and inclusion of overlooked sectors of populations.⁵¹⁵

Governments across the world continue to be a significant source of funding for basic AI research.⁵¹⁶ A joint-stakeholder approach, which sees governments playing a principal role particularly in enforcement, could be the recommended method to drive AI-based innovation and its widespread dissemination in the right direction.

513 Cath *et al* 2017 *Science and engineering ethics* 507.

514 Baldwin, Cave and Lodge *Understanding* 103-311; Hudson B and Rosenbloom JD "Uncommon approaches to commons problems: Nested governance commons and climate change" 2013 *Hastings law journal* 1316.

515 Hudson and Rosenbloom 2013 *Hastings Law Journal* 1316.

516 Asay CD "Artificial Stupidity" 2020 *William and Mary law review* 1255.

This will ensure that it will have the most positive influence on the lawmakers, the regulated groups and those who stand to benefit from regulation.⁵¹⁷

4.2. A broad regulation

From a technical perspective, AI involves a mixture of different disciplines, such as: information technology; mathematics; computer science; and data science. When the technology is practically applied in applications, AI can be used in a number of fields, such as healthcare, which extends its involvement to expertise in the specific fields of application. However, the algorithmic nature of AI to sort, filter and arrange on its own, remains constant irrespective of its diverse applications and the same algorithm can be applied in various fields.⁵¹⁸ In these circumstances, it would be futile to regulate on a sector-by-sector approach, as a decentralised regulation could ultimately lead to different answers in respect of an identical problem, and many inconsistencies. This would render regulation time consuming and costly.

However, with the onslaught of a new application of algorithm, “ML”, the issue of a central regulation has become complex. These algorithms allow AI technologies the capability to autonomously learn, adapt, and improve from various data sets and knowledge, and are designed and coded to solve problems for themselves.⁵¹⁹ A company can thus develop and sell only an algorithm’s code and the same algorithm can thereafter be reproduced, altered, and commissioned in many applications that its original developer could not have anticipated.

To address the issue, it is recommended that regulation should consolidate rudimentary rules about AI on certain aspects, such as its design and accountability, whereas, in the application of each particular industry, specific laws can be enacted in order to regulate AI based on its specific idiosyncrasies. Therefore, at the very least with, regards to issues around the design, teaching, and application of AI technologies, a broad regulatory framework set by government, for the different sectors, would be the best approach. This could entail rules that allow for an AI system to be overruled by an approved human controller or to produce passwords,

517 Cath *et al* 2017 *Science and engineering ethics* 507.

518 Tutt 2017 *Administrative law review* 84.

519 Tutt 2017 *Administrative law review* 85.

codes or automatic switches that will stop a system from causing damage or assigning liability based on its development stage.

For algorithms that are programmed for application in specific cases, government can consider fitting AI in prevailing regulatory frameworks within a sector, in conjunction to the general rulings, to deal with the idiosyncratic challenges that AI brings to public.⁵²⁰ For the healthcare sector in South Africa, the different legislative instruments⁵²¹ regulated by the NDoH, should be targeted as to address the specific issues of AI applications in healthcare.

Thus, this thesis recognises that, a central set of rules is required concerning the design, advancement and accountability issue of AI. The application of AI for specific cases can be regulated for each sector, so as to address the unique concerns that stem from smart AI systems in areas such as healthcare. Similarly, specific data sets are to be used to teach AI systems that will be applied in specific industries, such as health data to be applied for medical AI systems.

Globally, the healthcare sector needs larger markets such as the US and EU, as well as connected nations, in order to facilitate rapid and inexpensive innovation that offers novel ways to treat and care for patients globally.⁵²² Thus, worldwide governance approaches for AI should, as best practice, consider the approach of other governance systems.⁵²³ Therefore, a diversity in participation is required across governments globally (or between certain continents) in order to gain an inclusive and coordinated regulation in those cases for shared regulatory jurisdiction over certain aspects or uses of AI.

520 Cath *et al* 2017 *Science and engineering ethics* 509.

521 The primary legislation is South Africa for healthcare is the National Healthcare Act 61 of 2003 which offers a health system framework for South Africa. However, there are various secondary legislation governing the healthcare profession, for instance the Health Professions Act 56 of 1974 to regulate the training, registration and practices of health Professions Act 56 of 1974 or the Medical Schemes Act 131 of 1998 that regulates Medical Schemes.

522 Erixon, Ferracane and Van der Marel 2015 *European Centre for international political economy* 2-3.

523 Mialhe N "AI & Global governance: Why we need an intergovernmental panel for artificial intelligence" <https://cpr.unu.edu/ai-global-governance-why-we-need-an-intergovernmental-panel-for-artificial-intelligence.html> (Date of use: 29 August 2020).

Divergences in regulations should, however, be limited where governments share the regulatory objective.⁵²⁴ For example, the GDPR is one of the early attempts to introduce data protection regulations for member countries in the EU and for those countries that operate in the EU jurisdiction. Similarly, African nations could create legislation for AI use with due regard to the unique social and ethical challenges of the African continent. This geographical harmonisation would also avoid the creation of incompatible legislations across the African states, whilst reducing the challenges of innovation as scientists are not burdened with having to comply with different rules across the states.

4.3. *Civil liability in delict arising from AI systems*

As previously proposed, in order for a good regulation to exist, there should be lawfulness and transparency to the public in the development of the regulation in order to support decisions made. When an AI-driven system is designed and automated to execute a useful task, it can develop a harmful technique towards attaining its objective, because it was not designed to fully align its goals with that of humans.⁵²⁵ Much like the way in which a skilled human specialist may make a mistake in delivering advice, an AI system could also err in the performance of its tasks.

Imagine a radiologist who depends on the output generated using a defective AI system and as a result causes an overexposure of radiation, or a doctor who relies on an AI program to screen a patient's medical progress, and if the system is defective and fails there is insufficient notice regarding the patient's actual medical status. In these instances, the question is: who is to be held accountable for damage or injury inflicted upon a patient when the technology malfunctions? Although these are only a few instances involving AI errors in the healthcare sector, liability for such errors has a considerable potential as the subject matter for litigation. To assign liability for AI systems when things go wrong, regulatory approaches should be

524 Erixon, Ferracane and Van der Marel 2015 *European Centre for international political economy* 4.

525 Future of Life Institute "Benefits and risks of artificial intelligence" <https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/?cn-reloaded=1> (Date of use: 11 April 2020).

implemented as to allow for satisfactory redress for the patient, which is also essential in fostering trust in the technology.

Civil liability on account of medical malpractice is regulated by delict law. A delict is defined to be a “civil wrong” and specifically as the “wrongful and culpable act or omission of a person which causes harm to another” resulting in liability for the transgressor.⁵²⁶ Such an act is either an act of intent or arises from negligence. As AI becomes more complex and as we continue to implement the technology for the healthcare industry, the main concern is whether our traditional legal doctrines of liability based on delict⁵²⁷ are sufficient to handle loss or injury to patients due to the use of AI. The primary aim is to compensate a person who is harmed, usually the patient, for injury or loss experienced due to an error emanating from the AI system.

In South Africa, the issue of an AI system error that causes injury in healthcare is yet to be adjudicated by the courts, and there is no case law that can offer clarity to the AI creators or end-users regarding the potential legal responsibility in relation to each of these participants. Specific legislation concerning this issue is also absent. In order to address the issue, this research proposes a standard for civil liability by presenting traditional notions of liability and how they correlate to AI. The discussion on civil liability arising from a defective AI system is limited to delict(s) relating to causing of harm through negligent conduct and strict liability.⁵²⁸ Delictual liability emanating from fraudulent behaviour or misrepresentation based on intent, is excluded for the purposes of this research, as such practices will be intermittent and will not occur among legitimate producers and users of AI.

526 Neethling J, Potgieter JM and Visser PJ *Law of delict* 7th ed (Lexis Nexis Butterworths Durban 2015) 4; Midgley JR and Van der Walt JC *Principles of delict* 3rd ed (Durban: Lexis Nexis Butterworths 2005) 2. *H v Fetal Assessment Centre* 2015 (2) SA 193 (CC); *Oppelt v Department of Health, Western Cape* 2016 (1) SA 325; *Goliath v The MEC for Health, Eastern Cape* 2015 (2) SA 97 (SCA); *Links v MEC Department of Health, Northern Province* 2016 (4) SA 414 (CC)).

527 Neethling, Potgieter and Visser *Law of delict* 4; Midgley and Van der Walt *Principles of delict* 2.

528 Alheit K “Delictual liability arising from the use of defective software: comparative notes on the positions of parties in English law and South African law” 2006 *The comparative and international law journal of Southern Africa* 266; Reed C and Angel J *Computer law* 5th ed (Oxford University Press New York 2004) 102.

This thesis does not deal with liability arising from breach of a contract. It is submitted that liability on account of such a breach or due to warranties for faulty AI systems, will be governed in terms of most software contracts. Where no relationship in terms of a contract exists, the party who has suffered loss or injury is left only with a delictual action under law of delict, which offers redress due to breach of duty or infringement of a subjective right or strict product liability flowing from the CPA. The argument is that, the contractual liability of the developer on account of breach of contract for faulty software will be omitted or waived for a majority of contracts, and the party who has suffered loss or injury, in most cases, will need to depend on a delictual action or the CPA.

4.3.1. Conventional software versus AI software

Before discussing the issue of the possible delictual (tortious) liability actions emanating from defective AI, it is important to distinguish AI from conventional computer systems. The most prominent distinction between conventional software versus AI software, relates to the division concerning their knowledge base.⁵²⁹

Conventional software is coded and programmed in order to function in a straight lined or single dimension way, controlling data input by a user across an algorithmic procedure in order to produce a given result, requiring limited interaction between the user and program. In contrast, AI systems use the knowledge base exchanged between objects and incidents in a certain domain as a basis for problem-solving.⁵³⁰ AI systems manipulate knowledge instead of data through the use of heuristics – methodologies that are applied to direct a search procedure to a result.⁵³¹

Heuristics are rules of thumb derived from human experts that operate to progressively limit the program's search by focusing on specific areas, and do not give an accurate answer but provide a hint to the solution.⁵³² By way of inferring knowledge from the existing knowledge base concerning the problem field, the

529 Tuthill GS "Legal liabilities and expert systems" 1991 *AI expert* 45-51.

530 Tuthill 1991 *AI expert* 48.

531 Waterman DA *A guide to expert systems* 1st ed (Addison-Wesley Crawfordsville 1986) 22.

532 Rich E *Artificial Intelligence* (McGraw-Hill New York 1983) 35; Beutel RA "Government regulation of diagnostic software: A threat to artificial intelligence software developers" 1985 *The computer lawyer* 22.

result is derived from AI systems and allows it to deal with uncertainty. With conventional software, the output produced from input data is copied, distinct from AI software systems, where the amalgamation of sizeable quantities of heuristic rules, as well as a non-linear method to reach a solution allows the user to manipulate the output.⁵³³ By selecting data inputs which prompt different areas in the database, the user interactively engages with AI systems, allowing for two users working from the same knowledge base to choose dissimilar options and produce divergent outcomes.

This ability of AI to handle everyday obstacles in a manner that matches the intelligence of humans, renders it visibly different to that of conventional computers, rendering AI more beneficial because it is capable of performing more tasks on its own. On the other hand, for the very same reason, it also poses a risk of causing damage or injury if something goes wrong, leading to the question: who should be liable if something goes wrong?

4.3.2. Sources of system error

The instances where the application of AI systems could result in damage being inflicted, is restricted in this thesis to events that concern a system malfunctioning or failing due its ability to defy a rule due to a design fault, a fault in production, or the result of inadequate software design or coding. Unlike conventional software, AI systems tend to be more vulnerable to defects and errors due to their distinctive system design or problem-solving methods that accommodate inaccurate or incomplete data (as explained in the previous section).⁵³⁴

Issues on account of inaccurate data input, computer hardware or software malfunction, faults or “bugs” are prevalent in conventional software and AI software alike.⁵³⁵ Nevertheless, AI systems are prone to other faults. It could entail faults relating to incompatibility of the database, insufficient expert knowledge, the

533 Tuthill 1991 *AI expert* 48.

534 Tuthill 1991 *AI expert* 48.

535 Maule RM “Applying strict products liability to computer software” 1992 *Tulsa law review* 738.

knowledge engineers not providing correct and complete instructions and rules, or the failure to maintain or upgrade the system which generates an outdated output.⁵³⁶

With regard to this study, it must be accepted that in respect of defective AI systems, the cause of errors or failings is inherent in the AI system itself, and is not due to a defect related to other computer hardware or software. A distinction should also be made where injury is caused by a defective AI system, as opposed to injury that arose from the improper use of the system, or an unwarranted dependence on the AI system on the part of the end-user, such as self-assistance applications used without the involvement or assistance of any professional user and predominantly relies on the user's input. In the latter instance, where there is an improper handling of an AI system, in relation to this study it is assumed that the AI system itself is not defective and that it is fit for purpose.

4.3.3. The identification of possible liable parties if AI systems cause harm

The law consists of legal principles that are concerned with human behaviour. The traditional concept of negligence (as a cause of action) requires legal personhood that confers rights and obligations.⁵³⁷ As computers are not legally recognised as persons, and until such time as the issue of conferring legal recognition of computers as persons has been resolved through the courts, this research proposes a permissive regulatory framework that places liability on AI creators and health professionals based on prevailing delict-based principles.

Not only are there several participants employed in the design and application of sophisticated AI systems, but most of these systems are designed to function in a manner that involve human actors being kept "in the loop". Whilst the various tasks and roles that were formerly only performed by humans can now also be performed by intelligent machines, human actors will always be interconnected at the successive stages of the supply chain (encompassing design, testing, manufacturing, deployment and application). This gives rise to complexities in

536 Tuthill 1991 *AI expert* 46-47.

537 Bathae 2018 *Harvard journal of law and technology* 891.

identifying the correct allocation of accountability between human actors involved, due to the intricate and multifaceted dealings between them.

4.3.3.1. Distinction in software

An aspect that affects the ability to identify accountable parties when it comes to the issue of legal culpability, relates to the functional distinction in the nature of the software applied in AI, which can classify it either as products or services. In instances where the software is either incorporated into an apparatus, or in a robot where the software instructs the machine or robot to perform a task which produces a substantial or material result, the system effects a material output as the software functions like any other automated system, and such software may therefore be considered to be a component of the apparatus or robot and accepted as a product or device.⁵³⁸

This means that a causal connection will typically arise between a flaw in the software and the ensuing damage. On the other hand, where the AI system generates an intellectual outcome or result, the involvement of a human actor can lead to a disruption in the causation chain between a flaw in the system and ensuing harm. These apply to instances where software serves as an instrument for professional advice and decision rendering, and provides recommendations for the clinician, traditionally involving the rendering of a services although acquired as a product.⁵³⁹

This rendering of advice (by using an AI system for assistance) amounts to the rendering of professional and expert services. Thus, professional liability comes into play. Although AI systems fall into both material and intellectual output, the rapid intellectual development of AI systems often regard it as encompassing intellectual output. Therefore, AI's that generate an intellectual output are intangible in nature, and therefore renders it difficult to fit into traditional legal concepts of products

538 Alheit 2006 *The comparative and international law journal of Southern Africa* 267-268.

539 Alheit 2006 *The comparative and international law journal of Southern Africa* 267-268; McKinsey and Company Report March 2020 "Transforming healthcare with AI: The impact on the workforce and organization" https://eithealth.eu/wp-content/uploads/2020/03/EIT-Health-and-McKinsey_Transforming-Healthcare-with-AI.pdf (Date of use: 30 July 2020).

(goods) or services, as well as rendering it challenging to assign legal responsibility for its flaws and defects.⁵⁴⁰

Whether AI is categorised as products or services is an important consideration when assessing if it should be the focus of professional negligence, or strict product liability claims.⁵⁴¹ The intellectual output of AI that encompass characteristics of both products and services is problematic in Anglo-American law, as products are regulated by strict product liability laws and services in terms of negligence theory.⁵⁴²

The ways in which software can be acquired lends itself to be placed under three separate categories: software developed for "mass-marketing mode" as off-the-shelf or standard (turnkey system) that have a limited degree of expert knowledge; software that is custom-designed for a specific use; and software that is adapted for a specific user's requirements (hybrid system). The contact between the AI developers and the user is negligible in respect of the first listed category, the highest in relation to the second listed category, and moderate in respect of the third listed category.⁵⁴³

Medical AI systems, such as IBM's Watson, augments the doctor's own expertise and knowledge, where the system needs the medical record and laboratory outcomes of the patient from the treating doctor in order to produce a probable diagnosis and treatment plan. As such, Watson would be categorised under the second (customised) or third (hybrid) type of customised AI systems, as its knowledge base contains a number of rules that will guide a search process to a probable diagnosis.

Medical AI systems that are designed for in-home use, which can be operated by an ordinary person in the absence of the assistance or intervention of a doctor, are

540 Turley TM "Expert software systems; the legal implications" 1988 *Computer law journal* 455; Reed and Angel *Computer law* 102; Tuthill GS 1991 *AI expert* 45-51.

541 Scott MD "Tort liability for vendors of insecure software: has the time finally come?" 2008 *Maryland law review* 434.

542 Alheit 2006 *The comparative and international law journal of Southern Africa* 268.

543 Gerstner ME "Liability issues with artificial intelligence software" 1993 *Santa Clara law review* 244.

considered to be “mass- marketing mode” or standard systems that fall into the first category. These categories⁵⁴⁴ are important with regard to the civil liabilities of producers and end-users in relation to the demarcated use of the systems, namely, in cases where injury or harm ensues due to the use of a faulty AI system where the producers are accountable where an AI system is used by the end-user, who becomes liable.

Anglo-American authors suggest that standardised software (category 1 above) ought to be categorised as products governed by strict liability principles, and that customised software, including the hybrid form of customised software (categories 2 and 3 above), are regarded as services governed by negligence principles.⁵⁴⁵

According to the law in South Africa, the procurement approach is not contingent on the type of the services or products, but rather on the type of arrangement that was executed amongst the parties involved. Delictual remedies are grounded on common principles, irrespective of the type of products, services or transactions.⁵⁴⁶

4.3.3.2. Legal relationships

In order to determine the liability for injury or harm inflicted upon a party using defective AI systems, the legal relationships amongst the human actors involved with such systems, and the relationships between such human actors and the harmed parties must be considered. Therefore, it becomes essential to identify the individual actors that are part of the development and design of an AI system.⁵⁴⁷

In this regard, a distinction can be made amongst two main categories of individuals – the producers and end-users. Producers include the: designers;⁵⁴⁸ domain

544 Namely the mass-marketing mode, custom made or modified systems.

545 Alheit 2006 *The comparative and international law journal of Southern Africa* 269.

546 Neethling, Potgieter and Visser *Law of delict* 346.

547 The legal relationship is based on the conduct of the person which must have caused the harm. See *Smit v Abrahams* 1994 (4) SA 1 (A) and *Fourway Haulage SA v SA National Roads Agency* 2009 2 SA 150 (SCA) dealing with the wrongful or culpable act that must be linked sufficiently closely or directly to the loss for legal liability to ensue with due regard to policy considerations; Waterman A guide to expert systems 8-11.

548 Designers are the individuals for whom the AI system is commissioned.

experts;⁵⁴⁹ knowledge engineers;⁵⁵⁰ programmers;⁵⁵¹ developers;⁵⁵² or suppliers,⁵⁵³ which design and/or manufacture and/or supply the AI systems through knowledge engineering.⁵⁵⁴ The end-users⁵⁵⁵ can consist of persons that use the AI systems for assistance in the performance of their professional duties to third parties (e.g. a physician using a medical expert system such as MYCIN designed to enhance physicians in diagnosing and prescribing blood);⁵⁵⁶ or non-professional end-users themselves for instruction in a particular domain (e.g. lay-person uses a self-help AI system in a specific subject area such as for advice on a medical issue).⁵⁵⁷ Thus, the development of the AI system is collaborative in nature, relying on several people with different skills to produce a system that is sold by the supplier to the end-user.

The producers, or in some instances the end-user who are professionals (e.g., doctors) using the AI systems for assistance in the performance of their professional duties to third parties, are the possible participants that stand to be held accountable under civil law under delictual principles in South Africa (discussed in the sections that follow). The question of accountability is contingent on the legal relationships that arise between the abovementioned parties, and the relationships between them and the harmed parties – since different legal standards would apply to the various relationships.

549 Domain Experts provide the knowledge that is inputted into the AI system are experienced in problem solving within a domain.

550 Knowledge Engineers are the bridge between the domain engineers and programmers who are not experts in the particular subject area but build the AI system based on converting the domain expert's knowledge into rules for use by the programmers in the system.

551 Programmers write the program or code for the AI system once rules are created by the Knowledge Engineers.

552 Developers are involved in the production concerning the AI system such as manufacturers.

553 Suppliers supply the AI systems to the end-user and this could also be the developer.

554 Knowledge Engineering involves a system of designing "expert systems" by obtaining information from experts and incorporating it into expert systems.

555 End-users do not participate in the production of the AI systems and are the individuals for whose benefit the AI system has been created.

556 Waterman *A guide to expert systems* 29.

557 Waterman *A guide to expert systems* 29; Compton P and Jansen B "Knowledge in context: A strategy for expert system maintenance" 1988 *Proceedings of the Australian joint artificial intelligence conference* 1-2; Gill GT "Early expert systems: Where are they now" 1995 *MIS Quarterly* 51-81.

Where no contract exists between the person that has been harmed and the relevant producers involved, or in the event that a contract exists but excludes or waives liability by way of an exclusion or indemnity clause, the injured party's only recourse will be based on a delictual action. For example, the professional end-user, such as a physician, could also have purchased the software from the developer of the software, or directly with the supplier, and in this instance – if the system malfunctions, the developer or supplier could be accountable to the physician as the harmed party under delict where no contract exists.

If the harmed party is the patient, they would not have concluded a contract with the developer or supplier. Thus, the physician would be liable in delict towards the patient, where an AI system was employed by the physician for diagnosis or treatment and caused injury to the patient. In this instance, the injured party could also be an interested third party associated to the patient (for instance the dependents of the patient).

Since an error in the AI system may emanate from the human actors, referred to as the producers individually or jointly, it could be problematic to identify with certainty which one of them is accountable for design errors with regard to an AI system. Another difficulty in identifying the responsible party, is in relation to those systems that are designed for a specific application, where the end-user may comprise of both the knowledge engineer and AI system designer.

For instance, in the development of an application that assesses and screens the status of vital signs concerning a patient, a doctor is likely to be consulted to explain to the software programmer the parameters and ranges that are appropriate, and the required sequence thereof. As such, the doctor could be liable for medical malpractice based on delict, if he used the system for his patients and for errors in the AI system (if he was involved in the design and provided the knowledge base for the AI system). This distorts the difference amongst the producers and end-users. Thus, patients may under delict recover damages for injury or loss from

doctors, healthcare organisations, and producers of medical devices, since the aforementioned actors did not comply with lawfully recognised obligations.⁵⁵⁸

4.3.4. Negligence and product liability

As indicated previously, the law currently does not accord legal status to computers as with humans. As noted in *United States v Athlone Indus Inc*,⁵⁵⁹ “robots cannot be sued” because the robot lacks legal capacity, even though “they can cause devastating damage”. The increasing resemblance in the intelligence of persons and machines may ultimately prompt legal acceptance of intelligent computers as “persons”. Oxford Professor Nick Bostrom suggests that, machines “capable of independent initiative and of making their own plans [...] are perhaps more appropriately viewed as persons than machines.”⁵⁶⁰

A conceivable approach is to assign legal “personhood” to the AI system, acknowledging such system as a separate legal person in law. In such instance, the system will not be considered an agent of a person, but deemed a principal in the same way as a human being or entity under the notion of legal personhood, with rights and responsibilities in its own right capable of being litigated against directly for negligent acts.⁵⁶¹

This is particularly true where autonomous AI agents are no longer regarded as agents, as they are given the aptitude to think independently, and can defy the instructions of their human principal. To give effect to this, the AI system would need to be covered for insurance by its human principal, akin to how physicians are indemnified through medical negligence insurance, and any losses or injury will be covered by the insurance. The funds for insuring the AI system could be provided by the relevant producers or end-users, permitting a “different form of cost-spreading” which fosters justice, because it does not only place a sole burden on

558 Sullivan and Schweikart https://journalofethics.ama-assn.org/sites/journalofethics.ama-assn.org/files/2019-01/hlaw1-1902_1.pdf (Date of use: 4 June 2019).

559 746 F.2d 977, 979 (3d Cir. 1984). The case considered a manufacturer’s civil liability for a robotic pitching apparatus which was faulty due to the machine’s defects.

560 Bostrom N “When machines outsmart humans” 2003 *Futures* 763.

561 Vladeck DC “Machines without principles: liability rules and artificial intelligence” 2014 *Washington law review* 122.

the technology's creators to absorb the costs, but also convinces end-users to carry a portion of the expenses.⁵⁶² As indicted in Chapter 3 at section 3.2.5, Asimov's famous description of the law of robots, i.e. that "a robot may not injure a human being, or, through inaction, allow a human being to come to harm", would then be satisfied if AI machines were directly held accountable for their mistakes.

Until the issue of according legal recognition to AI technology is resolved, the solution would be to hold the producers or professional end-users of AI liable, as proposed in the preceding section, but allow the aforementioned to secure an indemnity or financial contribution from any other accountable actors.⁵⁶³ Dr Morley and Dr Lawrence from the Newcastle University Law School, in addressing the importance of regulating AI, state that, "if AI are not awarded legal personality then the government will need to decide who takes legal responsibility for these technologies, be it the developers (companies) or the owners".⁵⁶⁴

The regulatory system in most countries have dealt with transformational technological change in the past, adapting common-law delictual principles – most notably in the development of the internet. Generally, when considering the issue of legal liability for AI systems, typical delictual claims in the realm of medicine and healthcare will rely upon traditional liability concepts, such as negligence and product liability.⁵⁶⁵ Thus, when remote actors such as a producer and an end-user are associated on account of the acquisition of an AI system or solution, or in the event that the end-user places reliance on information delivered by the AI system, these scenarios would consider negligence or product liability.

562 Vladeck 2014 *Washington law review* 125.

563 *In re Toyota Motor Corp. Unintended Acceleration Mktg, Sales Practices, & Prods. Liab. Litig* 754 F. Supp. 2d 1145 (C.D. Cal. Oct. 7, 2013), the class of plaintiff's alleged, that some vehicles produced by Toyota contained a software flaw that affected the acceleration mechanism in the defendant's vehicles. Typically, the software is licensed by the manufacturer to the defendant, and thus the plaintiff sued the manufacturer for breach of product warranty amongst other claim, and the manufacturer, in turn, could have requested financial contribution or an indemnity from the software provider either in terms of breach of warranties or other contract remedies.

564 Morley S and Lawrence D "Written evidence (AIC0036)" <http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/artificial-intelligence-committee/artificial-intelligence/written/69435.html> (Date of use: 30 July 2020).

565 Arkfeld 2020 *Judges' Journal* 6; Maule 1992 *Tulsa law review* 736.

4.3.5. The test for negligence in South Africa

4.3.5.1. AI as product versus AI as a tool

When a system malfunctions, it may result in a delict act on part of the producers or end-users, in relation to any one of the three AI systems categories described in section 4.4.4. In terms of Anglo-American law, liability based on tort stemming from private or pecuniary harm sustained by software, is premised on whether software is considered to be goods or services.⁵⁶⁶ The importance of this classification, relates to the various concepts of liability that are relevant to both goods and services separately.

Anglo-American writers hold the view that standard software – earlier referred to as “mass-marketing mode” software – should be classified as products, and that customised or modified software amounts to the rendering of services.⁵⁶⁷ As a result, strict liability principles apply to “any product”, and negligence principles apply to services.⁵⁶⁸ For example, if the FDA in the US has approved an AI system as a medical device, strict product liability would be applicable. But in those instances where the AI system is used for decision making by the clinician, professional accountability giving rise to negligence principles will apply.⁵⁶⁹ In South Africa, under the delictual principles, it makes no difference whether AI systems are classified as a product or a service, since both will be actionable in terms of the *Aquilian* action⁵⁷⁰ or non-patrimonial loss⁵⁷¹ when damage is unlawfully caused.

566 Reed and Angel *Computer law* 102; Turley 1988 *Computer law journal* 455.

567 Alheit 2006 *The comparative and international law journal of Southern Africa* 269; Gerstner 1993 *Santa Clara law review* 250; Waterman A guide to expert systems 32-58.

568 Gerstner 1993 *Santa Clara law review* 250; Turley 1988 *Computer law journal* 455-457.

569 McKinsey and Company Report March 2020 https://eithealth.eu/wp-content/uploads/2020/03/EIT-Health-and-McKinsey_Transforming-Healthcare-with-AI.pdf (Date of use: 30 July 2020).

570 Where this harm results in patrimonial loss with a monetary value such as medical expenses, one uses the *Aquilian* action to recover compensation of patrimonial damage caused by negligence, see Loubser MM *et al The law of delict in South Africa* (Oxford University Press New York 2009) 4.

571 Non-patrimonial loss does not really have a monetary value such as pain and suffering or injury to personality.

In direct divergence to Anglo-American law of tort, in South Africa the law of delict⁵⁷² is founded on the Roman-Dutch law and aspects of English law.⁵⁷³ Accordingly, the law of delict is grounded on common standards of liability as a fundamental aspect, in terms of which any harm that is inflicted in a wrongful and culpable manner, is tortious and actionable in law. In *Perlman v Zoutendyk*⁵⁷⁴ the court held that,

Roman-Dutch law approaches a new problem in the continental rather than the English way, because in general all damage caused unjustifiably is actionable, whether caused intentionally (*dolus*) or by negligence (*culpa*).

These South African generalist principles for liability relate to any breach of personal interests, and it is proposed that this includes damage arising from AI systems that are defective.⁵⁷⁵ The question arising is whether liability based on negligence due to the use of faulty AI systems, could be dealt with under our prevailing common-law doctrines with no interference of the lawmakers.

In terms of our law of delict, liability for negligence may arise if five requirements are satisfied:⁵⁷⁶

- (1) There must be conduct or omission by the wrongdoer. An act may be a positive act of doing something or it may constitute an omission which entails an omission to take positive measures to avoid damage;
- (2) The conduct or omission must be wrongful or unlawful. The wrongfulness has to do with the violation of the victim's subjective right(s)⁵⁷⁷ weighed against the interests of the community or "*boni mores*";
- (3) There must be blameworthiness or fault on part of the defendant where he behaved wilfully or negligently. Negligence is weighed in terms of the principle of "duty of care", which involves the reasonable man test being in the same position as the wrongdoer;
- (4) A causal nexus must be present between the wrongdoer's actions, the harm caused to the subjective rights as well as the ensuing damages. This is determined on the factual evidence which must be established on a preponderance of probabilities; and

572 "Delict" and "tort" are used interchangeably. European and Roman-Dutch law jurisdictions such as South Africa use delict while English common law jurisdictions refer to tort.

573 Neethling, Potgieter and Visser *Law of delict* 9.

574 1934 CPD 151 155.

575 Neethling, Potgieter and Visser *Law of delict* 5, ft 11.

576 Midgley and Van der Walt *Principles of delict* para 2.

577 Subjective rights concern real rights (e.g., long-term leases), personal rights (e.g., right to personal safety), personality rights (e.g., impairment of dignity or privacy) and intellectual property rights (e.g., trademarks), see Neethling J, Potgieter JM and Visser PJ *Neethling's law of personality* 2nd ed (Lexis Nexis Butterworths Durban 2005).

- (5) Harm sustained by the plaintiff in the form of patrimonial loss (e.g., hospital expenditures), or non-patrimonial loss (e.g., injury to personality, pain and suffering).

This research deals with delictual liability that is limited to damage arising from a negligent act in the context of a defective AI system. Delictual liability emanating from misrepresentation or conduct associated to fraudulent is excluded for the purposes of this research. The primary intention is therefore to offer redress to an injured party where harm is caused on account of a defective AI system induced by negligence.

4.3.5.2. Wrongfulness and negligence

In South Africa, for a claim in negligence to succeed, it must be established whether the transgressor acted wrongfully in that his or her conduct caused harm in an unreasonable or unlawful way.⁵⁷⁸ Wrongfulness is established by way of a two-fold approach. Firstly, it requires forming a view as to whether any lawfully recognised right, was violated, and secondly if such infringement occurred in an unreasonable manner.⁵⁷⁹

The general test in determining whether conduct is wrongful, is the “objective reasonableness” standard or the *boni mores* test, commonly known as the legal convictions of the public.⁵⁸⁰ This is based on whether in the legal opinion of the public, the wrongdoer transgressed the interests of injured party in an unreasonable way.⁵⁸¹ In the case of products or services offered by AI systems, the “expectations of the ordinary consumer” is the proposed criterion.⁵⁸² Based on the reasonableness standard, a court assesses the wrongdoer’s behaviour in order to determine if it is socially appropriate by counterbalancing the conflicting interests of the complainant, the offender and of the public as a whole.

578 Neethling, Potgieter and Visser *Law of delict* 29; Boberg PQR *The law of delict: Aquilian liability vol 1* (Juta Cape Town 1984) 30.

579 *Universiteit Van Pretoria v Tommie Meyer Films (Edms) Bpk* 1977 4 SA 376 (T).

580 *Carmichele v Minister of Safety and Security* 2001 4 SA 938 (CC) [956-957].

581 In *Coronation Brick (Pty) Ltd v Strachan Construction Co (Pty) Ltd* 1982 4 SA 371 (D) it was held that, “in any given situation the question is asked whether the defendant’s conduct was reasonable according to the legal convictions or feeling of the community.”

582 Vladeck 2014 *Washington law review* 135.

In *Carmichele v Minister of Safety and Security*⁵⁸³ the court determined that, such a counterbalancing of interests is to be:

Carried out in accordance with the spirit, purport and objects of the Bill of Rights, and the relevant factors weighed in the context of a constitutional State founded on dignity, equality and freedom.

When applying the test for wrongfulness, the court has to determine and understand the broader legal opinions of the public with due regard to the legal rules, principles, legal authorities and the circumstances of each case before it. Thus, in terms of the *boni mores* test, one's behaviour will be prohibited if, for example, a personal right has been impaired, or a legally recognised obligation has been breached on account of a defective AI system.

The principle of foreseeability of harm, is an element that the court has due regard to when the test for wrongfulness is applied.⁵⁸⁴ There are grounds of justification⁵⁸⁵ that would on the face of it, justify violation of a legally accepted interest or right and exclude the wrongfulness element of a delict.⁵⁸⁶ However, no "*numerus clausus*" exists concerning such justifications, but it could be further developed based on unique circumstances.⁵⁸⁷ The grounds of justification are disregarded for the purposes of this thesis, as the relevance of the discussion is establishing standards of liability that can be applied when defective AI systems cause harm.

The standard which the South African legal system applies to determine negligence, involves the "objective reasonable person", the "*diligens paterfamilias*".⁵⁸⁸ A person is considered to be negligent when they do not apply the standard of care that a *diligens paterfamilias* would apply in comparable situations, and where the courts determine that the reasonable person would have behaved in a different way if the ensuing harm could be reasonably foreseen and prevented.⁵⁸⁹ Such a test for

583 2001 4 SA 938 (CC) [43].

584 *Coronation Brick (Pty) Ltd v Strachau Construction Co (Pty) Ltd* 1982 4 SA 371 (D) [384].

585 Necessity, private defence and consent to injury are some of the legally recognised grounds of justification.

586 Neethling, Potgieter and Visser *Law of delict* 81; Midgley and Van der Walt *Principles of delict* 95.

587 *Clarke v Hurst* 1992 4 SA 630 (D) [650]; *Argus Printing and Publishing Co Ltd v Inkatha Freedom Party* 1992 3 SA 579 (A) [589].

588 Neethling, Potgieter and Visser *Law of delict* 122; *Boberg Aquilian liability* 274.

589 Neethling, Potgieter and Visser *Law of delict* 122.

negligence is grounded on the established maxim of Holmes JA in *Kruger v Coetzee*.⁵⁹⁰

[F]or the purposes of liability *culpa* arises if (a) a *diligens paterfamilias* in the position of the defendant; (i) would foresee the reasonable possibility of his conduct injuring another in his person or property and causing him patrimonial loss; and (ii) would take reasonable steps to guard against such occurrence; and (b) the defendant failed to take such steps.⁵⁹¹

Whilst the negligence test entails that of the *diligens paterfamilias* when it comes to an expert or a professional such as doctor or software engineer, such a test is that of a “objective reasonable expert” (e.g. doctor or engineer) performing the same activity.⁵⁹² In *Van Wyk v Lewis*,⁵⁹³ this test is stated by the court to be “the general level of skill and diligence possessed and exercised at the time, by the members of the branch of the profession to which the practitioner belongs.” An expert or professional would have a minimum level of specialised expertise in a particular field, and undertakes work requiring special skills. The software industry refers to “software engineering” as the activity of designing and developing software, and the software producer may be regarded as an expert in the development of software.⁵⁹⁴ Likewise, the producer of AI systems may be regarded as an expert in the construction of such systems.

A physician’s actions will be regarded as an expert in the field of medicine based on the same knowledge, skills, and expertise of a reasonable physician under similar circumstances. Under the dictum *imperitia culpa adnumeratur*, a lack of knowledge or experience is considered as negligent when one assumes a task that requires professional skill and knowledge, whilst they know or should have known that they do not possess such expertise or knowledge.⁵⁹⁵

590 1966 2 SA 428 (A).

591 *Kruger v Coetzee* 1966 2 SA 428 (A) [430]. Also see *Santam Versekeringsmaatskappy Bpk Swart* 1987 2 All SA 443 (A) [819]; *Ngubane v SA Transport Services* 1991 1 SA 756 (A) [776].

592 Neethling, Potgieter and Visser *Law of delict* 129; Midgley and Van der Walt *Principles of delict* 71.

593 1924 (AD) [438-444].

594 Rowland D and Rowland JJ “Competence and legal liability in the development of software for safety-related applications” 1993 *Law, computers and artificial intelligence* 238.

595 Neethling, Potgieter and Visser *Law of delict* 130; Midgley and Van der Walt *Principles of delict* 70-71; Boberg *Aquilian liability* 346-347; Gerstner 1993 *Santa Clara law review* 247;

It is submitted that standards of care tend to evolve from time to time with new developments in medical knowledge and technology, and this could give rise to legal uncertainty for the courts in deciding on the prevailing standard at a given time. Another difficulty with establishing a standard of care when it comes to AI systems, also arises because the standards may differ depending on the array of human actors that are engaged in designing and producing the AI system, making it difficult to identify which one of them caused the error or defect.

Certainly, for experts that are appointed because of the specialised skills they possess, a higher degree of care is required, particularly given the potential risk for disaster⁵⁹⁶ with faulty programs. The public considers them to be accountable for staying well-informed when it comes to technological advancements, and for their expertise to create and oversee quality and safety controls for the systems. The role of the domain expert, knowledge engineer, and software programmer indicates that the standard of an expert is required, because the success and potential of the AI system is contingent upon the expertise involved in developing the system. Other human actors that may be part of the design and development of the system, such as a quality assurance engineer, could be subjected to a lower standard.⁵⁹⁷

The regular manifestation of emblematic factual cases (particularly involving dangerous things), has yielded an array of standards and best practices⁵⁹⁸ regulating cases, which are not necessarily binding, but has persuasive authority. A key factor in creating and setting a standard of care when it comes to AI systems, is to establish a body of standards regulating usage in the industry. In evaluating the standard of care, industry best practices ought to be considered.

Simon's Town Municipality v Dews 1993 1 SA 191 (A); *Savage and Lovemore Mining (Pty) Ltd v International Shipping Co (Pty) Ltd* 1987 2 SA 149 (W) [210].

596 One of the alarming aspects is that the potential for multiple or reoccurring harm caused by defective AI systems can be due to the conduct of one human expert because any mistakes or oversight on their part will become an essential part of the system that is applied to multiple users.

597 Gerstner 1993 *Santa Clara law review* 247.

598 "Best practices" denotes certain industry practices which over time have been established as successful and are applied by various industries, see Scott 2008 *Maryland law review* 446.

To this end, product safety standards such as the ISO 9000 and 9001, which are established and globally accepted by recognised bodies, for example, the South African Bureau of Standards (SABS),⁵⁹⁹ would have an essential role to play in establishing negligence. Certification issued by an authoritative agency, confirming that a product was created based on recommended quality assurance and control standards as with the ISO standard series, suggests that an acceptable standard of care had been applied by the producer to ensure the safety of end-users.⁶⁰⁰

As a minimum, employing best practices relating to the design and testing of AI platforms is recommended as it can offer a starting point for establishing an acceptable standard of care.⁶⁰¹ Thereafter, a risk analysis through on-going testing and debugging procedures of the AI application field should be applied, as to determine the appropriate specification methods and models that will offer the most reliability in the completed product. As a final measure, external quality assurance and control should be carried out through certification such as that offered by the ISO series, prior to the system being released to the public.

However, it should be borne in mind that conformity with industry guidelines does not *per se* establish the actual standard of care; it is only demonstration of a standard of due care and offers persuasive evidence of ordinary care.

Our courts have also, on occasion, ignored the reasonable man test for negligence premised on the Roman-Dutch regime and have employed the English “duty of care” doctrine.⁶⁰² Where a duty of care is said to have been infringed, the two questions to consider are, namely (a) whether the wrongdoer had a duty to behave reasonably towards the claimant (the duty question); and (b) whether the wrongdoer violated such a duty (the negligence question).

599 See “ISO 9000 and 9001 standards for quality management systems” available at <https://www.iso.org/standards.html> (Date of use: 29 July 2020).

600 Alheit 2006 *The comparative and international law journal of Southern Africa* 276.

601 Scott 2008 *Maryland law review* 446.

602 Midgley and Van der Walt *Principles of delict* 59-60, 148-151. Boberg *Aquilian liability* 274; Neethling, Potgieter and Visser *Law of delict* 158.

If the answer to both issues is in the positive, the defendant would have acted negligently. In *Administrateur Natal v Trust Bank van Afrika Bpk*,⁶⁰³ it was held that, reasonable foreseeability is not a consideration in the duty of care, but more the extent of interests to be protected against negligent harm. In deciding on the existence of a duty of care, due regard is given by the courts as to whether the wrongdoer employed the standard of care (which a reasonable person would implement under the same situation to prevent the damage).

Unless the claimant can prove that a wrongdoer had a duty of care to him or her, the claimant shall have no recourse. If such a duty has been demonstrated, the standard of care will then be tested against that of a reasonable man. There is little question that a producer of an AI system owes a duty of care towards an end-user, such as to physicians as end-users of AI medical systems who must ensure that the systems sold to them do not harm their patients.

Thus, the conduct of producers of defective AI systems will thus be *prima facie* unlawful when injury or harm arises by its use. Healthcare practitioners who use the systems in their professional duties also have a duty of care towards patients, and in such instance, their professional liability is at stake and actions instituted against them are based on “medical malpractice”.

4.3.5.3. Reasonable foreseeability and preventability

The negligence test is made up of two factors: (1) reasonable foreseeability of harm, and (2) reasonable preventability of harm.⁶⁰⁴ The standards for establishing reasonable foreseeability were articulated in *Lomagundi Sheetmetal and Engineering v Basson* where the court held:⁶⁰⁵

The sort of circumstances, however, which the Courts often look to in cases such as this in deciding what degree of foreseeability must be proved by the plaintiff before a defendant can be held responsible for the resultant damage are these: How real is the risk of the harm eventuating?; If the harm does eventuate, what is the extent of the damage likely to be; and What are the costs or difficulties involved in guarding against the risk?

603 1979 3 SA 824 (A) [833].

604 *Kruger v Coetzee* 1966 2 SA 428 (A).

605 1973 4 SA 523 (RA) [524-525].

The extent of the risk of the harm resulting from the error, and the extent of damage that such harm is likely to cause, therefore determines whether such harm is foreseeable.⁶⁰⁶ It is accepted that the sophistication of AI systems, and the multiplicity of applications under which they are used, create technical and management problems, and thus makes it impossible for producers to offer perfect and error free systems. However, this does not suggest that no duty should be imposed on them to employ all reasonable methods at their disposal to deliver safe and reliable systems.

For instance, there may be cases where the risks of foreseeable damage arising from the AI system may be limited by the implementation of a reasonable substitute model or design. It is thus not unreasonable to subject producers to a greater standard of care in cases of essential AI applications such as cancer diagnosis, than for non-essential applications such as video games or word processing. Consequently, the system is not required to be faultless, and it should only satisfy the standard of care for a reasonable expert under those conditions.

Once it has been established that the possibility of damage would have been foreseeable by the reasonable person, the next consideration is if the reasonable person would have employed precautionary steps in prevention of the foreseeable harm happening. This would be premised on what the level of risk created by the wrongdoer's conduct is likely to be;⁶⁰⁷ the seriousness of the potential losses or damages if the harm occurs; the effectiveness relating to the wrongdoer's action; and the difficulty in eradicating possibility of harm.⁶⁰⁸

If it has been established the extent of risk of harm is greater than the efficacy of the wrongdoer's action, the reasonable person will employ methods to avert the event of harm. Thus, in the event that a wrongdoer does not employ such methods, his conduct would be considered to be negligent. If the ability to eradicate a risk of harm is greater than the level of risk, it would not be expected of the reasonable person

606 Midgley and Van der Walt *Principles of delict* 143.

607 *Ngubane v SA Transport Services* 1991 1 SA 756 (A) [776]; *Herschel v Mrupe* 1954 3 SA 464 (A) [477].

608 *Lomagundi Sheetmetal and Engineering (Pvt) Ltd v Basson* 1973 4 SA 523 (RA); *Khupa v SA Transport Services* 1990 2 SA 629 (W).

to employ means to avoid the foreseeable harm from happening. In the case of injury arising due a defective AI system, the system can be said to be unsatisfactorily constructed where the foreseeable risks relating to harm caused by an AI system is particularly inordinate in comparison to its foreseeable advantages and benefits, such that reasonable producers or healthcare professionals having knowledge of any foreseeable risks would avoid recommending it to “any class of patients”.⁶⁰⁹

Thus, where physicians use AI systems for assistance in the performance of their professional duties to patients, the physician is considered the end-user of AI systems, because as the healthcare professional, he is the most suitable person to weigh the risk of harm against the likely advantages in employing AI systems. This would mean that the producers of the AI systems may fulfil their duty to reasonably prevent harm, by issuing a warning about the possible dangers of the systems, or providing instructions for safe use to the physicians who will be using them in their professional environment.

The producers’ liability may thus only result from cases where critical material or data concerning the product and its fitness for deployment, was absent or deceptive.⁶¹⁰ A negligent claim will arise if a physician uses the AI system in support of his professional duties, without the required level of care and oversight that is required from a reasonable physician in his situation, or where a reasonable physician would not have considered the system as appropriate and/or of an acceptable standard for the purposes that it was used. Physicians have a responsibility to scrutinise the AI systems before using it. If they consequently neglect to inform patients of the risks and advantages linked to the AI system, the physician could be accountable for negligence under delict.⁶¹¹

609 This accords with the Restatement (Third) of Torts: Products Liability, section 6 of the American Law Institute (1998).

610 *Bayer South Africa (Pty) Ltd v Viljoen* 1990 2 SA 647 (A).

611 Sullivan and Schweikart https://journalofethics.ama-assn.org/sites/journalofethics.ama-assn.org/files/2019-01/hlaw1-1902_1.pdf (Date of use: 4 June 2019).

4.3.5.4. Onus of proof

The claimant is required to demonstrate on a preponderance of probabilities that a wrongdoer acted negligently. The exact standard of care will be decided on the facts presented. In terms of the Anglo-American doctrine in tort law, *res ipsa loquitur*,⁶¹² a court can deduce or presume negligence premised on the facts presented to it, where there is no direct proof as to the cause of the incident.⁶¹³

The courts in South Africa have been hesitant to follow the dictum of *res ipsa loquitur* in medical malpractice lawsuits in light of the judgement under *Van Wyk v Lewis*, where the dictum was overruled.⁶¹⁴ In *Goliath v The MEC for Health, Eastern Cape*,⁶¹⁵ the court for the first time opened the door for the *res ipsa loquitur* dictum in medical malpractice cases. It becomes challenging to demonstrate the negligence of manufacturers, due to the complexity involving the technological development and production process of AI systems.

Neethling *et al*⁶¹⁶ state that this difficulty should be alleviated in a similar manner as with the Anglo-American legal system, which entails the recognition of the *res ipsa loquitur* dictum. They propose that presumption of negligence ought to be considered for those circumstances where a plaintiff is able to demonstrate that he or she suffered some prejudice due to a product being defective at the time that the manufacturer relinquished control and possession thereof.

The development of AI involves a complicated technological process, and this may render it problematic to demonstrate negligence of the producers of the system. In this regard, the doctrine of inference of negligence is particularly useful as it absolves the wrongdoer from having to prove that he did not act negligently, as he

612 The term means “the facts speak for themselves”.

613 Neethling, Potgieter and Visser *Law of delict* 50-54.

614 1924 (AD) [438-444].

615 2015 (2) SA 97 (SCA).

616 Neethling, Potgieter and Visser *Law of delict* 348-349.

is only required to demonstrate that the relevant facts of a case support a finding not involving negligence.⁶¹⁷

The critical legal issues involved in all AI applications are whether its provable outcome is based on accurate and adequate data, and whether the algorithms used to perform and produce the outcome are correctly coded. Therefore, the input data, training data, algorithms, and outcomes must be analysed to determine whether or not the AI has produced a harmful outcome. In establishing liability, a court will need to understand the design process that form the basis for the AI decision-making outcome, such as which data and algorithms were selected and whether or not they were tested by an independent third party; what data were selected for the training; how the algorithm was validated and tested; and how old the algorithm is and if it has been recently tested.⁶¹⁸

A plaintiff will face extreme challenges to demonstrate negligence because they lack the insight into the “black box” associated with AI software and its design methods. The complicated and enigmatic landscape of AI’s design, which is not properly grasped by most except for its producers, prompts end-users to place their trust and reliance on the merchant’s depiction of AI systems. Consequently, a plaintiff seeking damages for negligence will likely be subjected to significant fees because of having to hire technology lawyers, engineers and other expert witnesses.

4.3.5.5. Contributory negligence

In South Africa under the common law, the claimant’s contributory negligence prevented the plaintiff from recovering damages against the defendant who was also negligent, is referred to as the “all or nothing rule”.⁶¹⁹ The Apportionment of Damages Act⁶²⁰ has now changed the situation, and courts may apportion the damage in respect of each party in terms of their respective extent of negligence.⁶²¹

617 In *Bayer South Africa (Pty) Ltd v Viljoen* 1990 2 SA 647 (A) the appellate division held that, policy requirements may motivate the acceptance of the *res ipsa loquitur* dictum regarding matters concerning product liability.

618 Arkfeld 2020 *Judges’ Journal* 6-7.

619 Neethling, Potgieter and Visser *Law of delict* 149.

620 Act 34 of 1956.

621 Section 1(1)(a).

Given the array of human actors that are implicated in either the design or application of AI systems, any contributory negligence in respect of each would become highly relevant in a claim of negligence. It would mean that the producer (such as a domain expert), knowledge engineer, programmer, etc., could only be contributory liable to the extent of his or her respective degree of negligence, or if the physician used a defective AI system on his patient that he did not test himself.

Contributory negligence would not be regarded as a legal defence as it does not absolve the defendant from liability, but rather aims to limit the damages awarded to the plaintiff.

4.3.5.6. Factual and legal causation

Negligence will only ensue if a causal link exists between the offender's conduct and the claimant's damages. In *Tuck v Commissioner for Inland Revenue*,⁶²² Corbett JA held that,

factual causation is whether there is a factual nexus of cause and effect between the act or omission of the party concerned and the harm for which he is sought to be held liable.

In the absence of any evidence that an individual's conduct caused the injury to the injured party, such a person will escape delictual liability. In *Minister of Police v Skosana*,⁶²³ the Supreme Court of Appeal accepted that factual causation can be confirmed by employing the “*conditio sine qua non*” (“but for”) test, where it must be established if the offender's actions caused, or significantly caused the harm inflicted on the claimant. Thus, the defendant's conduct should be the “*sine qua non*” in relation to the ensuing harm.⁶²⁴

Legal causation is concerned with the remoteness of damage which entails a sequence of consequences put into action by the defendant's behaviour, where it must be decided if the defendant must be held responsible for any “remote” consequences. Numerous tests to determine legal causation were proposed, but in

622 1988 SA 819 (A) [832-833].

623 1977 1 SA 31 (A) [34-35].

624 *International Shipping Co (Pty) Ltd v Bentley* 1990 1 SA 680 (A).

*S v Mokgeth*⁶²⁵ the appellate division opted for a flexible criterion based on determining if the factual link between the defendant's actions, and its ensuing consequences is strong enough to impute such consequence to such defendant according to policy deliberations based on equality, justice, foreseeability and causation.⁶²⁶

4.3.6. Product liability in South African law

Historically, in South Africa, reference to “product liability” tends to be restricted to law of delict. Product liability is defined to be “the liability that arises when a product contains a defect which leads to damage to property (patrimonial loss)” or an infringement of subjective rights, such as violation of dignity or physical harm that was negligently inflicted.⁶²⁷ Liability will likely arise where a product contains design or production defects that would render it as unacceptably perilous or hazardous and where pertinent information or warnings concerning the product, its appropriateness for deployment and application, were absent or deceptive.⁶²⁸

As such, unlike the Anglo-American law of tort where strict liability principles apply to product, and negligence principles apply to services, the South African law of delict does not provide for such separation of grounds of liability. There is no distinction between products and services since both will be actionable under the general delictual principles. Under the Anglo-American legal system, products liability is premised on strict product liability, whereas in South Africa, a manufacturer's liability is fault-based.

However, in South African law, the possibility of regarding AI software as a product motivates a review of the *Aquilian* liability principles necessitating fault when dealing with products liability, especially when consideration is given to the disadvantage of

625 1990 1 SA 32 (A).

626 The flexible criterion was also confirmed in *International Shipping Co (Pty) Ltd v Bentley* 1990 1 SA 680 (A).

627 Neethling, Potgieter and Visser *Law of delict* 52-54, 228, 246-250; *Universiteit Van Pretoria v Tommie Meyer Films (Edms) Bpk* 1977 4 SA 376 (T).

628 *Bayer South Africa (Pty) Ltd v Viljoen* 1990 2 SA 647 (A).

the plaintiff in having the burden of proving the negligence of a producer, and also involving systems that are too complex to explain.

There are AI systems that squarely fall in the category of either a product or service and will be classified based on that. A dilemma arises when AI systems generate an intellectual output that are difficult to explain, and blur the lines between products and services because the system's output may arguably serve as both. This is commonly encountered in the case of professional services performed by doctors, who use AI systems to assist them with patient diagnosis or treatment. Thus, the problem is exacerbated when medical AI systems are involved as they could also be categorised as a "medical device" in terms of the Medicines Act.⁶²⁹

A "medical device" is defined to be "any instrument, apparatus, implement, machine, appliance, implant, reagent for in-vitro use, software, material or other similar or related article, including Group III and IV Hazardous Substances contemplated in the Hazardous Substances Act 15 of 1973, intended by the manufacturer to be used, alone or in combination, for humans or animals, for one or more of the following:

- diagnosis, prevention, monitoring, treatment or alleviation of disease;
- diagnosis, monitoring, treatment, alleviation of or compensation for an injury;
- investigation, replacement, modification or support of the anatomy or of a physiological process;
- supporting or sustaining life;
- control of conception; disinfection of medical devices; or
- providing information for medical or diagnostic purposes by means of in vitro examination of specimens derived from the human body; and
which does not achieve its primary intended action by pharmacological, immunological or metabolic means, in or on the human or animal body, but which may be assisted in its intended function by such means".

Where AI software is built into a product and is integral to any "medical device", or forms the component part to a product applied in the production and preservation of such medical device or which on its own is a medical device,⁶³⁰ and causes the

629 101 of 1965, as amended.

630 In the US the term for standalone medical software is described to be "software as a medical device", which the International Medical Device Regulators Forum (IMDRF)

product to operate defectively, the product as a whole will be subjected to product liability.

The current definition broadly covers software, hardware or the combination thereof and can apply to all types of AI devices that is used in healthcare and related applications as a regulated device. To classify medical devices in accordance with the definition under the Medicines Act, due regard is given to the manufacturer's intended purpose, design and functionality of the medical device. Therefore, when a manufacturer intends for an AI system to be applied towards examining or changing part of the anatomy in a human, then plausibly that system would be accepted as a medical device under the intended purposes of the Medicines Act, and as with any other medical device, it could be subjected to the certification and registration standards under the Medicines Act. This would also mean that medical AI systems may in general be regarded as a product in South Africa, for which the producers could incur delictual liability based on product liability.

However, the difficulty arises in the endorsement of AI for licensing under a medical device in terms of the current regulatory framework. There are "locked" AI devices that deliver identical outputs on each occasion that the same input is supplied, and where any algorithm modifications will possibly necessitate premarket review with regard to modifications outside of the initial market authorisation.⁶³¹ Nevertheless, the traditional regulatory paradigm of medical device licensing before premarket release, was not designed for AI systems that can continuously learn, change and acclimate in real-time after they reach consumers and where the output may vary prior to and subsequent to modifications that are applied.

Adaptive AIs are particularly used for clinical recommendations, and decision support for patient diagnosis or treatment. The highly autonomous and flexible landscape of AI systems necessitates novel regulatory approaches that enables a rapid progression of technology driven innovation, and that will enable these devices

631 Gerke S *et al* "The need for a system view to regulate artificial intelligence/machine learning-based software as medical device" 2020 *NPJ digital medicine* 1-4.

to constantly develop while offering effective safeguards. In South Africa, medical devices are administered and registered under the Medicines Act read with its Regulations.

The medical device division of the South African Health Products Regulatory Authority (SAHPRA) was established under regulations to the Medicines Act. SAHPRA regulates the licencing of medical device establishments such as producers, wholesalers or suppliers in South Africa, to ensure that medical devices are certified and meet an adequate standard of safety, efficacy and quality control prior to market release.⁶³² The SAHPRA guidelines recommend a quality and safety management system and handbook, which the applicant of a licence must compile, submit and maintain up to date, encompassing detail about the applicant, its operations, key employees, quality review and control policies and procedures, and work directives as to validate its fitness to participate in the offerings of medical devices and associated services in line with regulatory requirements.⁶³³

As a part of the application, the applicant must indicate the nature and risk class of medical devices to be produced, supplied, or wholesaled, and a declaration must be made regarding the level of the quality control procedures that are employed.⁶³⁴ The quality management system recommended under the SAHPRA guidelines are aligned to the ISO13485 framework.⁶³⁵

The ISO13485 sets out quality control system requirements that can be implemented by an organisation in the design, manufacture, storage and distribution, installation, examining, final decommissioning, and disposal of medical devices.⁶³⁶ No medical device may be produced or traded with without a licence

632 Section 22C(1)(b) of the Medicines Act as amended read with regulation 5.

633 SAPHRA "Medical device quality manual" https://www.sahpra.org.za/wp-content/uploads/2020/01/Medical_Device_Quality_Manual_Nov19_v2.pdf (Date of use: 12 October 2020).

634 SAPHRA https://www.sahpra.org.za/wp-content/uploads/2020/01/Medical_Device_Quality_Manual_Nov19_v2.pdf (Date of use: 12 October 2020).

635 ISO "Medical devices – Quality management systems – Requirements for regulatory purposes" <https://www.iso.org/obp/ui#iso:std:iso:13485:ed-3:v1:en> (Date of use: 12 October 2020).

636 Section 0.1: Introduction ISO <https://www.iso.org/obp/ui#iso:std:iso:13485:ed-3:v1:en> (Date of use: 12 October 2020).

issued by the SAHPRA.⁶³⁷ Medical devices, save for those that are custom designed, is required to be registered in terms of SAHPRA prior to being marketed or used in the Republic.⁶³⁸ Custom made medical devices refer to devices that are:

(a) specifically made in accordance with a written prescription or order given by a person authorised for the same by virtue of professional qualifications; (b) specifically made in accordance with specific design characteristics; (c) which is intended for the sole use of a particular user; and (d) which excludes mass produced medical devices that only need adaptation to meet the specific requirements of the health professional user.⁶³⁹

Thus, any AI medical devices that are deemed to be “mass-marketing mode”, off-the-shelf or standard (turnkey) systems, will not require licensing as such device would have a limited degree of expert knowledge and are not subject to variances because they generate the same outcome for every occasion that the same input is offered.⁶⁴⁰

The registration of AI devices would concern the classification of devices ranging from moderate to high risk-based, particularly those that evolve and adapt to changes based on data where the manner in which they arrive at their recommended outcomes, may tend to be opaque to doctors.⁶⁴¹ For medical devices that are under the classification of “Class C (medium to high risk)” and “Class D (high risk)”, holders thereof should be able to produce detailed technical credentials⁶⁴² when requested by SAHPRA.⁶⁴³

637 SAHPRA <https://www.sahpra.org.za/medical-devices/> (Date of use: 13 August 2020). SAHPRA took over the functions of the Medicines Control Council (MCC) mentioned in the Medicines Act.

638 Regulation 11 of the Medicines Act, as amended.

639 Regulation 1 of the Medicines Act, as amended.

640 Refer to section 4.3.3.1 where the categories of software were defined.

641 Regulation 11(1)(a) of the Medicines Act, as amended, refers to classes as “(a) class a low risk; (b) class b low-moderate risk; (c) class c moderate-high risk; (d) class d high risk”. South Africa risk-based classification scheme is based on the Global Harmonization Task Force (GHTF) scheme, an organisation consisting of members from various countries that promotes standards concerning the safety and quality of medical devices.

642 Technical documentation is defined as “the documented evidence, normally an output of the quality management system that demonstrates compliance of a device to the Essential Principles of Safety and Performance of Medical Devices”, see SAHPRA “Guideline for a license to manufacture, import, export or distribute medical devices and IVD’s” https://www.sahpra.org.za/wp-content/uploads/2020/01/Licence_Medical_Devices_IVDs_Nov19_v3.pdf (Date of use: 12 October 2020).

643 Section 4.2.1(d) SAHPRA https://www.sahpra.org.za/wp-content/uploads/2020/01/Licence_Medical_Devices_IVDs_Nov19_v3.pdf (Date of use: 12 October 2020).

Since its establishment not too long ago in June 2017, SAHPRA has not registered medical devices that relate to AI and thus, the challenges that arise as a result of AI's as medical devices have not been addressed yet. In contrast, the US FDA has by now endorsed and approved stand-alone medical software for market consumption, referred to as "software as a medical device".⁶⁴⁴ However, the FDA has admittedly stated that US conventional paradigm of regulation when it comes to medical devices was not intended for progressive AI technologies. Consequently, in 2019 a document was presented by the FDA, "proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD) – Discussion paper and request for feedback" that, introduces its groundwork regarding a prospective strategy on premarket analysis on AI and ML variations.⁶⁴⁵ Likewise, South Africa needs to consider a novel framework for adaptations to AI driven devices, that describes an innovative approach with due regard to patients – which is critical to the licensing of adaptive AI medical devices.

Change will conceivably further expand as, and when, AI driven medical devices begin to work together more dynamically with doctors – for instance by reacting to the doctor's commands and also becoming more acclimated to the requirements of a user.⁶⁴⁶ With the interaction of humans, the AI system perception necessitates assessing how the new data relates with human application and organisational aspects.

The SAHPRA has to develop an approach for exactly how "adaptive" AI algorithms must be regulated as the technology changes and adapts as, and when, they are introduced to new data sets.⁶⁴⁷ The FDA proposes a "total product lifecycle

644 "Software as a Medical" is defined by the IMDRF as "software intended to be used for one or more medical purposes that perform these purposes without being part of a hardware medical device." US FDA "Software as a medical device" <https://www.fda.gov/medical-devices/digital-health/software-medical-device-samd> (Date of use: 12 August 2020).

645 "US FDA: Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD)" <https://www.fda.gov/media/122535/download> (Date of use: 11 August 2020).

646 Gerke et al 2020 NPJ digital medicine.

647 The FDA gives an example of adaptive technology in its discussion paper: It looks at AI/ML-based SaMD's which "receives electrocardiogram, blood pressure, and pulse-oximetry signals from a primary patient monitor" and thereafter "signals are processed and analysed to detect patterns that occur at the onset of physiologic instability", encompassing

regulatory approach for AI and machine learning based software”, that enables evaluation and monitoring from development phase to performance, based on consideration that offset the benefits and dangers and allows increased openness to users for upholding safe and effective AI based devices.⁶⁴⁸ A similar approach may well work for the South African market, given that the real-world performance of AI devices and ancillary risks are unlikely to vary between jurisdictions.

Until such time as a new regulatory framework has been proposed, AI medical devices could be regulated in term of the traditional principles of “products liability”, that deals with the manufacturer’s accountability for damage ensuing from defective or faulty products. According to Neethling *et al*,⁶⁴⁹ the growing technological developments of contemporary society, gives rise to perpetual risk to the public caused by defective customer products.

Technological commerce and trade involving AI systems, are bound to increase and this will prompt the necessity for expert knowledge and skill transfer. Increasingly, South Africans will engage in products and associated services that integrate AI, whether as merchants, consumers, or as observers. With AI systems that generate intellectual outcomes, the technological shortcomings of the users are heightened, due to the complicated outputs of the technology. The average consumer will not possess specialised knowledge, nor have insight into the complex development processes to identify any defects in an AI application that may impact him. Due regard should further be given to the fact that in most scenarios, no contract between the producers and consumers of products exists.

In the case of *Kroonstad Westelike Boere-Ko-operatiewe Vereniging Bpk v Botha and Another*,⁶⁵⁰ the court articulated the principle concerning a “merchant seller’s” accountability relating to deficiencies in the products marketed as follows:

[L]iability for consequential damage caused by latent defect attaches to a merchant seller, who was unaware of the defect, where he publicly professes to have attributes

a warning for the doctor “that prompt clinical action is needed to prevent potential harm to the patient”. See <https://www.fda.gov/media/122535/download> (Date of use: 11 August 2020).

648 <https://www.fda.gov/media/122535/download> (Date of use: 11 August 2020).

649 Neethling, Potgieter and Visser *Law of delict* 345.

650 1964 3 SA 561 (A).

of skill and expert knowledge in relation to the kind of goods sold [...]. Whether a seller falls within the category mentioned will be a question of fact and degree, to be decided from all the circumstances of the case. Once it is established that he does fall within that category, the law irrebuttably attaches to him the liability in question, save only where he has expressly or by implication contracted out of it.

In *Langeberg Voedsel Bpk v Sarculum Boerdery Bpk*,⁶⁵¹ Schutz J expresses his disagreement with the principle in *Kroonstad* and points out that,

[t]he merchant is denied the opportunity to see, to feel or to smell the produce that passes through his hands. He can as little examine the metal in the bearings as the beans in the tin or the chip in the computer [...]. It seems to me cumbrous, wasteful and uncertain of result, and therefore unjust, to require a buyer to prove and a seller to resist in case after case the proposition that the latter publicly professes to have attributes of skill and expert knowledge in relation to particular goods.

Based on conflicting views from the above-mentioned authorities, the law on products liability in South Africa is considered to fall short in tackling the demands of the public. This has become more prevalent in the contemporary era of AI. It could be claimed that an injured party has an unreasonable onus to establish negligence, based on a preponderance of probabilities, in favour of the producers or distributors. As indicated earlier, it is challenging for a claimant to establish negligence when sophisticated products are involved, and such claimant is disadvantaged as they do not have access to the knowledge-base in the same way as the producers of AI systems would have.

The question concerning whether goods are unreliable and deficient must be addressed taking into account the lawful convictions of the community (“*boni mores*”). It would assist our courts to apply the spirit of fairness and justice of a specific community, and to provide some redress to a claimant in considering the Anglo-American regime of the dictum *res ipsa loquitur*,⁶⁵² where the claimant avers that a product is “unreasonably dangerous” and/or “unsafe” and the factual circumstances support a presumption of negligence.

Usually, unreasonably dangerous products will be deemed to be defective, and such products would be unreasonably dangerous, if they fail to address the needs of a reasonable customer or user with regard to its reliability and safety.⁶⁵³ As a result,

651 1996 2 SA 565 (A).

652 For a discussion of the maxim refer to section 4.4.5.4.

653 Neethling, Potgieter and Visser *Law of delict* 306.

the claimant will only have the onus of demonstrating that a product is “unreasonably dangerous” and/or “unsafe”, leaving the courts to make the presumption of negligence in relation to the producers. The burden will thereafter transfer to the producers to demonstrate with factual evidence that reasonable care was taken in the design, development, or distribution to ensure a safe product.

A producer of AI systems would thus have a lawful duty of care, with due regard to the beliefs of the community towards deterring the supply of faulty systems into the marketplace (which is considered to be a risk to consumers).⁶⁵⁴ If the defective system causes harm, it will be regarded as wrongful in delict on account of a violation of this lawful duty of care. Therefore, the system ought to be flawed or defective for wrongful conduct to be imputed on the producer.⁶⁵⁵ The general set of defects – as established in the product liability laws of other countries – are also notable in South Africa, namely manufacture, design and training defects.⁶⁵⁶ Based on the criterion of reasonableness, the following obligations are illustrious in determining the duty of care of the producers to:⁶⁵⁷

- The duty to implement reasonable measures to avoid faulty goods from being introduced in the market, or to ensure that they are withdrawn from the market in the event of their release;
- The duty to have due regard to up-to-date information and knowledge accessible for the purposes of planning and development;
- The duty to examine and implement controls over the product once produced; and
- The duty to furnish prospective end-users with written instructions of usage and warnings against the potential dangers and risk of the product.⁶⁵⁸

654 Neethling, Potgieter and Visser *Law of delict* 347.

655 Midgley and Van der Walt *Principles of delict* 241.

656 Alheit 2006 *The comparative and international law journal of Southern Africa* 297.

657 Alheit 2006 *The comparative and international law journal of Southern Africa* 297.

658 For example, in AI applications developed to aid of doctors in the execution of their professional duties it can be inferred that the doctor will read and apply the enclosed instructions and warnings, whereas with standard software marketed to the public, the instructions will have to be simple enough so that the consumer understands it.

For negligence, the infringement of a lawful duty of care is assessed against a reasonable man.⁶⁵⁹ The question is whether a reasonable and prudent producer could in a given situation have reasonably anticipated the possibility that his conduct would result in harm, and whether he would have employed reasonable safety measures to avert it. In the case of AI systems, the producers hold themselves out to possess the requisite skills and expertise relative to the kind of AI they produce.

They can thus be held liable for a latent defect⁶⁶⁰ in an AI system, if they failed to issue written instructions on use of the product or warnings regarding its safety. Professional end-users, such as physicians, may also become negligently liable by handling and using unsafe systems to support them with patient care, without the physician having satisfied himself of its reliability and safety. For producers to be liable for harm, the test for “reasonable foreseeability and preventability” must be satisfied.⁶⁶¹

The test is based on whether, during the act, both the consequences (factual causation) and the causal link that exists between the act and such consequences (legal causation), can be said to have been reasonably foreseeable. This test is not the same as the “duty of care” test to determine negligence. Negligence can also be imputed when there is a failure to implement control systems or product guidelines linked to quality assurance for any new technology.

Therefore, negligence can be imputed on a producer, if essential quality assurance measures were neglected. Observance of applicable product guidelines and codes of good practice such as the ISO series,⁶⁶² may set the standard for the skill and care expected during production and manufacturing. The international standard ISO 9001 for software, offers general guidance regarding the manner in which to execute, sustain and enhance an effective assessment structure that will generate superior quality software. Other countries have adopted the ISO series of standards.

659 Refer to discussion on the reasonable man test in section 4.4.5.2.

660 A latent defect refers to a fault that could not be detected even by way of a reasonable inspection.

661 Refer to section 4.4.5.3 for the discussion on reasonable foreseeability and preventability.

662 Refer to n 578.

In South Africa, the SABS regulates the process, and the standard is identified as SABS ISO 9001.⁶⁶³

In conclusion, in the case of product or manufacturers liability, any producer of AI systems can be held accountable for inherent deficiencies in a program – along with the manufacturer of a product. This liability is grounded on the principles of delict. According to the law in South Africa, manufacturer's liability is premised on the broad principles found in delict, which is fault based. In this regard, the action is established by the producer's conduct concerning the design, production and application of the system.

Such producer's conduct can comprise of either a commission or omission, for instance, the knowledge expert who designed the inference engine of an AI system could be accountable for defective components, or the domain expert who delivered the knowledge base of the system may be liable for omitting to provide sufficient knowledge. The producers' conduct will be tested against the standard of care applied by other producers and consequently, quality guidelines and assurance practices akin to that of the "ISO series", have a valuable purpose when establishing the negligence of the producers.

Any failure by the producer to apply standards and assurance systems, may indicate negligence. Negligence may also arise if an AI system is produced or manufactured without the required expert skills and knowledge, such as an unskilled person who performs the function of the domain expert. As suppliers do not participate in the construction of AI systems, they would not be liable in terms of manufacturer's liability.

However, where there was a faulty or defective product that caused harm, the supplier would be accountable in delict if a duty existed on them to examine the product, and they neglected to do so. The same argument holds true for a physician that uses the AI system in support of his professional duties towards patients, and failed to apply the proper level of care and oversight as required from a reasonable

663 See "ISO 9000 and 9001 standards for quality management systems"
<https://www.iso.org/standards.html> (Date of use: 29 July 2020).

professional in his position. The professional negligence of the physician comes into play where such system causes harm.

4.3.7. Strict product liability under the CPA

4.3.7.1. Definitions

Once it became apparent that the negligence framework would not be adequate to safeguard the consumer concerns and interests, it was only a matter of time that strict liability measures would be instituted. The South African CPA was promulgated to include such measures. The CPA is relevant to all dealings that take place within South Africa relating to the sale of services or goods,⁶⁶⁴ except where a transaction is specifically exempted.⁶⁶⁵ In terms of the definitions, transactions involving the sale of services or goods within the healthcare sector, will be covered under the CPA.

The broad definition of “consumer” under the Act, also encompasses patients as a “consumer” in the application of this legislation.⁶⁶⁶ Depending on the situation, both patients and healthcare suppliers may be recognised as “consumers”. Thus, a healthcare provider will be regarded as a consumer if medical devices are procured from an importer, distributor, vendor or manufacturer.

The CPA defines “*supplier* as a person who markets any goods or services”. In terms of this definition, “*market* means to supply or promote goods or services”, whereas “*supply* means to sell services, or to perform services or cause them to be performed or provided”.⁶⁶⁷ With regard to goods, “*supply* means to sell, rent, exchange or hire for consideration”.⁶⁶⁸ A healthcare facility or healthcare expert will, premised on the situation, also be considered to be a “supplier” in relation to “goods” or “services” under the CPA.

664 See section 1.

665 Section 5(1).

666 See section 1 where the definition of “consumer” is wide.

667 Section 1.

668 Section 1.

The CPA defines “goods to include anything marketed for human consumption”.⁶⁶⁹ Having such a wide connotation it therefore, does not only comprise of medicines, but includes medical devices.⁶⁷⁰ It is submitted that AI medical systems would qualify as goods in the context of the CPA, subject to it being tangible in nature which is normally associated with “goods” and where it was obtained as with normal transactions involving goods, i.e., through sale or rent. An importer, distributor, manufacturer or vendor of medical AI systems, will also qualify as suppliers of goods in terms of the broad definition of “goods”. Computer software relating to AI medical systems may be regarded as intangible, however, once they are built into the product as a whole or they form a component of such product, they can be regarded “goods” as envisaged in the CPA.

According to the CPA “service includes but is not limited to any work or undertaking performed by one person for the direct or indirect benefit of another”.⁶⁷¹ In the healthcare context, services would include professional advice issued by a doctor, or any remedial procedures or treatments undertaken by the practitioner, regardless of the fact that the healthcare practitioner incorporates AI systems in executing the above engagements.

Arguably, a medical AI system renders services due to its intangible quality. Moreover, as with physicians and attorneys who are regarded as service providers, a producer of medical AI system is also considered to be a service provider, because they are experienced professionals who assign their knowledge as an expert into intangible software codes.

Based on these definitions under the CPA, it is clear that essentially all dealings in the healthcare sector between healthcare providers, importer, distributor, manufacturers, vendors or patients, will be regarded as a transaction under the CPA. AI systems, if used in healthcare settings, would fall within the ambit of “goods” in terms of the wide definition accorded. Any activities rendered by healthcare practitioners or producers in the design or implementation of AI medical systems as

669 Section 1.
670 Section 1.
671 Section 1.

part of their professional duties, would be deemed as services. However, the CPA envisages services as being performed by healthcare providers. This limitation would create difficulties in terms of the strict liability enforced by the CPA for performance of medical services performed by AI that causes harm to patients and the scope of the definition for services in the CPA would need to be expanded, to include the sale and performance of services by AI systems.

4.3.7.2. Strict liability for goods

Some aspects of common law concerning consumer interests and rights can be said to have been codified in the CPA. Supplementing the common-law remedy for delict:

[E]very consumer has a right to receive goods that are reasonably suitable for the purposes for which they are generally intended, are of good quality; in good working order and free of any defects.⁶⁷²

In terms of services, “the consumer is also entitled to the performance of services in a manner and of a quality that persons are generally entitled to expect”.⁶⁷³ The Act further states that a “producer or importer, distributor or retailer of goods” is accountable in respect of damage or injury ensuing on account of supplying “unsafe goods, a product failure, defect or hazard in goods, or insufficient instructions or warnings to the consumer relating to any hazard arising from or associated with the use of the goods”, regardless of whether such damage or injury was due to the negligence of such parties.⁶⁷⁴

This provision deviates from the legal position that existed before the enactment of the CPA, as prior to that, a consumer only had recourse in terms of contract or delict law against a manufacturer responsible for the product that caused harm to the consumer. Based on the delictual remedies, the consumer is required to demonstrate fault.⁶⁷⁵ The consumer may find it problematic to demonstrate fault relating to the manufacturer, where fault is absent in the production stages, or where the manufacturer is not easily identifiable, or where the purchaser was not privy to

672 Section 54(1)(c), read together with section 55(2).

673 Section 54(1)(b).

674 Section 61.

675 Fault refers to blameworthiness or culpability consisting of either intent or negligence.

the production process, particularly when it comes to complex aspects of technology.⁶⁷⁶

In 2003 the SCA in *Wagener v Pharmacare Ltd*,⁶⁷⁷ held that the acknowledgement relating to strict product liability would be the assignment of the lawmakers. With the enactment of the no-fault liability in the CPA, the “producer or importer, distributor or retailer of goods” are deemed to include an implied warranty regarding the standard of goods. Thus, to succeed in a product-defect dispute, the claimant need only demonstrate that the product was defective, dangerous, or did not encompass sufficient warning relating to a risk and ensued in damage.

The CPA thus allows for the strict (no-fault) liability, in respect of producers or sellers of AI systems used in healthcare, due to a defective product or if there are insufficient directions or cautionary notices to the purchaser relating to any danger or threat related to the goods, notwithstanding fault on the part of such professional. The CPA also provides that,

[a] supplier of services, who in conjunction with the performance those such services, applies, supplies, installs or provides access to any goods, must be regarded as a supplier of those goods to the consumer.⁶⁷⁸

Thus, strict liability may apply to a healthcare practitioner who applies AI in the rendering of healthcare services. Since a healthcare practitioner provides the care and is the most obvious person to be recognised and depicted in the supply chain, under the CPA, they can be held strictly accountable if the AI system is defective and harm ensues. However, the patient may elect to take legal action against either the “producer, importer, distributor or retailer” (inclusive of the healthcare practitioner) or all of them jointly or severally, if these parties are part of the supply chain. A causal connection must still be established between the defective goods and resultant damage⁶⁷⁹ on a preponderance of probabilities, but the common-law onus necessitating evidence relating to negligence is not required. As indicated

676 Jacobs W, Stoop PN and Van Niekerk R “Fundamental consumer rights under the Consumer Protection Act 68 of 2008: A critical overview and analysis” <http://www.scielo.org.za/pdf/pej/v13n3/v13n3a09.pdf> (Date of use: 5 August 2020).

677 2003 4 SA 285 (SCA) [298-300].

678 Section 61(2).

679 Refer to the discussion on factual and legal causation required for negligence.

earlier, the definition of services under the CPA would need to be extended to include healthcare services being rendered with the assistance of AI systems to hold those in the supply chain responsible.

The CPA contains defences to the “no-fault” liability structure. Liability will not occur if, amongst others, “(a) the unsafe product characteristic, hazard, failure or defect is the result of compliance with any public regulation; (b) the alleged unsafe product characteristic, hazard, failure or defect did not exist in the goods at the time they were supplied to another person alleged to be liable; (c) it is unreasonable to expect the distributor or retailer to have detected the unsafe product characteristic, failure, defect or hazard; and finally, (d) if the claim for damages is brought more than three years after the death or injury of a person[.....]”.⁶⁸⁰

Notably the CPA provides that, should goods be distributed in South Africa under a transaction which would be excluded under the CPA, such “goods, and the importer or producer, distributor and retailer of those goods, respectively”, would remain subjected to strict product liability.⁶⁸¹ Thus, even if AI systems were deemed to be excluded under the CPA, a consumer would still have recourse against the “importer or producer, distributor or retailer” in terms of strict product liability.⁶⁸²

In conclusion, actions for liability of defective AI systems causing harm in South African law can be invoked either on negligence under delict where there is wrongful and blameworthy conduct that ensues in harm or based on product liability (strict liability) contained under the CPA where no fault is required on part of the wrongdoer and the burden that rests with the plaintiff to demonstrate negligence is alleviated.

For the standard of care expected in delictual negligence, specialists are subjected to a greater standard of care – being that of a reasonable expert relevant to the occupation. A malpractice lawsuit could follow in the event of a departure from this standard. Generally, malpractice claims are common to physicians, although many other professions are liable to the same rules. With AI systems, a malpractice lawsuit

680 Section 61(4).
681 Section 5(5).
682 Section 61.

is likely to first be made against the healthcare provider as the professional user who used the system and secondly, against the producers of a defective AI system.

In performing his professional duties of rendering healthcare services, the physician is deemed to have the expertise and skill normally held and applied by colleagues of the profession that are considered to be in good standing. It is therefore required of physicians to practice reasonable care to ensure that all facts upon which their recommendations and judgement are founded, are sound. This extends to cases where the physician bases his or her decision on information derived from an AI system, and where the physician is under an obligation to make sure that such a system is safe and dependable.

This obligation can be achieved through, for example, ensuring that the AI system undergoes testing and certification through quality assurance procedures (such as the ISO series), that can be applied as best practice. By not ensuring that a medical AI system is safe and reliable, or failing to detect an outcome containing an error which would have been palpable to another in the same occupation, would evidence a lack of care giving rise to negligence.

Computer malpractice lawsuits are to be permitted against the producers of defective medical AI systems, if they do not exercise the level of skill and care of an average reasonable expert from the same industry. However, the complications that arise when it comes to claims regarding computer malpractice, relate to the lack of a standard against which experts in such industry should be judged; swift advancements in the technological arena, particularly in AI resulting in any standard becoming archaic in only a few years, and finally, the challenge in determining which of the profession's standards should be recognised in light of the varied classes of professionals that contribute to the design of AI systems.

4.4. A comparative analysis on delictual liability arising from AI systems

AI-driven solutions are transforming the traditional approach to service delivery, including the way professional healthcare is rendered. AI exemplifies skilled or professional practices by way of consumer related goods, which is readily offered on the common market and for most part, lacks oversight. The probability of harm

is enormous, particularly because many AI systems are now being implemented in healthcare services, a domain with high risk of injury.

The South African legal system has been slow to recognise in general “computer and information technology law” as a fundamental discipline. Consequently, issues concerning AI liability remain dormant in the judicial precedents of South Africa to date. Thus, the approach adopted by other legal systems can assist with legal solutions for South Africa’s systems, regarding the universal ethico-legal challenges that AI presents. Civil liability emanating from the application of AI systems has already been acknowledged and remarked upon at length in Anglo-American technology regulation.⁶⁸³ For the objectives of this thesis, the applicable frameworks in the Anglo-American legal systems of the US and UK are analysed.

It is emphasised that, as there are thus far no noteworthy decisions on liability in delict emanating from the application, and the use of AI in the foreign legal systems considered, the ultimate deductions reached remain presumed and qualified, grounded on the analysis of the foreign legal approaches.

4.4.1. United Kingdom

The English tort is established on the common law and a civil wrong.⁶⁸⁴ The most important feature of a tort, is that it infringes a general obligation enforced by law and it is a private right of action which offers redress in the form of damages to an injured party.⁶⁸⁵

Torts embody legally recognised damages that are actionable within the special liability rules of wrongdoing.⁶⁸⁶ There are three categories of torts: (1) intended

683 Turley 1988 *Computer law journal*; Tuthill 1991 *AI expert*; Reed and Angel *Computer law*; Callaghan D and O'Sullivan C “Liability for software errors” 2005 *Computer Law & security report*; Gerstner 1993 *Santa Clara law review*; Scherer 2016 *Harvard journal of law and technology*.

684 Delict is used South Africa and it is known as “tort” in Anglo-American law.

685 Fleming JG *The law of torts* 8th ed (Law Book Co Sydney 1992) 1.

686 Fleming *The law of torts* 1.

violation of a claimant's legally recognised rights; (2) negligence; and (3) product liability.⁶⁸⁷ For this study, intentional violation of rights is not considered.⁶⁸⁸

4.4.1.1. Negligence

For tortious liability, producers of defective AI systems would be negligently liable both due to an infringement of the duty of care and strict product liability. As stated earlier, liability premised on negligence and strict product liability depend on whether an AI system is a product or a service.⁶⁸⁹ It is based upon whether AI-based software comprises of "goods" in context of a "product" described under the common law⁶⁹⁰ and the UK CPA.⁶⁹¹ Where software is regarded as a product, principles relating to product liability would be relevant, but if it is viewed to be a service, liability principles for negligence will be fitting.

The tort of negligence gained prominence following the verdict of *Donoghue v Stevenson*⁶⁹² where the court determined that "duty of care", "breach of duty" and "breach causing harm" have to be established for the purpose of proving liability when it comes to negligence. In this case, to establish if a duty of care existed Lord Atkin famously referred to the neighbour test:⁶⁹³

You must take reasonable care to avoid acts or omissions which you can reasonably foresee would be likely to injure your neighbour. Who, then, in law is my neighbour? The answer seems to be persons who are so close and directly affected by my act that I ought reasonably to have them in contemplation as being so affected when I am directing my mind to the acts or omissions which are called in question.

4.4.1.2. Breach of the duty of care

Where there is a required duty of care, a claimant must then demonstrate that this duty was violated. The test is to establish whether the wrongdoer behaved less competently than a "reasonable person" would have in the same situation, being the

687 Fleming *The law of torts* 15.

688 Refer to section 4.3.

689 See section 4.3.5.1.

690 In common law, the manufacturer's liability was used for various products, including products which are not deemed to be justifiably safe to persons, health or property, see Fleming *The law of torts* 483.

691 British Consumer Protection Act 1987.

692 (1932) UKHL 100 562.

693 [580].

so-called objective test.⁶⁹⁴ Reasonable care depends on the kind of the risk involved taking in account the circumstantial facts of each matter.⁶⁹⁵

Thus, when applying a standard of care in one case, it will not automatically become a precedent for another, as each case is adjudicated on its own facts. As with the law of delict in South Africa, the yardstick for the standard of reasonable care is that of the reasonable person in the same situation as that of the offender. Where a professional such as a doctor is involved, the standard of care is centred on the way a “reasonable doctor” would have behaved under the same conditions.⁶⁹⁶

A professional is deemed to hold a minimum standard of specialist knowledge and training, and who performs services requiring specialised expertise.⁶⁹⁷ In *Carr v Inland Revenue Commissioners*,⁶⁹⁸ Du Parcq LJ suggested that, undertakings that are viewed as skills or talents in a specific period of time, could later be ranked as a profession. Thus, computer professionals and producers of AI systems would be regarded as professionals in the same way as doctors, who would be liable for breach on account of their profession – known as medical malpractice.

The general standard applied to professionals is therefore that of a “reasonable professional” of that profession.⁶⁹⁹ Bingham LJ (dissenting in *Eckersley v Binnie*)⁷⁰⁰ provided a comprehensive view of the degree of knowledge required for a professional individual, namely:

[...] a professional man should command the corpus of knowledge which forms part of the professional equipment of the ordinary member of his profession. He should not lag

694 *Blyth v Company Proprietors of the Birmingham Water Works* (1856) 11 Ex Ch 781. The court held as follows: “Negligence is the omission to do something which a reasonable man, guided upon those considerations which ordinarily regulate the conduct of human affairs, would do, or doing something which a prudent and reasonable man would not do. The defendants might have been liable for negligence, if, unintentionally, they omitted to do that which a reasonable person would have done, or did that which a person taking reasonable precautions would not have done.”

695 *Nettleship v Weston* (1971) 2 QB 691; *Overseas Tankship (UK) Ltd v The Miller Steamship Co or Wagon Mound No 2* 1967 1 AC 617.

696 *Bolam v Friern Hospital Management Committee* (1957) 1 WLR 582.

697 Tuthill 1991 *AI expert* 50.

698 (1944) 2 All ER 163.

699 In *Bolam v Friern Hospital Management Committee* (1957) 1 WLR 582 McNair J held: “The test is the standard of the ordinary skilled man exercising and professing that special skill; it is well established law that it is sufficient if he exercises the ordinary skill of an ordinary competent man exercising that particular art.”

700 1988 18 Con LR 1 at 80.

behind other ordinary and assiduous and intelligent members of his profession in his knowledge of new advances, discoveries and developments in his field. He should have such awareness as an ordinarily competent practitioner would have of the deficiencies in his knowledge and the limitations on his skill. He should be alert to the hazards and risks in any professional task he undertakes to the extent that other ordinarily competent members of the profession would be alert. He must bring to any professional task he undertakes no less expertise, skill and care that other ordinarily competent members of his profession would bring but need bring no more. The standard is that of the reasonable average. The law does not require of the professional man that he be a paragon combining the qualities of a polymath and a prophet.

Producers will thus be in breach of their duty if they did not apply the same degree of care in producing an AI system, as that of a reasonable man of the same skills under the same conditions. The foreseeability of risk materialising,⁷⁰¹ or the determination on whether a wrongdoer complied with standard practice in the situation, are elements in establishing if there was any breach of duty. Where risks are foreseeable, the risk must be offset against the cost or burden needed to prevent the risk.

Breach arises if the expense of implementing preventative actions, is overshadowed by scale of the risk and severity of potential injury or loss.⁷⁰² While a case of negligence may not arise in cases involving minimal risk or minor damage, the test for negligence is elevated with regard to advanced technological and/or high-risk designs.⁷⁰³

With regards to the question of whether the defendant conformed to common practice, there is no universal principle which could be applied for the various professions, as many professions have had a tendency to create their own specialised regulatory standards. The standard of care when an expert is involved would be one of a reasonable expert in the same profession.⁷⁰⁴

To determine if a duty of care exists for professionals, consideration ought to be given to standards created by regulation to assess their conduct that give due regard to the nuances of particular professions. The key feature of a profession relates to

701 *Bolton v Stone* (1951) AC 850.

702 *Latimer v AEC* (1952) 2 QB 701.

703 *Independent Broadcasting Authority v EMI Electronics* (1980) 14 Build LR 1.

704 *Bolam v Friern Hospital Management Committee* (1957) 1 WLR 582: "The test is the standard of the ordinary skilled man exercising and professing that special skill. It is well established law that it is sufficient if he exercises the ordinary skill of an ordinary competent man exercising that particular art."

its knowledge base and it is based on the social and cultural assessment of, and the approaches for, managing that knowledge base in the case of computer technology.⁷⁰⁵

According to Diane Rowland, the dependence by consumers on a specific knowledge base could illustrate a method of distinguishing professions against that of others and this point has been rather persuasive in various decisions⁷⁰⁶ on professional negligence.⁷⁰⁷ Rowland and Rowland subdivide the kinds of professions between: (a) professionals who are not able to provide assurance of their results or labour, for instance the medical or legal profession; and (b) professionals who by implication guarantee to provide a specific result, for instance architects.⁷⁰⁸ With regard to category (b) which would apply to producers of software, they are deemed to have a certain knowledge base of the outcomes he produces, and by virtue of this, it suggests a higher degree of care is required.

It however remains challenging to test complex technology exhaustively, or to guarantee that they will be 100 per cent error free, and may stay undetected for some time.⁷⁰⁹ Producers must, however, perform a risk assessment of the functional area and must employ appropriate methods to reach the desired integrity status of the final product.⁷¹⁰ A duty of care can be violated in various ways where AI systems malfunction, such as faults in the programming design that the developer could have noticed; an improper or insufficient knowledge base inputted by the domain expert; insufficient manuals or warnings from the manufacturer; not updating the knowledge base, or the end-user unduly relying on the output of the system. A way in which to achieve safety, is through devising safety standards for the construction and

705 Rowland D “Negligence, professional competence and computer systems” https://warwick.ac.uk/fac/soc/law/elj/jilt/1999_2/rowland/#Carr (Date of use: 7 August 2020).

706 *Hedley Byrne v Heller and Partners* (1964) AC 465; *Home Office v Dorset Yacht Co Ltd* (1970) AC 1004 and *Anns v Merton London Borough Council* (1978) AC 728.

707 Rowland https://warwick.ac.uk/fac/soc/law/elj/jilt/1999_2/rowland/#Carr (Date of use: 7 August 2020).

708 Rowland and Rowland 1993 *Law, computers and artificial intelligence* 229-243.

709 Rowland and Rowland 1993 *Law, computers and artificial intelligence* 237.

710 Rowland and Rowland 1993 *Law, computers and artificial intelligence* 244.

application of AI technologies. In *Bevan Investments v Blackhall and Struthers*,⁷¹¹ it was held that,

a design which departs substantially from relevant engineering codes is *prima facie* a faulty design unless it can be demonstrated that it conforms to accepted engineering practice by rational analysis.

Certification issued by a qualified and well-known regulating agency confirming that, a notable quality assessment and control system such as ISO series was applied during the development of an AI system, may serve as persuasive evidence that reasonable care was exercised by the producer to prevent harm to a user, and consequently non-compliance constitutes evidence of negligence.⁷¹²

The International Electrotechnical Commission (IEC), publishes standards similar to that of the ISO series, but concerning electrical technologies for engineers working on safety-related systems.⁷¹³ If the producer of AI systems is a member of a recognised profession, preferably government-certified such as software or computer engineering, and does not comply with the codes of conduct or guidelines set by the profession when developing an AI product, this could evidence that a duty of care was breached by a producer, although it does not guarantee the competence of the individual.⁷¹⁴

In the UK, guidance concerning the duties and associated accountability of software engineers, is obtainable from the Engineering Council that encompasses code of good practice for engineers and supervisors working with and are responsible for safety-related technologies.⁷¹⁵ Liability is premised upon the experience and skills of the professionals concerned. However, the current problem is that globally-

711 No 2 (1973) 2 NZLR 45.

712 As discussed in section 4.3.5.2 the adoption of product safety standards like that of ISO 9000 and 9001 series, could serve an important function in establishing product safety.

713 IEC “Standards” <https://www.iec.ch/standardsdev/publications/is.htm> (Date of use: 7 August 2020).

714 Rowland https://warwick.ac.uk/fac/soc/law/elj/jilt/1999_2/rowland/#Carr (Date of use: 7 August 2020).

715 UK Engineering Council “Guidance for institution codes of conduct” <https://www.engc.org.uk/engcdocuments/internet/website/Guidance%20for%20Institution%20Codes%20of%20Conduct%202017.pdf> (Date of use: 7 August 2020).

recognised and established codes of good practice and rules, are currently not available for complex technology such as AI.

This may be largely due to the fact that research and development implications for novel designs, are at risk of being subdued when the designs are subjected to negligence without argument. Ultimately, the integrity of any new technology is largely attained by employing quality assurance standards during design and application, and producers have a duty to guarantee the adequate safety of AI systems.

Professional negligence claims against the producers of AI systems, will be based on the fact that they have not exercised the standard of reasonable care during the development of the AI system, which inflicted harm. Similarly, an action of medical malpractice against a healthcare provider due to medical errors emanating from the use of, and reliance on, a faulty AI system will be due to the fact that the healthcare professional omitted to take due care when using the system in his professional duties. To avoid culpability for negligence, these actors will have to provide evidence that they exercised every reasonable care to prevent the harm.

4.4.1.3. Breach causing damage – causation and foreseeability

For liability in negligence to arise, there must be an infringement of a duty of care that has resulted in harm. Actual harm is required in tort for a claim of negligence, and in the absence of actual damage, no cause of action ensues.⁷¹⁶ Harm includes bodily and mental impairment, property loss, or even pure economic loss under certain circumstances.⁷¹⁷ An infringement of the duty of care must ensue in damage.

The UK law relating to negligence, requires a causal nexus between the act of negligence and injury suffered. Causation consists of two parts, factual causation and legal/proximate causation. Factual causation has to do with the determination

716 *Smith v Leech Brain & Co* (1962) 2 QB 405.

717 Fleming *The law of torts* 191.

on whether [but for] the defendant's blameworthy act would the harm have resulted, for which the well-known test of the *conditio sine qua non* is employed.⁷¹⁸

If, on a preponderance of probabilities, damage would or could have ensued irrespective of the wrongdoer's negligence, he will not be liable for having acted negligently. Because AI systems in certain ways can be regarded as solely information centred, there may be conceptual problems with the idea that AI can cause damage – being an essential element for negligence.

During the 1980s, due to shortcomings relating to software introduced into a radiotherapy system called Therac-25, several patients were exposed to high levels of radiation resulting in serious sickness and death.⁷¹⁹ There was no real difficulty with establishing the causal nexus between the defective software and the ensuing harm. However, there is also the scenario where the patients could have been subjected to under-doses of the radiation, and would not have been harmed due to the radiation, but instead from an increased degree of cancer which was being treated with the assistance of the radiotherapy system. The competence of the designers could still be questioned in the latter scenario, but they would still escape liability for negligence.⁷²⁰

On second part of legal causation, it must be determined to what degree the defendant ought to be liable for the consequences of their conduct, and this is based on policy considerations.⁷²¹ In the case of defective AI systems, the difficulty is that given the various human actors involved in the design and application thereof, it may be difficult or unreasonable to assign blame entirely to one party.

Where the AI system is used to make decisions, or recommends a course of action in a particular situation, there is at least one other actor involved, making it difficult

718 *Barnett v Chelsea & Kensington Hospital Management Committee* (1968) 2 WLR 422.

719 Leveson NG *Safeware: System Safety and Computers* (Reading, Mass: Addison-Wesley Crawfordsville 1995).

720 Rowland https://warwick.ac.uk/fac/soc/law/elj/jilt/1999_2/rowland/#Carr (Date of use: 7 August 2020).

721 *Lamb v Camden LBC* (1981) QB 625. In this case Lord Denning held: "The truth is that all these three – duty, remoteness and causation – are all devices by which the courts limit the range of liability for negligence [...]. All these devices are useful in their way. But ultimately it is a question of policy for the judges to decide".

to establish causation. For instance, the knowledge base may have been provided by the domain expert, it was improperly executed by the knowledge expert, or it was incorrectly programmed by the software programmer. Thus, when it comes to causation and apportionment of damage, it will be essential to establish the roles and contributions of each human actor in relation to AI systems that cause harm. Thus, causation may be easier to prove in instances where the AI system itself goes beyond recommending an action, but where it takes the action itself.

A defendant is only liable for harm that was reasonably foreseeable. It must be established if the nature of the harm caused was reasonably foreseeable at the time of the defendant's conduct.⁷²²

4.4.1.4. Duty of care

In *Donoghue*, it was stated that with regard to a manufacturer of food or medicine or the like, the manufacturer has a legal obligation towards a consumer to apply reasonable care, so as to safeguard against the merchandise having defects that may ensue in personal injury or harm.⁷²³ For negligence, there must be a direct nexus between the defendant and plaintiff where that injury or harm should have been reasonably foreseen.

Where there is personal injury or property damage, the presence of a direct relationship between the offender and complainant is established by virtue of such injury or damage having occurred.⁷²⁴ Where the extent of possible harm is excessive resulting in damages than a reasonable person is required to employ measures to avoid it.⁷²⁵ Where pure economic losses are claimed,⁷²⁶ it may not be recoverable due to policy considerations – despite an existing nexus creating a duty of care.⁷²⁷

722 *Hughes v Lord Advocate* (1963) UKHL 31. It was held by the court in this matter that even where the damage is not itself foreseeable, liability ensues on the proviso that the actual loss is within a “foreseeable class of harm”.

723 [562].

724 *Caparo Industries Plc v Dickman* (1990) UKHL 2; *Overseas Tankship (UK) Ltd v Morts Dock and Engineering Co Ltd or Wagon Mound (No. 1)* (1961) AC 388.

725 *Overseas Tankship (UK) Ltd v The Miller Steamship Co or Wagon Mound (No. 2)* (1967) 1 AC 617.

726 “Pure economic loss” involves financial loss that excludes personal harm or damage to property; Fleming *The law of torts* 173; Neethling, Potgieter and Visser *Law of delict* 280.

727 *Rondel v Worsley* 1969 1 AC 191.

The producers of AI systems, therefore, have a duty of care to implement systems that do not cause injury to the end-

user, or damage their property.⁷²⁸ Although for a specialist application, the producer of AI systems will in most cases have a contract with the consumer. They could also be liable in tort – not only to the consumer, but also to those whose safety and health are affected on account of the system being defective.

Tort negligence thus permits recovery if a defect can be proven to result from the absence of due care by the producers' when creating or applying the system. Similarly, a professional end-user, such as a physician, must take due care that in the application or use of the AI system, no patients are harmed. If such physician uses the system as part of his professional duties, and renders advice based on the output generated by the system, it will be accepted as complimentary to the task that the physician performed. What must be determined is whether reasonable care was employed by the physician in performing such a task.⁷²⁹

If it is demonstrated that the physician had knowledge or ought to have known that a system employed by him was faulty in design or function, in trusting such a system any erroneous outcomes generated by it may suggest that the physician acted negligently. According to Reed,⁷³⁰ these situations offer such a compelling suggestion that a task was performed with negligence, that the *res ipsa loquitur* doctrine must apply. In the matter of *The Lady Gwendolen*,⁷³¹ a ship loaded with cargo set sail at maximum velocity in heavy fog.

Although the ship was equipped with new technological radar, the controller who was experienced and competent only monitored the radar occasionally, and the radar settings were incorrect. The court held that the new radar technology was negligently applied, that the owners of the ship were at actual fault, and their liability could not be limited as they had failed to train their workers in the use of the new

728 Reed and Angel *Computer law* 87-91.

729 Reed and Angel *Computer law* 87-91.

730 Reed and Angel *Computer law* 87-91.

731 1965 3 WLR [294-296].

technology. In *Montgomery v Lanarkshire Health Board*,⁷³² it was held that, all healthcare practitioners have an unequivocal duty to act with reasonable care, to ensure that the risks of a treatment are explained to their patients so that they appreciate the consequences.

4.4.1.5. Onus of proof

For negligence claims based on tort, a claimant has to prove the elements of negligence in question – which includes demonstrating a *prima facie* case relating to the defendant's negligence.⁷³³ However, where the reasons for the harm remains entirely within the offender's influence, the dictum referred to as *res ipsa loquitur* ("the facts speaks for themselves"), will apply.⁷³⁴ In terms of this maxim, the court can infer or presume negligence from the facts of a matter where no direct evidence exists as to the cause of the incident.⁷³⁵

In situations where it is impossible to demonstrate negligence, but the facts of the case persuasively implicate the offender, a court can on a presumption of negligence rule for the claimant – unless the offender tenders evidence to rebut such a presumption. The maxim finds its origin in *Byrne v Boadle*,⁷³⁶ where the plaintiff as a pedestrian was struck and rendered unconscious by a container of flour which fell out of the window of a warehouse. A witness testified that he saw the barrel fall, but had no knowledge as to how it fell.

The plaintiff was unable to provide evidence proving that the barrel fell as a result of the defendant's negligence. The lower court dismissed the claim as there was no *prima facie* case. The higher court overturned the ruling of the lower court, invoking *res ipsa loquitur* on the basis that the circumstances of the incident in question, strongly suggested negligence. This presumption of negligence can be challenged if there is evidence to suggest that the incident was not due to the defendant's negligence, or that reasonable care was employed. If the defendant is successful in

732 (2015) UKSC 11.

733 Fleming *The law of torts* 314.

734 This maxim would apply for product liability if a manufacturer is entirely in charge of the manufacturing procedures.

735 Fleming *The law of torts* 314.

736 2 H. & C. 722, 159 Eng. Rep. 299 (Exch. 1863).

rebutting the presumption of negligence, the duty shifts to the plaintiff to demonstrate negligence.⁷³⁷ Therefore, the maxim of *res ipsa loquitur* serves as a method for achieving a result similar to the no-fault based negligence for strict liability.

4.4.1.6. Contributory negligence

Contributory negligence may be raised as a mitigating defence against legal action for damages ensuing from AI systems, which reduces the offender's accountability concerning damages. A claim can be decreased based on a "just and equitable" value relative to the ratio of the plaintiff's contribution to the damage endured.⁷³⁸ The court will compute the actual damages suffered, and offset the damages payable by the offender equal to the plaintiff's involvement in same.

Contributory negligence is, furthermore, effective as a justification where the plaintiff was entirely at fault.⁷³⁹ Given the amount of participants engaged in the construction and application of AI systems, it is further possible for the defendant to only be liable to the extent of his or her respective degree of negligence, in relation not only the plaintiff's, but other defendants' contribution to the loss or damage. This applies to the position where more than one wrongdoer is individually or collectively responsible in causing the same damage or loss.⁷⁴⁰

4.4.1.7. Product liability

While negligence cases in tort serve as the groundwork for liability standards, many other areas of tort have become distinctive through statutory reform. Following the American example,⁷⁴¹ most countries have chosen to reform their laws to deal with

737 Fleming *The law of torts* 325; *Byrne v Boadle* 2 H. & C. 722, 159 Eng. Rep. 299 (Exch. 1863).

738 Fleming *The law of torts* 268-269.

739 In *Jayes v IMI Kynoch* (1985) ICR 155 where the court found that the fault was 100 per cent that of the plaintiff.

740 Fleming *The law of torts* 255.

741 The US product liability regime is a leading model on strict liability relating to defective products, with many countries who follow the product liability regime adopting the US model. See Reimann M "Liability for defective products at the beginning of the twenty-first century: Emergence of a worldwide standard" 2003 *The American journal of comparative law* 751.

product liability using legislative approaches.⁷⁴² After concluding that the negligence system was not adequately addressing consumer concerns, the UK resorted to adopting strict liability structures, and consequently the UK Consumer Protection Act of 1987 (CPA)⁷⁴³ was passed.⁷⁴⁴

The UK CPA establishes a civil law right of redress damage resulting from the use of defective products (i.e., “product liability” terms). It seeks to protect the consumer from products that are considered to not pass a satisfactory degree of safety. The damage must be due to a defect relating to the product.⁷⁴⁵ Thus, the UK CPA imposes a strict product liability on producers or sellers offering products that are defective and that cause damage, irrespective of no fault on the side of such actors.⁷⁴⁶ The liable parties in terms of the UK CPA are the “producer or anyone who holds himself out to be the producer, or the importer of the product into the EU or against the supplier”.⁷⁴⁷

The supplier will only be liable if they are not willing upon request, to provide the identity of the producer or importer within an acceptable period after a loss or injury arises.⁷⁴⁸ Such liability is not capable of exclusion in respect of the injured parties, or their dependents by way of contractual provisions.⁷⁴⁹ Whilst negligence is concerned with the actions of the parties in determining liability, product liability under the UK CPA focuses on the defective condition in relation to product placed in the market by the seller. The UK CPA defines “product” as:

Any goods or electricity and includes a product which is comprised in another product whether by virtue of being component parts or raw materials or otherwise.⁷⁵⁰

This suggests that products must be tangible. Based on this broad definition of product, it suggests that AI systems may be incorporated in the UK CPA, on the

742 Reimann 2003 *The American journal of comparative law* 751.

743 The Act is available at <https://www.legislation.gov.uk/ukpga/1987/43/contents> (Date of use: 8 August 2020).

744 Part I of the Act implements into UK law the provisions of the Product Liability Directive of the Council of the European Communities dated 25 July 19 (85/374/EEC).

745 Section 2(1).

746 Section 2.

747 Section 2(1). Under section 1(2) the “producer” is defined to be a manufacturer.

748 Section 2(3).

749 Section 7.

750 Section 1(2).

provision that it encompasses a tangible aspect and it was acquired in the same way as products, such as a sale, rental or lease and not as part service. The marketing of software is not inherently different to other items of commerce. For example, software downloaded from the internet will lack necessary tangibility to qualify as a product.⁷⁵¹

Where such software is assembled into a product, or becomes a component or unprocessed substance of a product, and causes such a product to operate defectively, the product as a whole will concern product liability in accordance with the CPA. With regard to this, the CPA states that,

there is a defect in a product for the purposes if the safety of the product is not such as persons generally are entitled to expect; and for those purposes 'safety', in relation to a product, shall include *safety with respect to products comprised in that product* [...].⁷⁵²

A product is regarded to be defective where it fails to provide the degree of safety that individuals generally have the right to demand considering the circumstances, intended purposes of promoting the product, appearance of the product, directions or cautionary instructions regarding the product's application or handling, the expectation relating to the purposes of the product, and the period relating to the release of the product into the supply chain.⁷⁵³ This test broadly covers manufacturing as well as design errors in the product. The CPA contains the defences which consist of the following:⁷⁵⁴

- The compliance defence: that the defect is attributable to compliance with any requirement imposed by or under any enactment or with any community obligation;⁷⁵⁵
- The non-supply defence: that the person proceeded against, did not at any time supply the product to another (e.g., a stolen product);⁷⁵⁶

751 Reed and Angel *Computer law* 114.

752 Section 3(1).

753 Section 3(2).

754 Section 4.

755 Section 4(1)(a).

756 Section 4(1)(b).

- The non-profitable defence: that the supply of the defective product was not made in the course of business” for commercial objectives;⁷⁵⁷
- The non-existence defect defence: that the defect did not exist in the product at the relevant time of supply;⁷⁵⁸
- The unforeseeable-defect defence: that the state of scientific and technical knowledge at the relevant time was not such that a producer of products of the same description as the product in question might be expected to have discovered the defect if it had existed in his products while they were under his control;⁷⁵⁹
- The subsequent-product defence: that the defect constituted a defect in a product (“the subsequent product”) in which the product in question had been comprised, and was wholly attributable to the design of the subsequent product or to compliance by the producer of the product in question with instructions given by the producer of the subsequent product.⁷⁶⁰

The CPA provides for a strict (no-fault) product liability of producers or sellers of AI systems used in healthcare for defective products, in contrast to the conventional fault-based liability.⁷⁶¹ Thus, the producer or seller are deemed to include an implied warranty of quality. Under the CPA, to succeed on a product liability (no-fault) complaint, the claimant is to only prove that the products were faulty or defective, and the existence of a causal link between the defective product and the ensuing harm. This reduces the burden of the claimant to establish negligence relating to the production or implementation process. Liability is concerned with personal injury, death or property damage ordinarily.⁷⁶² No compensation is allowed for pain and suffering or pure economic loss as it can be recovered under the normal tort liability regime.⁷⁶³

757 Section 4(1)(c).

758 Section 4(1)(d).

759 Section 4(1)(e).

760 Section 4(1)(f).

761 This supports the maxim of “*res ipsa loquitur*” which denotes an assumption of negligence.

762 Section 5(1) and 5(3).

763 Section 5(2).

In keeping with the notion of regulating medical devices as “products” under the CPA, the question is where AI medical devices would be regulated under the UK’s Medical Devices Regulations (MDR). The country has adopted the EU’s regulations for medical devices regulated in terms of the Medicines and Healthcare Products Regulatory Agency (MHRA). The MDR⁷⁶⁴ entered into force on 25 May 2017, and is anticipated to transitionally come into full implementation by 26 May 2021, which significantly contributes towards a regulatory structure for medical devices.⁷⁶⁵ Similar to the South African definition, medical device in the MDR is defined as:

“Medical device means any instrument, apparatus, appliance, software, implant, reagent, material or other article intended by the manufacturer to be used, alone or in combination, for human beings for one or more of the following specific medical purposes:

- diagnosis, prevention, monitoring, prediction, prognosis, treatment or alleviation of disease,
 - diagnosis, monitoring, treatment, alleviation of, or compensation for, an injury or disability,
 - investigation, replacement or modification of the anatomy of a physiological or pathological process or state,
 - providing information by means of in vitro examination of specimens derived from the human body, including organ, blood and tissue donations,
- and which does not achieve its principal intended action by pharmacological, immunological or metabolic means, in or on the human body, but which may be assisted in its function by such means.

The following products shall also be deemed to be medical devices:

- devices for the control or support of conception;
- products specifically intended for the cleaning, disinfection or sterilisation of devices as referred to in Article 1(4) and of those referred to in the first paragraph of this point”.⁷⁶⁶

The new regulatory framework introduces guidelines for medical software that are regarded as a device. The regulation defines medical software devices as:

Software, which drives a device or influences the use of a device, shall fall within the same class as the device. If the software is independent of any other device, it shall be classified in its own right.⁷⁶⁷

An example of the former is that software that operates for a general purpose in a healthcare setting, or software for health fitness applications that function as

764 Regulation (EU) 2017/745 of the European Parliament and of the Council <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32017R0745&from=EN> (Date of use: 27 August 2020).

765 EU regulations for Medical Devices <https://www.ema.europa.eu/en/human-regulatory/overview/medical-devices> (Date of use: 13 August 2020),

766 Article 2(1).

767 Annex VIII, Chapter II of the regulations.

recommendation tools to assist the healthcare provider to make a diagnosis, or decide on a treatment option where they intend to ultimately rely on their own knowledge.⁷⁶⁸

An example of the latter relating to independent (stand-alone) software, are those used for making recommendations for cancer diagnoses or predicting diseases that would be classified in their own rights.⁷⁶⁹ The MHRA has issued guidelines by way of various examples to indicate which stand-alone software and application could be regarded as a medical device requiring certification by the MHRA, and those which do not.⁷⁷⁰

The regulation places an obligation on manufacturers, authorised representatives of the manufacturer, importers and distributors, or persons that offer to the market a medical device in their name.⁷⁷¹ The devices must comply with all standard safety and performance conditions, and the appropriate specifications and structures necessary to ensure the safety of user.⁷⁷² This guidance contains important requirements that in all likelihood will be applicable to AI medical devices. The UK has adopted the “risk-based classification system” also used by EU countries to classify medical devices based on risk.⁷⁷³

Devices in the UK are classified as follows: “I, IIa, IIb, and III”. Class I are considered to be devices with the least risk, while Class III devices are deemed to have

768 MHRC guidance on Medical device stand-alone software including apps (including IVDMDs)
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/890025/Software_flow_chart_Ed_1-06_FINAL.pdf (Date of use: 13 August 2020).

769 According to the MHRC guidance on Medical device stand-alone software: “Decision support software is usually considered a medical device when it applies automated reasoning such as a simple calculation, an algorithm or a more complex series of calculations”
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/890025/Software_flow_chart_Ed_1-06_FINAL.pdf (Date of use: 13 August 2020).

770 MHRC guidance on Medical device stand-alone software including apps (including IVDMDs)
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/890025/Software_flow_chart_Ed_1-06_FINAL.pdf (Date of use: 13 August 2020).

771 Article 10, 11, 13, 14, and 16.

772 Annex I of the regulations.

773 Article 2 defines “risk” as “the combination of the probability of occurrence of harm and the severity of that harm.”

maximum risk concerning the device's proposed use and their innate risks.⁷⁷⁴ In order to be able to be certified for marketing by the regulatory agency (MHRA), manufacturers⁷⁷⁵ must prove that their medical devices comply with the accreditation requirements, by carrying out "an assessment of the conformity" on the device by the agency based on relevant assessment procedures.⁷⁷⁶

The risk classification of the device determines whether the assessment route is required. In terms of the assessment, approval is required for every Class of device with the exception of Class I, such as custom-made devices, which do not require a conformity assessment, and manufacturers are free to certify their products with an agency of their choice. As the classification rules under the MDR apply individually to each medical device, many devices that are AI enabled would fall under Class I under the current regime, and may under the MDR shift up a Class based on the conformity risk assessment procedures.

However, the conformity assessment procedure was designed for traditional medical devices that did not envisage the adaptive AI technology of the future (that changes on the real-time environment). The assessment and monitoring process for quality, safety and efficacy may be thwarted by having to wrestle with AI devices that have the ability to continuously update and change their output, even after they reach consumers and end-users. As with medical product regulators in other countries, the MHRA will need to shift more towards a centred regulatory assessment framework for AI based medical devices, and conceivably further impending ways where they can be merged with medical products.

In conclusion, actions for liability in English law emanating from harm caused by defective AI systems, can be made under tort (i.e negligence) or product liability under the CPA. Where an AI system is an essential component to the services offered, the producers of a defective AI system could be held liable under tort of negligence where the plaintiff would have to satisfy all the elements, i.e., a duty of

774 Article 51.

775 Article 2 defines manufacturers as "a natural or legal person who manufactures or fully refurbishes a device or has a device designed, manufactured or fully refurbished, and markets that device under its name or trademark."

776 Article 52 read with Annex IX to XI.

care owed to the plaintiff, the violation of the duty of care and damage inflicted. Where AI software is regarded to be a product, strict product liability provisions under the CPA will be appropriate where death, personal or property loss is suffered by the claimant.

4.4.2. United States

American law of torts is primarily based on English common law, which encompasses standard rules, whilst the individual states have independent rules of civil wrongs that offer redress under an action for damages. Most of the American tort law was designed by court judgments in certain cases, making it flexible. As with English law of torts, three categories of torts are identified: intentional torts, strict liability, and negligence torts.⁷⁷⁷ For purposes of this thesis, grounds (2) and (3) are relevant.⁷⁷⁸

4.4.2.1. Negligence

In American law, the most common torts involve harm caused by carelessness referred to as negligence.⁷⁷⁹ In terms hereof, “negligence is the failure to use such care as a reasonably prudent and careful person would use under similar circumstances”. To prove negligence the following elements must be satisfied:

- (1) A defendant has a duty of care towards the claimant;
- (2) A defendant breached such duty of care by his actions or his omission to act;
- (3) A breach of such duty inflicted harm upon the claimant; and
- (4) The existence of causation, where the harm inflicted is a “reasonably foreseeable consequence” of the defendant's violation of duty.⁷⁸⁰

777 Keeton WL *et al Prosser and Keeton on the law of torts* 5th ed (West Publishing Co. Saint Paul, Minnesota 1984) 152.

778 Refer to section 4.3

779 Feinman JM *Law: Everything you need to know about American law* 4th ed (Oxford University Press 2014) 152.

780 Keeton *et al Law of torts* 164. See Fleming *The law of torts* 183 where the English tort comprises of the same elements.

The abovementioned elements must be demonstrated to establish an action for negligence for injury endured, due to using a medical AI system.

4.4.2.2. Duty of care

The main notion for a claim in negligence is that reasonable care ought to be exercised in an act, with due regard to possible damage that could be foreseeably caused to persons or property.⁷⁸¹ It is only an idea and not a rule, and the courts have great flexibility in deciding when there is a duty of care.⁷⁸² Without the benefit of precedent, the determinant is predominantly a policy-based determination which will require the courts to take into account a number of considerations, including reasonable foreseeability of injury or loss.⁷⁸³ It is foreseeable that any sophisticated technology embedded with software, is bound to be affected by errors.⁷⁸⁴ The concern however, is that it is not foreseeable precisely what type of errors will occur, or what their likely effect will be on any person.⁷⁸⁵

American courts have established that the manufacturer of products has a duty of care towards the consumer.⁷⁸⁶ In *MacPherson v Buick Motor*,⁷⁸⁷ Cardozo J concluded that the vehicle manufacturer was negligently liable in tort for a faulty wheel. Following this decision, manufacturers or suppliers may be accountable based on negligence for: (1) making or neglecting to detect a mistake or error; (2) not warning consumers of any associated risks related to the products; (3) distributing a product that is defective or faulty. With respect to AI systems, there are two possible duties in relation to producers: (i) a duty to create AI systems that are safe, and (ii) a duty to provide directions to end-users on exactly the manner to safely use the AI system, and to alert them to inherent dangers associated to the system.⁷⁸⁸ In relation to healthcare providers that use the AI systems in offering

781 Feinman *Everything* 153.

782 Feinman *Everything* 153.

783 Scott 2008 *Maryland law review* 443.

784 Brooks FP Jr. *The mythical man-month: Essays on software engineering* (Addison-Wesley Crawfordsville 1995) 182. The author discusses how software emanates in technical issues on account of its inherent complexities.

785 Scott 2008 *Maryland law review* 443.

786 Keeton *et al Law of torts* 682; Fleming *The law of torts* 482.

787 (1916) 217 NY 382, 111 NE 1050.

788 Scott 2008 *Maryland law review* 443.

services to patients, their duty would be to ensure that the system functions correctly in accordance with its intended purpose.

4.4.2.3. Breach of the duty of care

The defendant's duty of care is measured against a standard of care that should be employed. If people do not act with reasonable care taking in account possible harm, they would be negligent. In earlier decisions, the American courts held that, reasonable care was tested against that of the "reasonableness" of the defendant's actions. Under negligence, the test to determine reasonableness was satisfied if the cost to eradicate the risk of harm was greater than the likelihood and scale of the risk.⁷⁸⁹ In *United States v Carroll Towing Co.*⁷⁹⁰ the court suggested a test to determine the standard of care relating to negligence. In terms of this test, a formula was created, namely whether $B < PL$. In simple terms, the test requires:

If (Burden < Cost of Injury × Probability of occurrence), then the accused will not have met the standard of care required. If (Burden ≥ Cost of injury × Probability of occurrence), then the accused may have met the standard of care.⁷⁹¹

In later decisions, the law developed another yardstick for determining reasonable care, benchmarked against the conduct of the reasonable person⁷⁹² in similar circumstances, i.e., an objective standard.⁷⁹³ The reasonable person standard, endeavoured to give some latitude to real people by having due regard to the circumstances in which the defendant, whose conduct is being judged, had to act.⁷⁹⁴

For negligence, the reasonableness in relation to the manufacturer's conduct is essential.⁷⁹⁵ In assessing the standard of care concerning AI systems, the primary issue is whether producers, as software or computer professionals, are to be held to a standard of care that is greater to that of a reasonable person. A professional

789 Nguyen FD "Regulation of medical expert systems: A necessary evil" 1994 *Santa Clara law review* 1197.

790 159 F.2d 169 (2d. Cir. 1947).

791 *United States v Carroll Towing Co* 159 F.2d 169 (2d. Cir. 1947).

792 The English law of torts also uses the reasonable person as the standard of care.

793 *Vaughan v Menlove* 1837 3 Bing NC 468; 132 Eng Rep 490; *Fancher v Southwest Missouri Truck Center Inc* (1981) Mo App 618 SW 2d 271; Keeton *et al Law of torts* 173;

794 Feinman *Everything* 159.

795 Nguyen 1994 *Santa Clara law review* 1197.

is held to a higher standard, due to their training and skills in comparison to that of the ordinary individual.⁷⁹⁶

It can be argued that in the case of software or computer professionals, their standard of care is akin to that of other professionals, for instance physicians who are exposed to malpractice claims. When any professional violates a duty of care, his conduct is to be tested against the standard of care expected of a specialist that has comparable skills, experience and training. Likewise, with regard to producers of AI systems, their duty of care is to be measured against that of software producers who possess equal skills, training and experience.⁷⁹⁷

In *Data Processing Services Inc v LH Smith Oil Corporation*,⁷⁹⁸ it was held that, an oral agreement for the development and delivery of a software program was a service contract, and that the programmer breached its implied undertaking by not producing the reasonable skill and diligence possessed by knowledgeable members of such profession.

Non-conformance to the standard of reasonable care, would constitute professional malpractice.⁷⁹⁹ Malpractice cases are particularly significant in relation to AI systems, since many of these claims originate from the healthcare profession. If the physician relies on information generated by a medical AI system, upon which he basis and makes his decisions, the physician should verify that the system is dependable. The physician's failure to establish that such system is reliable would evidence a violation of the duty of care, where negligence may be imputed.

Similarly, consideration is to be given as to whether the producers of an AI system can be sued for computer malpractice. American courts have hesitated to rule on such cases. In *Triangle v Underwriters, Inc v Honeywell Inc*,⁸⁰⁰ the claimant unsuccessfully relied on computer malpractice. Although not explicitly referenced as computer malpractice, the allegations comprised of a "failure to supervise and

796 *Prooth v Walsh* 1980 105 Misc 2d 653; 432 NYS 2d 668 (Sup); *La Vine v Clear Creek Skiing Corp* (1977) 557 F 2d 730 (10th Cir).

797 Nguyen 1994 *Santa Clara law review* 1198.

798 492 N.E.2d 314 (Ind. Ct. App. 1986).

799 Keeton *et al Law of torts* 188.

800 (1979) 604 F 2d 737 (2d Cir).

correct deficiencies in the system”, as well as the “wrongful withdrawal of support personnel.” In *Chatlos Systems Inc v National Cash Register Corp*,⁸⁰¹ a claim based on “computer malpractice” was dismissed. The court stated that,

[t]he novel concept of a new tort called ‘computer malpractice’ is premised upon a theory of elevated responsibility on the part of those who render computer sales and service. Plaintiff equates the sale and servicing of computer systems with established theories of professional malpractice. Simply because an activity is technically complex and important to the business community does not mean that greater potential liability must attach. In the absence of sound precedential authority, the Court declines the invitation to create a new tort.

A similar claim was likewise later denied in *Invacare Corp v Sperry Corp*,⁸⁰² *Hospital Computer Systems Inc v Staten Hospital*,⁸⁰³ and later in *Columbus McKinnon Corp v China Semiconductor Co Ltd*.⁸⁰⁴ According to Gerstner,⁸⁰⁵ courts have always been hesitant to enforce an expert standard of care concerning claims for computer malpractice, due to an absence of any licensing process for software programmers to serve as a measure of a basic ability, or because programming is seen to comprise of only a minor aspect of the entire software development process.

Establishing a standard of care for AI systems is further convoluted by the reality that different standards may need to be applied to the various human actors that participate in development and application of the system, to identify which one of them created the defect of error.⁸⁰⁶ However, if there is evidence to suggest that an AI system is unsuitable for a specific case as simple logic is required for the solution, or if the knowledge base provided by the domain expert or knowledge engineer is inadequate or insufficient, or if the problem is exceedingly complicated, then the design, implementation, promotion and application of the system could be regarded as negligent due to poor design and inaccurate or inadequate warnings on the part of the producers.

801 (1979) 479 F Supp 738 (DNJ) 740.

802 612 F. Supp. 448 (N.D. Ohio 1984). The court found that, workers in the computer trades must be measured against the general standard of care such as with “machinists, electricians, carpenters, blacksmiths, and plumbers”.

803 788 F Supp 1351 (DNJ 1992).

804 867 F. Supp. 1173, 1182-83 (1994).

805 Gerstner 1993 *Santa Clara law review* 247.

806 Gerstner 1993 *Santa Clara law review* 247-248.

For negligence, the producers' duty is not an expectation of precision, but merely one of reasonableness. Thus, the AI system need not be fault-free. It should merely satisfy the standard of care relating to a reasonable producer of similar systems under the conditions. Furthermore, if the end-user such as the physician, unduly relies on the outcomes generated by an AI system and fails to apply reasonable care in examining the solutions, the physician may be negligent and accountable for any harm caused due to errors.⁸⁰⁷

Challenges in the ability to sue computer engineers for malpractice, has influenced an argument in favour of developing a generally recognised professional licensing system for computer and software engineers, where ownership of a professional license in the field confirms that an engineer has at least complied with the minimum level of standards, experience and knowledge needed to qualify as a licensed engineer.⁸⁰⁸ In a 2008 survey conducted by the IEEE Computer Society, more than more than 60 per cent of its members supported the development of a professional engineering licensing exam for software engineers.

Shortly thereafter, the Software Engineering Licensure Consortium took steps to implement the licensing exam,⁸⁰⁹ and as a result, organisations such as the IEEE-USA Licensure and Registration Committee's (LRC) mainly focus on cases of licensing in regard to professional engineers, and is dedicated to advancing professional licensure for engineers and continuously evaluating the industry's credentialing.⁸¹⁰

Thus, a key factor for considering a standard of care, is to establish if a custom or practice exists and is applied in the technological and computer sector in the context of safety standards for operating systems and associated software. In this regard, industry acceptable practices such as the ISO series should be studied. As the very

807 Gerstner 1993 *Santa Clara law review* 248-249.

808 Software Engineer Insider "The Movement to License Software Engineers" <https://www.softwareengineerinsider.com/articles/software-engineer-licensure.html> (Date of use: 12 October 2020).

809 Software Engineer Insider <https://www.softwareengineerinsider.com/articles/software-engineer-licensure.html> (Date of use: 12 October 2020).

810 IEEE-USA <https://ieeusa.org/volunteers/committees/lrc/> (Date of use: 12 October 2020).

least, employing best practices relating to AI design, application and evaluation may assist with determining the standard of care.

However, depending on the factual circumstances of a case, a court could still direct that a defendant should be measured against a greater standard in relation to that determined by the relevant industry – should it hold the view that the standard imposed by industry falls short. In the case of *TJ Hooper v N Barge Corp.*,⁸¹¹ a proprietor of a tugboat was held accountable due to losing flatboats it was towing, because of its failure to install a radio that would report on weather conditions. The owner submitted that the application of a radio was not customary practice in the nautical industry, and therefore, the absence thereof was not contrary to standards in the industry. Hand J rejected the argument and stated that,

[c]ourts must in the end say what is required; there are precautions so imperative that even their universal disregard will not excuse their omission.⁸¹²

The precise standard of care should thus be decided based on each case.

4.4.2.4. Causation

The causation element must also be considered. For negligence, the defendant's liability is limited to only those injurious defects that he could have discovered and rectified through "reasonable" practice, the so-called causation element. Causation is based on a two-pronged test. First, the claimant must establish on a preponderance of probabilities that there is a reasonable nexus between the offender's act or omission, and the ensuing harm endured by the claimant.⁸¹³ Secondly, the negligence should be the proximate or legal cause of the harm, as established in terms of the foreseeability test, where the injury to the claimant should be reasonably foreseeable.⁸¹⁴

811 60 F.2d 737 (2d Cir. 1932).

812 [739-740].

813 Scott 2008 *Maryland law review* 448.

814 *Salomey v Jeppesen & Co* 707 F.2d 671 (2d Cir. 1983). The court held that, a navigational chart maker's dependence on incorrect data used for its navigational charts was the legal cause of an aircraft crashing. In *Ultramares Corporation v Touche* 174 NE 441 1932 the court held that, "the law should not admit to a liability in an indeterminate amount for an indeterminate time to an indeterminate class" as it would open the "floodgates" of claims.

Foreseeability serves as a restriction both for causation, as well as the degree and type of damages that is recoverable for negligence.⁸¹⁵ Because of the various potentially negligent participants that may be involved, and the sophistication of AI programs which limit the plaintiff's visibility into AI's "black box" software and development process, proving that the program or that any of the parties were responsible for any defects in AI systems, may be challenging for the plaintiff.

Where there are various defendants involved, if a plaintiff cannot establish with certainty which of them are responsible for the inflicted harm, the courts may apply the principle of "alternative liability", where the onus of proving causation for injury or loss is shifted to various defendants, even though only one of them could have been responsible. This shifting of the burden is provided for under the *res ipsa loquitur* maxim where it moves to the defendant once a defect in the AI system has been demonstrated.⁸¹⁶

4.4.2.5. Breach causing damage

A breach of a duty ought to have caused the claimant's injury.⁸¹⁷ All damages, for instance physical or property damage proximately caused as a of the defendant's negligence (in some cases pure economic losses) are recoverable. In cases of gross negligence, punitive damages can also be awarded against healthcare practitioners and providers based on medical malpractice, and also against manufacturers for faulty goods.⁸¹⁸

The damage recoverable is limited by the foreseeability requirement, and in terms of this, it is not expected of a defendant to anticipate every possible situation under which a person could be harmed. In *Perrine v Pac Gas & Elec Co*,⁸¹⁹ the court held that,

815 Scott 2008 *Maryland law review* 448.

816 Keeton *et al Law of torts* 143; *Ybarra v Spangard* 25 Cal.2d 486, 154 P.2d 687. The *res ipsa loquitur* doctrine is also prevalent in English tort and South African delict.

817 Fleming *The law of torts* 186.

818 *Milwaukee & St. Paul Ry. v Arms* 91 U.S. 489, 492–93 (1875).

819 186 C.A.2d 442, 449 (1960).

[e]ven one who maintains so dangerous an instrumentality as a high-power line need not anticipate at his peril every possible fortuitous circumstance under which someone may make contact with the wires causing injury [...].

It would be unfair to expect a producer of an AI system to heed or safeguard against every calamity which may arise from the use of the system. The limitation placed by the foreseeability requirement, is relevant in the computer field where a software system is designed as non-customised. For instance, with a mass-marketed operating system,⁸²⁰ it is foreseeable that same will be applied without any modifications to run a computer system, and any software flaws could generate unauthorised interferences into and cause destruction or loss to the computer system or its data.⁸²¹

4.4.2.6. Onus of proof

The burden of proving negligence, is that of the plaintiff who must do so on a balance of probabilities.⁸²² Where there is only circumstantial evidence, the *res ipsa loquitur* principle will enable a court to make the presumption of negligence based on the facts, and pass the duty to the defendant to demonstrate the absence of negligence.⁸²³ The doctrine leads to liability without fault. Before the statutory codification of products liability in consumer protection legislation, the doctrine was widely relied upon by the courts who shifted the onus of establishing negligence to the defendant.⁸²⁴ The *res ipsa* doctrine created an onerous sense of duty, because the courts, who had to determine negligence, in most cases found in favour of the plaintiff.⁸²⁵

4.4.2.7. Contributory negligence

Contributory negligence was a leading principle in US jurisprudence during the 19th and 20th century, and historically was the rule in all states. Today, the states that still apply pure contributory negligence are Alabama, North Carolina and Virginia. According to the contributory negligence doctrine, when multiple parties are alleged

820 Refer to section 4.3.3.1 on the different software systems.

821 Scott 2008 *Maryland law review* 450.

822 Keeton *et al Law of torts* 239.

823 *Hyder v Weilbaecher* 1981 54 NC App 287 283 SE 2d 426, Keeton *et al Law of torts* 258.

824 Keeton *et al Law of torts* 681.

825 Keeton *et al Law of torts* 257.

to be at fault, the jury will reduce the quantum of damages recoverable by the party who has suffered loss or injury. The plaintiff can only recover from any given defendant in extent of the defendant's degree of liability. In other US states (such as Arizona, New Mexico and California), under principles of pure comparative negligence, damages are simply reduced by the degree to which the plaintiff contributed to the harm. Some states (such as Georgia, Colorado and Kansas), adopt a mixed model of "comparative and contributory negligence" whereby a plaintiff whose degree of fault is below fifty percent, may claim damages that is reduced relative to the extent of the plaintiff's fault, but a plaintiff who is above fifty percent at fault, is not entitled to recover damages.⁸²⁶

4.4.2.8. Strict product liability

The US has fundamentally contributed to principle on tort with regards to the advancement of strict product liability in relation to defective products.⁸²⁷ It spurred the implementation of the EU Directive on Product Liability,⁸²⁸ particularly after the judgment of *Greenman v Yuba Power Products Inc*,⁸²⁹ where Supreme Court of California publicly pronounced and accepted strict product liability in tort. *Greenman* influenced a vital movement towards a principle of enterprise-based liability, rather than predicating liability based on the negligence of the defendant.⁸³⁰ The notion of strict liability was later codified under the Restatement (Second) of Torts.⁸³¹

-
- 826 Larson A "Negligence and tort law" https://www.expertlaw.com/library/personal_injury/negligence.html (Date of use: 11 August 2020). *Li v Yellow Cab Co* 13 Cal.3d 804, 532 P.2d 1226 1975, accepted comparative negligence over that of contributory negligence; *I-lawsuit* "Comparative and contributory negligence by state" <https://www.i-lawsuit.com/comparative-and-contributory-negligence-by-state/> (Date of use: 12 October 2020).
- 827 Reimann M "Product liability" in Bussani M and Sebok, AJ (eds) *Comparative tort law: Global perspectives* (Cheltenham: Edward Elgar Publishing 2015) 250-278.
- 828 85/374/EEC.
- 829 377 P.2d 897 (Cal. 1962).
- 830 Under the strict liability concept, the defendant's liability is simply based on whether injuries to person was part of a business enterprise based on public considerations.
- 831 Restatement (Second) of Torts section 402A (1966) "special liability of seller of product for physical harm to user or consumer one who sells any product in a defective condition unreasonably dangerous to the user or consumer or to his property is subject to liability for physical harm thereby caused to the ultimate user or consumer, or to his property, if (a) the seller is engaged in the business of selling such a product, and (b) it is expected to and does reach the user or consumer without substantial change in the condition in which it is sold. (2) The rule stated in subsection (1) applies although (a) the user has exercised all

Section 402A, allows for recovery only in the event that the subject matter inflicting the harm: (1) is a product; (2) is defective; (3) the product is employed in commerce by a supplier; and (4) the defect caused injury and damages. For strict product liability under tort, the focus is on the product that is defective, and not on the breach of a duty of care in the design and manufacture of a product.⁸³² In *Winter v GP Putnam's Sons*,⁸³³ it was held by the court that,

[for strict liability] it is not a question of fault but simply a determination of how society wishes to assess certain costs that arise from the creation and distribution of products in a complex technological society in which the consumer thereof is unable to protect himself against certain product defects.⁸³⁴

As strict product liability is only concentrated around the products, the question is whether medical AI programs are regarded as products or services. The Restatement (Third) of Torts Product Liability indicates that,

a product is tangible personal property distributed commercially for use or consumption. Other items, such as real property and electricity, are products when the context of their distribution and use is sufficiently analogous to the distribution and use of tangible personal property that it is appropriate to apply the rules stated in this Restatement.⁸³⁵

It is clear from the Third Restatement that the description of products is not meant to be static or narrow, and in each instance a court must decide as a question of law whether an item is to be regarded as a product or not. A number of judgments have held that computer software and data are tangible products, and hence would be accepted as a product when it concerns strict liability.⁸³⁶

832 Keeton WL *et al Products liability and safety* 2nd ed (Foundation Press Inc. Westbury, N. Y. 1989) 189.

833 938 F.2d 1033, 1035 (9th Cir. 1991).

834 [1035].

835 Restatement (Third) of Torts: Products Liability (1997) section 19(a).

836 *MW Mfrs Inc v Friedman Corp* No. 97-C-8319, 1998 WL 417501 (held that, software was regarded as tangible in nature because “the end result that Plaintiff sought was a product (software package) with certain identifiable capabilities”); *WalMart Stores Inc v City of Mobile* 696 So. 2d 290, 291 (held that, software is considered as tangible and any sale thereof would be liable to gross receipts for taxation purposes); *MAI Basic Four Inc v Generic Bus Solutions Inc* CIV. A. No. 9908, 1990 WL 3665 (“It is my view that documents or other physical objects containing confidential information, as well as computer disks or tapes containing software are tangible and thus able to be replevied.”); *S Cent Bell Tel Co v Barthelmy* 643 So. 2d 1240, 1245 (“As computer software became more prevalent in society, and as courts’ knowledge and understanding of computer software grew, later cases saw a shift in courts’ attitudes towards the taxability of computer software, and

With AI systems, the issue is whether computer software would be regarded as “tangible personal property”. Arguably, AI systems render a service due to their intangible nature. Distinct from computer hardware, computer software has no physical composition and is not tangible. Furthermore, as with doctors who are skilled service providers, developers of medical AI systems are also skilled service providers because they are experienced professionals capable of transferring their knowledge as a technology or computer expert, into system codes that are not tangible.⁸³⁷

As a result, medical AI systems can be classified as services. Conversely, the packaging and marketing of software would not inherently be different in comparison to other objects of trade, which are regarded as products. In deciding whether a transaction has to do with goods or services, the courts would have regard to some of the following considerations: (1) if the transaction has to do with a “sale”; (2) if the offender was involved in the selling or promotion of the article; and (3) if the article was produced in large quantities.⁸³⁸

It has also been suggested by experts, that AI systems are complex with regard to design and function, and they can be categorised by analysing their purpose and end product.⁸³⁹ Thus, if an AI system is employed to perform a service that is ordinarily performed by a human, e.g., where it is used as a tool by doctors in treating patients, the system could be regarded as a service. On the other hand, if the system merely offers repetitive data evaluations such as laboratory analysis, the system is likely a product. Mass-marketed programs (turnkey programs) will most likely be branded as products to which the strict liability principles would apply, whilst custom produced or modified software packages with distinctive characteristics, are expected to be branded as services to which the negligence principles will apply.⁸⁴⁰

courts began holding computer software to be tangible for sales, use and property tax purposes”).

837 Nguyen 1994 *Santa Clara law review* 1195.

838 Keeton et al *Products Liability* 760.

839 Turley 1988 *Computer law journal* 457.

840 See section 4.3.3.1 where the three different categories for software products were discussed.

The second aspect for strict product liability to apply, is to establish if the relevant product is defective and unreasonably hazardous. A defect can emanate from several areas of an AI system, i.e., the knowledge base, inference system and incorrect interface language. The rationality of the product design and safety, is to be assessed with regard to the moment of design, as opposed to the moment at which the damage happens.⁸⁴¹ Products are considered to be defective when they are “unreasonably dangerous” to the ordinary end-user.⁸⁴² In *Rudisaille v Hawk Aviation Inc*,⁸⁴³ it was held that,

[t]o prove liability under section 402A the plaintiff need only show that the product was dangerous beyond the expectations of the ordinary consumer. The reasonableness of the acts or omissions of the plaintiff is never considered in determining whether a product is ‘defective’.

The Restatement (Third) of Torts: Product Liability redeveloped strict product liability by redefining what a product defect means, namely:

A product:

contains a manufacturing defect when the product departs from its intended design even though all possible care was exercised in the preparation and marketing of the product (i.e., manufacturing defect);
is defective in design when the foreseeable risks of harm posed by the product could have been reduced or avoided by the adoption of a reasonable alternative design by the seller or other distributor, or a predecessor in the commercial chain of distribution, and the omission of the alternative design renders the product not reasonably safe (i.e., design defect); and
is defective because of inadequate instructions or warnings when the foreseeable risks of harm posed by the product could have been reduced or avoided by the provision of reasonable instructions or warnings by the seller or other distributor, or a predecessor in the commercial chain of distribution, and the omission of the instructions or warnings renders the product not reasonably safe.⁸⁴⁴

The due care standard as a negligence analysis in the manufacturing defect of the product, is clearly omitted from the definition as it encompasses a strict product liability assessment. It entails establishing the “intended design” concerning the product. In contrast, a design defect emanates due to the omission to implement a “reasonable alternative design” which intended to achieve the reasonable safety of

841 *Ward v Hobart Manufacturing Co* 1971 450 F 2d 1176 (5th Cir).

842 Restatement (Second) Of Torts Section 402A 1966.

843 92 NM 575, 592 P.2d 175 (1979).

844 Restatement (Third) of Torts: Products Liability (1997) section 2.

the product, denoting a departure from the strict product liability principles to the traditional standard for negligence.

It is, therefore, necessary to first establish if the defect emanates from either a design or manufacturing fault or error. Generally, the creation of software for programs would involve a few stages before reaching the end-user, which is categorised to be: “(i) the design phase, (ii) the coding phase, (iii) the testing phase, and (iv) the replication and distribution phase”.⁸⁴⁵ It is clear that any defect presented during the “design phase” is to be considered as a “design defect”, and similarly, a defect presented at the reproduction and supply phase, is to be considered as a “manufacturing defect”.⁸⁴⁶

The matter is however not as clear when determining whether the coding phase is to be considered as part of the design, or the manufacturing stage. Thus, it can be argued that, the process prior to the “replication and distribution phase” is design, hence – the design defect standard of negligence ought to be applicable to defects originating during the “design phase”. Everything occurring after the design phase should be accepted to be part of the manufacturing process, and be subject to the manufacturing defect of strict product liability. Software design is usually undertaken before the coding, and hence it would be part of the replication and distribution phase and a strict liability standard (manufacturing defect).⁸⁴⁷

845 Scott 2008 *Maryland law review* 459.

846 Scott 2008 *Maryland law review* 459.

847 Scott 2008 *Maryland law review* 460.

Table 1:⁸⁴⁸

Applicable Test	Development Phases
Design Defect / Negligence Standard	Design Phase
Manufacturing Defect /Strict Liability Standard	Coding Phase
	Testing Phase
	Replication/Distribution Phase

Strict liability also passes to the seller, distributor or persons deemed to be part of the distribution chain for failing to provide sufficient directions or warnings. The standard of care is if a reasonable person that is in the situation of the seller or distributor, would have warned the consumer. In *Borel v Fibreboard Paper Products Corporation*,⁸⁴⁹ the court held that,

the decision to market a product requires a balancing of the product's utility against its known or foreseeable danger [...]. Even when such balancing leads to the conclusion that marketing is justified, the seller still has a responsibility to inform the user or consumer of the risk of harm. The failure to give adequate warnings in these circumstances renders the product unreasonably dangerous.

The motivation behind this principle, is because consumers of products are at liberty to decide on whether the product's usefulness or advantages warrant subjecting him or her to the possibility of harm. Therefore, an obligation to warn emanates, when a reasonable person would expect to be apprised of the probability of harm to enable him to choose whether he wants to be exposed to it.

Should there be no reasonable alternative safe design, there are two options available to the seller: (1) to not market the product, or (2) to ensure that it is fairly safe by way of warnings and directions for the products use. However, the Third Restatement prescribes a different liability requirement in respect of sellers of

848 The table is a summary denoting product defects under the Restatement (Third) of Torts: Products Liability (1997) section 2.

849 493 F.2d 1076, 1088-89 (5th Cir.1973) [760].

medications or medical devices. Thus, in terms of Section 6(c) a rule is adopted that exempts medical device sellers from complying with the alternative safe design requirement relevant to other products. This exemption states:

A prescription drug or medical device is not reasonably safe due to defective design if the foreseeable risks of harm posed by the drug or medical device are sufficiently great in relation to its foreseeable therapeutic benefits that reasonable health-care providers, knowing of such foreseeable risks and therapeutic benefits, would not prescribe the drug or medical device for any class of patients.⁸⁵⁰

In terms of the rule, producers of medical devices will escape liability, even in cases where their products were reasonably capable of being made safer. If it is proven that the product has more benefits than risks in respect of at least one category of users, that will motivate a reasonable doctor to prescribe it, it will not be regarded as defective. Section 6(c) entails a standard of care below that of reasonable care for medical devices. Thus, the question is whether an AI medical system is deemed to be a medical device.

The Federal Food, Drug and Cosmetic Act⁸⁵¹ defines a “medical device” as:

[...] an instrument, apparatus, implement, machine, contrivance, implant, in vitro reagent, or other similar or related articles, including a component part, or accessory which is: recognised in the official National Formulary, or the US Pharmacopoeia, or any supplement to them, intended for use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment, or prevention of disease, in man or other animals, or intended to affect the structure or any function of the body of man or other animals, and which does not achieve any of its primary intended purposes through chemical action within or on the body of man or other animals and which is not dependent upon being metabolized for the achievement of any of its primary intended purposes.⁸⁵²

“Software as a Medical Device” (SAMd) is defined under the International Medical Device Regulators Forum (IMDRF)⁸⁵³ to be:

Software intended to be used for one or more medical purposes that perform these purposes without being part of a hardware medical device.⁸⁵⁴

850 Restatement (Third) of Torts: Products Liability (1997) section 6(c).

851 US Code Title 21: Food and Drugs.

852 Section 201(h).

853 The IMDRF are made up of regulators of medical devices globally that creates documents which include topics affecting medical devices.

854 FDA “Software as a medical device” <https://www.fda.gov/medical-devices/digital-health-center-excellence/software-medical-device-samd> (Date of use: 12 August 2020).

AI based SAMD varies from other medical devices, because, (1) of their capacity to constantly learn and adapt; (2) they are becoming pervasive in medical interventions, and in contrast to robotic systems, they deliver recommendations to the extent where skilled and trained healthcare practitioners are not involved; and (3) the method in which they arrive at their recommendations can be enigmatic to healthcare practitioners.⁸⁵⁵ Under US law, certain AI or ML based software functions in term of section 520(o)(1)(E) of the Food, Drug and Cosmetic Act, do not fall under the medical device under section 201(h) of the Act, and do not fall within the device definition. These software functions, known as “clinical decision support software”, are designed to deliver decision assistance for the diagnosis, therapy, deterrence, treatment, or prevention of various illnesses.⁸⁵⁶

The FDA established the governing pathways for new medical devices and traditionally, conduct evaluations on medical devices through specified routes, such as the De Novo classification,⁸⁵⁷ or section 510(k)⁸⁵⁸ premarket clearance in the Food, Drug and Cosmetic Act. In terms of the section 510(k) clearance, in order to market in the US, a Class I,⁸⁵⁹ II,⁸⁶⁰ and III⁸⁶¹ device for the purposes of human use, must observe the requirements of the Section 510(k) clearance, unless the device is exempt from the Section 510(k) requirements.

855 FDA “Artificial Intelligence and ML in software as a medical device” <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device#transforming> (Date of use: 11 August 2020).

856 520(o)(1)(E) of the Food, Drug & Cosmetic Act.

857 A “de novo” pathway is for devices that are automatically categorised as Class III because there is no existing device in the market that could be used to establish a section 510k submission.

858 A 510(k) clearance are for less invasive and lower risk medical devices in respect of that of a “predicate device” which is a “substantial equivalent” of that currently being offered in the public domain. Available at Cohen Healthcare Law “FDA policy on when software that uses artificial intelligence and ML qualifies as a medical device” <https://cohenhealthcarelaw.com/2020/05/fda-policy-on-when-software-that-uses-artificial-intelligence-and-machine-learning-qualifies-as-a-medical-device/> (Date of use: 12 August 2020).

859 Class I devices are not subjected to ‘premarket approval or clearance’ but are required to adopt ordinary controls. An example of this is dental floss.

860 Class II devices prescribe premarket clearance in applying the 510(k) process. Hearing aids would be an example of this.

861 Class III devices require premarket approval as with a new drug application and are generally devices permanently embedded into a human body or is required to support body functions. A replacement heart device falls under both.

This is not approval, but merely clearance for the device to be sold in the US. A premarket submission is required to the FDA to prove the reliability and effectiveness of a device, similar to an already approved device in the market. The role of the FDA also includes reviewing and authorising adaptations to medical devices, including SaMD's, which will be contingent on the magnitude of the risk presented to patients. A De Novo clearance for Class III novel medical devices, allows the FDA to examine and scrutinise new technologies to ensure their safety. A few AI medical systems have already been approved under the De Novo and Section 510(k) pathways, which include:

- (1) IDx's IDx-DR is an AI software system for the autonomous detection of diabetic retinopathy in adults who have diabetes;
- (2) Imagen's OsteoDetect, uses an AI algorithm to scan X-ray images for a common type of wrist bone fracture, known as distal radius fracture;
- (3) The cloud based DreaMed Advisor Pro is a diabetes treatment decision support product that analyses data from continuous glucose monitors, insulin pumps and self-monitoring to determine an insulin delivery recommendation;
- (4) The Coronary Calcium Scoring algorithm from Zebra Medical Vision, offers a coronary artery calcification score from a patient's ECG-gated CT scan.⁸⁶²

The FDA concedes that its conventional structures for the regulation of medical devices, were not developed with AI health technologies in mind.⁸⁶³ Under the FDA's existing software modification framework, it foresees that most AI driven software adjustments made to a device, are likely to require premarket review.⁸⁶⁴ On 2 April 2019, the FDA released a document that "proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD) – Discussion paper and request for feedback", which addresses the FDA's proposed framework to approach "premarket review for AI and ML-driven software modifications".⁸⁶⁵ The framework is intended to allow the FDA to assess and oversee a software device from the pre-market to post-market stages,

862 Cohen Healthcare Law <https://cohenhealthcarelaw.com/2020/05/fda-policy-on-when-software-that-uses-artificial-intelligence-and-machine-learning-qualifies-as-a-medical-device/> (Date of use: 12 August 2020).

863 US FDA <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device#transforming> (Date of use: 11 August 2020).

864 US FDA <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device#transforming> (Date of use: 11 August 2020).

865 <https://www.fda.gov/media/122535/download> (Date of use: 11 August 2020).

and in so doing, to promote AI-driven SAMD's, and also promising the safety of patients. In line with its proposed agenda, the FDA recommends a "predetermined change control plan" in pre-market proposals. This proposal envisages modifications known as "software as a medical device pre-specification", and the relevant strategy to be executed to give effect to those modifications in a structured way, that ensures the safety of patients – known as the "algorithm change protocol".⁸⁶⁶

As seen from the earlier examples, the FDA at present endorses AI and ML SAMD's as medical devices in terms of the De Novo classification and section 510(K) submissions. All AI and ML SAMD reviewed and endorsed by the FDA to date, were regarded as "locked" algorithms prior to marketing, and described to be:

An algorithm that provides the same result each time the same input is applied to it and does not change with use.⁸⁶⁷

A major problem lies with approving "unlocked" AI and ML SAMD's, which change or adapt following delivery thereof to consumers or healthcare providers and have "learned from real-world experience."⁸⁶⁸ They are subject to more variance in the application and real-life situations due to human intervention, the intricacies of the systems, and how they adapt to their ecosystem. Thus, they present greater potential for harm and uncertainty in contrast to the benefits they may offer.⁸⁶⁹

The FDA is addressing a few of these concerns by imposing training and accreditation programs, but more nuanced and complicated systemic problems still need to be examined. The FDA's new regulatory framework for AI medical devices, will require approval as it still subject to public comment, and may require Congressional approval. Until the new regulatory framework comes into effect, AI medical devices are only authorised for release into the market once they are cleared or approved through the traditional pathways, and subject to this, they will be considered as a medical device as per the Food, Drug and Cosmetic Act and the FDA. Under these circumstances, the section 6(c) exception in terms of design

866 <https://www.fda.gov/media/122535/download> (Date of use: 11 August 2020).

867 <https://www.fda.gov/media/122535/download> (Date of use: 11 August 2020).

868 <https://www.fda.gov/media/122535/download> (Date of use: 11 August 2020).

869 Gerke *et al* 2020 NPJ digital medicine.

defect under the Third Restatement that exempts sellers of medical devices from complying with the alternative safe design requirement, will apply to medical AI systems.

In terms of strict liability provisions, sellers are defined as individuals that are involved in the enterprise of selling the product.⁸⁷⁰ The definition excludes those that only intermittently sell the product, or those who do not derive a livelihood from the sale of the product.⁸⁷¹ Those involved in the production, manufacturing, promotion and supply chain would be regarded as sellers. Therefore, a manufacturer of a defective component is strictly accountable as with a producer, unless the defect has to do with the manner of its use in an already constructed product, in which event the manufacturer would be liable.⁸⁷²

Any person that is injured by a product may rely on strict liability, including consumers, end-users and onlookers.⁸⁷³ Under strict liability, the claimant must prove that a defect caused the harm on a balance of probabilities of the evidence, i.e., such a defect should be the proximate or legal cause of the harm.⁸⁷⁴ This will not always be easy to establish, given the complex design and development process.

The requirement of causation is essential, as without its application, producers could be held accountable for any issue that ensues in relation to the systems that they design. The intention is that sellers ought to be held accountable for their defective products that ensues in harm. Only damage to person or property is remunerative under section 402(A),⁸⁷⁵ and pure economic losses are excluded.⁸⁷⁶ If the defective product itself is damaged, this would also be compensable.⁸⁷⁷

In conclusion, actions for liability in American law, emanating from harm caused by defective AI systems, can be pursued in terms of the negligence principles under

870 Restatement (Second) of Torts section 402A (1966).

871 Keeton *et al Law of torts* 705.

872 Keeton *et al Law of torts* 705-706.

873 Keeton *et al Law of torts* 704.

874 Keeton *et al Law of torts* 712.

875 Restatement (Second) of Torts section 402A (1966).

876 Keeton *et al Law of torts* 708.

877 Keeton *et al Law of torts* 708.

tort law and strict product liability regime provided for under the Second and Third Restatement. In the case of the physician who is considered as a professional end-user of medical AI systems, a plaintiff can rely on the traditional negligence-based tort, which is akin to a medical malpractice claim. While producers can be held liable for negligence due to defective AI systems, plaintiffs will have a challenge when pursuing a traditional tort claim against the producers, since a “computer malpractice” claim is not recognised in American law at the current development phase of the AI industry.⁸⁷⁸

In an era in which prospects for progressive legislation are stark, and in which innovation is prioritised over notions of individual responsibility, there is reluctance to hold AI producers to an elevated responsibility of care. Producers are therefore subjected to the normal standard of care of persons within their industry, although a court may decide to hold them to greater standard if it is of the opinion that the industry standard is ineffective. The same considerations would apply to strict product liability. A plaintiff also has the option to institute action based on strict product liability where an AI system is used, which is considered to be a defective or hazardous product in accordance with the standards prescribed under US law for strict liability negligence.

4.4.3. Conclusion on AI liability based on comparative jurisdictions

An AI system can be regarded as a stand-alone system, or a system linked to a product and is thus regarded as a component of the assembled and marketed product. Civil liability issues can arise from defective AI systems that are deployed in healthcare settings that produce a material output in the form of advice or assistance, offering solutions or the performance of tasks. In such situations the liable parties must be identified, the grounds for their liability must be established, and whether the liability standards in South Africa’s civil law are designed to address these issues must be established. This research does not consider any liability arising from fraudulent behaviour or misrepresentation based on intent, but deals with the issue of negligence.

878 Refer to section 4.4.2.3

In South Africa, where harm is caused due to defective AI systems, the party who suffered harm would have a claim against the producers or professional end-users of the systems – either in delict, based on negligence (where the plaintiff carries the duty of establishing that the standard of care of a reasonable expert or professional) has been breached, or based on product liability (as codified under the CPA).

Similarly, in the UK and US, the producers and professional users of software can be held accountable in terms of strict (no-fault) product liability standards,⁸⁷⁹ or the (fault-based) tort regime based on negligence where the plaintiff has the onus of proving all the requirements, that is a duty of care owed by the defendant, a breach of such duty and ensuing injury or losses endured by the plaintiff due to the use of defective AI systems. Thus, producers and healthcare practitioners as professional end-users, could be exposed to liability for injuries or losses caused by defective AI systems, that arises from delictual negligence or the strict product liability regime.

In the UK and US, if an AI system is categorised as a product, the strict product liability principles will be applicable, and if classified as a service, the negligent tort liability principles would apply. Thus, the relevance of strict product liability standards to AI, is firstly based on whether AI systems are deemed to be products. This issue has incited fierce discussion among computer and technology law analysts concerning the actual landscape of AI systems.

The controversial issue, is whether AI systems must be described as products or as services - as this prescribes whether actions could be premised either on strict liability or on negligence standards. Aspects that contribute to the categorisation of AI systems as products or services are their tangibility nature, their functional characteristics, their informational nature, and their specific application of the software. The English and the US approach centres on the policies, where producers are held liable for product defects without fault or negligence on their part. The above approach as is also applied by US courts – premised on the fact that the manufacturer is in a position to reduce the possibility of defects in a product, is most knowledgeable of the likely presence of a products defect once placed into the

879 In the UK, strict liability is codified under the CPA and in the US under the Statement contained in the Second and Third Restatement.

market, and can safeguard against liability through insurance. Safety and quality assurance standards can significantly assist in establishing if a product contains a defect. Compliance concerning safety and quality standards highlights the safety controls that are offered to the consumer, and may suggest the absence of negligence.

Unlike the US or UK, delictual law in South Africa does not make the distinction as to whether AI systems are classified as a product or a service, since both will be actionable in terms of the *Aquilian* action or non-patrimonial loss when damage is unlawfully caused. Similarly, this distinction does not apply under the CPA.⁸⁸⁰

In supplementing the common law remedy for delict, in terms of the CPA every consumer has a right to receive goods that are reasonably suitable for the purposes for which they are generally intended, are of good quality; in good working order and free of any defects.⁸⁸¹

In terms of services: “the consumer has a right to the performance of services in a manner and of a quality that persons are generally entitled to expect”.⁸⁸² Thus, South African delictual law allows a claimant greater flexibility in terms of pursuing an action, either under the negligence principles, or in terms of the strict liability principles (as opposed to the jurisdictions analysed).

Delictual principles in South Africa require proof of the producer’s or professional end-user’s negligence in the case of defective AI systems, which in most cases is unachievable due to the highly technological and intricate processes associated with such systems. The doctrine of *res ipsa loquitur* may, however, offer assistance to the plaintiff in South Africa. This doctrine is also recognised in Anglo-American doctrine tort law where a court can presume negligence premised on the facts of at hand, whether there no direct evidence as to the cause of the incident. Neethling *et al*⁸⁸³ state that, this difficulty should be addressed in a much similar way as the Anglo-American legal system, by way of the acceptance of the doctrine.

880 68 of 2008.

881 Section 55(2)(a) & (b).

882 Section 54(1)(b).

883 Neethling, Potgieter and Visser *Law of delict* 348-349.

The authors suggest that, inference of negligence should be considered where a plaintiff proves that he suffered prejudice, due to a product being defective at the time that the manufacturer relinquished possession over the product. However, the burden of the plaintiff in South Africa may also be alleviated by seeking to impose liability on the producers or professional end-users for defective AI systems, based on strict product liability which alleviates a plaintiff's burden to demonstrate the defendant's fault.

What is a huge issue in South Africa is that the current liability frameworks under delict seems to hold the human actors responsible. However, as the technology becomes more sophisticated and can be classified as strong AI. the autonomy of the technology will become more prevalent. In this regard, the traditional delictual principles will need to consider according legal personhood to AI that confers upon it, rights and obligations.

The idea of "computer malpractice" liability in relation to producers of AI systems that have defects, has been considered by the other countries investigated in this research. An action can only be pursued against the producer whose conduct causes loss or injury in the performance of his or her professional duties. A professional's standard of care is considered to be greater than that of an average person. Of the two legal jurisdictions examined in this research, only the UK acknowledges software engineering to be a profession. In terms hereof, software developers may be accountable under negligence for not complying to the professional standard of care.

The UK's Engineering Council has issued guidelines on duties and legal liabilities of software engineers, which incorporates a code of good practice that applies to engineers and supervisors engaged with security-linked technologies. Similarly, a professional status for the software and computer sector in South Africa should be considered, as this may encourage the implementation and observance of increased standards for the creation and advancement of AI-based software programs and those that would primarily be "fit-for-use". This would provide a benchmark that registered software engineers and technicians have met globally recognised professional standards.

AI systems that used either a stand-alone system, or a system which forms part of a commercial product as medical devices, could also be regulated under the Medicines Act of South Africa. “Medical device” as defined under the Act, is broad enough to consider AI systems that are used in healthcare environments as medical devices. The difficulty arises in the endorsement of AI for licensing as a medical device under the current regulatory framework. The traditional regulatory paradigm of medical device licensing before premarket release, was not designed for AI systems that are continuously absorbing, changing and adapting in real-time after they reach consumers.

South Africa has not tackled the challenges that arise as a result of AI’s as medical devices. New Medical Device Regulations in the UK, introduces rules for medical software intended to bring noteworthy reforms to the regulation of medical devices. However, these regulations may need to shift more towards a centred regulatory assessment framework for AI based medical devices, and possibly novel and futuristic ways in which they can be introduced into medical products.

Of the comparative jurisdictions, America’s FDA has made more progress concerning the regulation of AI systems that are considered to be medical devices, and has cleared or approved stand-alone AI systems as medical devices for market consumption. Furthermore, the FDA issued a discussion document in 2019 that offers the FDA’s approach to pre-market review for AI software adaptations.⁸⁸⁴ Likewise, South Africa needs to consider a new framework that describes an innovative approach, which is critical to the licensing and regulation of adaptive AI medical devices. Until such time as a new regulatory framework has been proposed, AI medical devices should be regulated under the traditional principles of strict product liability, or delictual negligence liability.

4.5. *Ethics guidelines for achieving trustworthy AI in healthcare*

“Ethics” as used in this thesis, refers to a concept of morality which encompasses,

884 <https://www.fda.gov/media/122535/download> (Date of use: 11 August 2020).

a system of rules and values for guiding human conduct, as well as principles for evaluating those rules.⁸⁸⁵

Accordingly, ethical behaviour denotes compliance with specific values recognised as being inherent to human ideals (e.g., safety of people, equality, autonomy and dignity), or being a moral requirement representing views and principles of particular clusters of societies (e.g., cultural practices), or individual demands (e.g., social justice).⁸⁸⁶

This research does not consider AI in a moral sense or from a particular normative perspective, but aims to enhance the dialogue around the development of a suitable regulatory regime that will entrench ethics into AI systems. This is because ethical values can be construed as directives to behave in distinct ways and environments, for a specific population.

As discussed in Chapter two, AI holds significant potential in terms of precise, efficient and cost-effective diagnosis and treatments in healthcare. However, there is also increasing recognition of ethical and societal implications⁸⁸⁷ that may ensue from unsupervised developments in AI, which is a key consideration in daily healthcare practice, given its massive data concentration.

Mireille Hildebrandt⁸⁸⁸ contends that, currently we are operating in between the realms of an “information society” and a “data-driven society”. She underlines how the ubiquitous use of machine-learning systems that propels the “data-driven agency”, endangers ideals such as privacy, individuality, freedom, equality, and due process. Because intelligent technologies such as AI tend to be fundamentally unique in comparison to other types of technologies that we are accustomed to, the difficulty for society is dealing with their divergence and power.⁸⁸⁹

885 Boddington P *Towards a code of ethics for artificial intelligence* (Springer International Publishing Cham 2017) 10; Waltz and Firth-Butterfield 2019 *Duke law & technology review* 184.

886 Waltz and Firth-Butterfield 2019 *Duke law & technology review* 184.

887 The implication were discussed in Chapter two are data privacy violations, data inequity, bias, accountability and safety.

888 Hildebrandt M *Smart technologies and the end(s) of law: novel entanglements of law and technology* (Edward Elgar Publishing Cheltenham UK and Massachusetts USA 2015).

889 Hildebrandt *Smart technologies* 3.

AI applications, particularly for the healthcare sector, must be presented and implemented in ways that foster trust and awareness, and that respects basic human rights.⁸⁹⁰ Trust is a precondition in order for the public to create, implement and use AI systems, and it is not only centred on the innate properties of the technology, but includes characteristics of the “socio-technical” practices relating to AI products.⁸⁹¹

Strategies that enhance the capacity to understand and develop AI technologies and to enjoy their use in an accountable manner, may assist in tackling the issues. Guiding standards are being considered globally in order to encourage trustworthy advancements and use of AI systems in general, and to mitigate against undesirable risks.⁸⁹²

These guiding principles can support and provide certainty and standards to producers of the system and governments when making choices about the development and application of AI. The debate regarding the comparative merits of principles and rules appear to indicate that they are mutually exclusive. Some authors do not support this view, as they believe that principles tend to encompass rules that support their implementation, whilst rules normally comprise of options where varying actions and outcomes still accomplish the goals of the rules.⁸⁹³

This thesis prefers to make use of principles that will direct the conduct of individuals and society, instead of adopting comprehensive or precise set of rules and will thus,

890 Saxenian *et al* Artificial intelligence and life in 2030 49.

891 These systems comprise of various elements such as humans, governments, corporations, infrastructure, software, technology protocols, regulatory frameworks, prevailing laws, incentive schemes, auditing processes, best practices, reporting and others, see “European Commission’s Ethics guidelines for trustworthy AI” <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

892 Other examples of guiding principles can be found in Asilomar “AI principles” <https://futureoflife.org/ai-principles/> (Date of use: 14 August 2020); OECD Principles <http://www.oecd.org/going-digital/ai/principles/> (Date of use: 9 July 2019). G20 AI Principles (which draws from the OECD principles) available at <http://www.g20.utoronto.ca/2019/2019-g20-trade.html> (Date of use: 9 July 2019). Also see Martinovs D “10 organizations leading the way in ethical AI” <https://ocean.sagepub.com/blog/10-organizations-leading-the-way-in-ethical-ai> Date of use: 16 August 2020).

893 Thomadakis https://www.ifac.org/system/files/downloads/30th_anniversary_Thomadakis_Pres_Nov_07.pdf (Date of use: 15 July 2020).

consider the case for regulating AI in healthcare in furtherance of broad principles. There is a practical advantage in the adoption of a principle-based framework instead of a rule-based framework, as the first-mentioned allows regulations to react effectively to the evolving landscape of AI without the need for continuous amendment, and also denotes relationships based on authentic trust rather than suspicion.⁸⁹⁴

It will further explore principles-based regulation in which principles are used to outline regulatory objectives and values, and upon which regulators can develop their own systems aimed at observing such principles.

4.5.1. International context

4.5.1.1. European Union

A good example of ethics principles are observed in the European Commission's "Ethical guidelines for trustworthy artificial intelligence" (EU guidelines), of the High-Level Expert Group on AI.⁸⁹⁵ Launched during April 2019, the report aims to foster and secure the ethical "designing, developing, deploying, implementing or using AI products and services" for the EU market.

The fundamental principle behind the EU guidelines is to introduce a "human-centric" context for AI, which acknowledges human ideologies and basic social rights, and those rights protected under the treaties and charters of the EU.⁸⁹⁶ The EU, in its approach, is thus committed to its increased protection against the social and ethical risks presented by AI, especially those that threaten issues of privacy, data breaches, autonomy and equality.

894 Thomadakis
https://www.ifac.org/system/files/downloads/30th_anniversary_Thomadakis_Pres_Nov_07.pdf (Date of use: 15 July 2020).

895 "European Commission's Ethics guidelines for trustworthy AI" <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

896 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

The EU guidelines are grounded in four principles of ethics entrenched in basic human rights, which is important to be upheld to ensure a trustworthy outlook to the design and advancement of AI systems. The four principles include:

(i) respect for human autonomy; (ii) prevention of harm; (iii) fairness; and (iv) explicability.⁸⁹⁷

In noting the key aspects to achieve trustworthy AI, the EU guidelines also adopt a non-exhaustive “assessment list” for directing the design, deployment or implementation AI systems, which include:

(1) Human agency and oversight, (2) Technical robustness and safety, (3) Privacy and data governance, (4) Transparency, (5) Diversity, non-discrimination and fairness, (6) Societal and environmental wellbeing, and (7) Accountability.

Whilst the EU guidelines are intended to be complementary to other frameworks, these are arguably highly significant to achieve and encourage forceful enactment of the ethical principles in key sectors, such as healthcare, where human autonomy over AI decision driven solutions is imperative. For this reason, in October 2019, the EU guidelines underwent an early scrutiny through European Institute of Technology’s Innovation Community on Health and Ageing (EIT Health), a partnership between the private and public industry of health innovators supported by EU. EIT Health presented outcomes from a survey it conducted at the request of the High-Level Expert Group on AI.⁸⁹⁸

The EIT Health survey conducted amongst industries and EIT Health stakeholders, revealed that merely 22 percent of participants were already aware of the EU guidelines. Over 60 percent of participants were cognisant of the fact that their AI solutions will require regulatory endorsement. Of the seven assessment requirements for achieving trustworthy AI under the EU guidelines, privacy and data governance ranked the highest (which includes data protection and access to data),

897 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020); European Commission “The assessment list for trustworthy artificial intelligence (ALTAI)” <https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment> (Date of use: 25 August 2020)

898 EIT Health pilot of the AI & ethics guidelines of the European Commission’s (EC) high-level group “AI and ethics in the health innovation community” <https://eithealth.eu/wp-content/uploads/2020/01/AI-and-Ethics-in-the-Health-Innovation-Community.pdf> (Date of use: 14 August 2020).

followed by technical robustness and safety (cyber-resilience and duplicability of the AI), and followed by “human agency and oversight”.

Lower ranked were: “diversity; non-discrimination; and fairness”; followed by “accountability”; and lastly “social and environmental well-being”. Moving forward, it was noted that EIT Health should consider issuing specific guidance and/or training on AI and ethics in health innovation, within the frame of the general AI and ethics guidelines, and should focus on addressing specifically sensitive issues for health and ageing, such as human oversight and transparency, accountability, privacy and data governance and technical robustness and safety.⁸⁹⁹

4.5.1.2. United States

The US strategy on AI has advanced largely from initiatives of the private industry and through self-regulation. In the US, the larger industry participants have already created codes of conduct on ethics and AI. Joint initiatives, for instance the Partnership on AI that comprise of partners from industry such as technology giants Microsoft, Amazon, Facebook and Apple, as well as academia and non-profit organisations, have all pledged to develop and share best practices, including on ethics.⁹⁰⁰

In 2018, the Association for Computing Machinery (ACM) also issued a code of ethics and professional conduct for computing professionals.⁹⁰¹ In addition, some companies are adopting their own ethical practices. Examples of this are Microsoft that has formed an internal AI advisory panel to direct AI Ethics,⁹⁰² and Google has a charter on AI principles,⁹⁰³ in order to direct the accountable design and application of AI in research and development and products. OpenAI, is a non-profit

899 <https://eithealth.eu/wp-content/uploads/2020/01/AI-and-Ethics-in-the-Health-Innovation-Community.pdf> (Date of use: 14 August 2020).

900 “Partnership on AI” <https://www.partnershiponai.org/#> (Date of use: 6 June 2020).

901 ACM “Code of ethics and professional conduct” <https://www.acm.org/code-of-ethics> (Date of use: 15 August 2020).

902 “Microsoft FATE: Fairness, accountability, transparency, and ethics in AI” <https://www.microsoft.com/en-us/research/theme/fate/> (Date of use: 15 August 2020).

903 “Google AI principles” <https://www.blog.google/technology/ai/ai-principles/> (Date of use: 15 August 2020).

company with a charter that describes the principles it uses to execute its mission of building AI that is beneficial and safe for all of humanity.⁹⁰⁴

Nevertheless, there is increasing apprehension that self-regulation is not adequate in order to address the ethical concerns that arise from the design and implementation of AI. The AI Now Institute issued an advisory report during 2018, highlighting that the governance systems in most technology corporations are neglecting to safeguard accountability for AI technologies that they design or develop. Therefore, it contends that government bodies must be empowered with much more authority to regulate, examine and monitor AI technologies.⁹⁰⁵ In the context of healthcare, the American Medical Association (AMA) during June 2018, introduced a policy, “augmented intelligence in health care”, presenting a wide agenda to “promote development of thoughtfully designed, high-quality, clinically validated healthcare AI”, in order to be able to ensure that AI is able to deliver the benefits that is anticipated for patients, doctors, and other healthcare stakeholders.⁹⁰⁶ The ethical values recognised in this policy consist of “best practices in user- centred design, transparency, leading standards for reproducibility, addressing bias, and preserving of individuals privacy and personal information”.

4.5.1.3. China

The Chinese strategy on AI is essentially government-led with private-public backing into AI technologies.⁹⁰⁷ Over recent years, interest in China has risen towards developing an AI ethical framework. During 2017, China issued the “next generation

904 <https://openai.com/charter/> (Date of use: 6 June 2020).

905 “AI Now Report 2018” https://ainowinstitute.org/AI_Now_2018_Report.pdf (Date of use: 15 August 2020).

906 American Medical Association “Augmented intelligence in health care H-480.940” <https://policysearch.ama-assn.org/policyfinder/detail/augmented%20intelligence?uri=%2FAMADoc%2FHOD.xml-H-480.940.xml> (Date of use: 16 August 2020)

907 Zhihao Z “AI development plan draws map for innovation” <http://www.chinadaily.com.cn/a/201908/05/WS5d476b48a310cf3e35563d0d.html> (Date of use: 16 August 2020).

artificial intelligence development plan”, setting out future strategic objectives for the development AI in China to be implemented by 2030.⁹⁰⁸

One of those objectives, is to achieve responsible development of AI in China by creating a framework for ethics in AI, to be realised by 2025. It also has a guarantee measure to improve research on ethico-legal concerns, and that promotes the accountable advancement of AI. The country also intends to encourage self-regulation amongst the technology industries in China, and to impose stricter penalties for data infringement, personal privacy infringement, and other unethical actions.

During May 2019, the Artificial Intelligence Industry Alliance (launched in 2017) comprises of Chinese technology companies and colleges, and published its draft guidelines for dealing with AI’s self-regulation.⁹⁰⁹ The guidelines recommend principles for “human-oriented”, “secure/safe and controllable”, and “transparent and explainable” AI, much like the principles found in the EU guidelines.

In addition, in June 2019, the New Generation AI Governance Expert Committee (associated to China’s Ministry of Science and Technology), published a paper containing eight principles for directing AI governance.⁹¹⁰ These principles mainly emulate the EU guidelines for AI.⁹¹¹ For example, AI’s advancements should adhere to “human values, ethics, and morality”, should “be based on the premise of safeguarding societal security and respecting human rights”, and should “respect and protect personal privacy”.

908 The Foundation for Law and International Affairs “Next generation artificial intelligence development plan” <https://flia.org/wp-content/uploads/2017/07/A-New-Generation-of-Artificial-Intelligence-Development-Plan-1.pdf> (Date of use: 16 August 2020).

909 New America “Translation: Chinese AI alliance drafts self-discipline ‘joint pledge’” <https://www.newamerica.org/cybersecurity-initiative/digichina/blog/translation-chinese-ai-alliance-drafts-self-discipline-joint-pledge/> (Date of use: 16 August 2020).

910 New America Chinese expert group offers ‘governance principles’ for ‘responsible AI’” <https://www.newamerica.org/cybersecurity-initiative/digichina/blog/translation-chinese-expert-group-offers-governance-principles-responsible-ai/> (Date of use: 16 August 2020).

911 Refer to section 4.5.1.1.

4.5.1.4. United Kingdom

The UK Committee on Standards in Public Life,⁹¹² issued its report in February 2020 on public standards concerning the application of AI, in order to ensure an acceptable level of conduct as advances in decision-making AI technology are heightened in the public domain.⁹¹³ The principles embody “selflessness, integrity, objectivity, accountability, openness, honesty and leadership”.

The Committee further made recommendations to lawmakers and government, to support a robust and comprehensible regulatory AI framework in the public sphere comprising of, amongst others:

Ethical principles and guidance, articulating a clear legal basis for AI, data bias and anti-discrimination law, transparency and disclosure, upholding responsibility, establishing oversight and diversity.

The above recommendations are embodied in the EU AI ethical guidelines on AI.⁹¹⁴ The Seven Principles (aptly called the “Nolan Principles”), were established by the Committee in its report (chaired by Lord Nolan), and is applicable to public sector officials or anyone employed in other sectors delivering public services. Thus, these principles extend to the UK public healthcare sector, and apply to all public healthcare providers.

Furthermore, the UK House of Lords Select Committee on Artificial Intelligence, established by the government to oversee the financial, societal, and ethical effects of developments relating to AI, issued the report “AI in the UK: ready, willing and able?” on 16 April 2018.⁹¹⁵ Central to the report is the ethical use of AI, embodying five main principles:

912 The Committee serves the UK government, formed in 1994 to advise on ethical standards in public life.

913 “UK Artificial Intelligence and Public Standards Report”
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/868284/Web_Version_AI_and_Public_Standards.PDF (Date of use: 16 August 2020).

914 Refer to section 4.5.1.1.

915 UK House of Lords Select Committee on Artificial Intelligence Report on “AI in the UK: ready, willing and able?”
<https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf> (Date of use: 16 August 2020).

(1) Artificial intelligence should be developed for the common good and benefit of humanity, (2) Artificial intelligence should operate on principles of intelligibility through technical transparency and explainability, (3) Artificial intelligence should not be used to diminish the data rights or privacy of individuals, families or communities, (4) All citizens have the right to be educated to enable them to flourish mentally, emotionally and economically alongside artificial intelligence, and (5) The autonomous power to hurt, destroy or deceive human beings should never be vested in artificial intelligence.⁹¹⁶

According to the report, due to the lack of wider awareness and co-ordination, a public and private cross-sector code of good conduct on AI ethics should be developed, which will provide a basis for statutory regulation. Following the advisory in the report, a draft framework for the sharing of NHS data has been published for consultation; describing the expectations for distribution of patient data under a protected system; the safeguards required due to such distribution; and educating staff on the value of such data and the manner in which it should be used.⁹¹⁷

Focusing on the healthcare sector, the UK released a code of conduct on 5 September 2018 for “data-driven health and care technology”, which comprises of a set of ten principles based on what the government expects from sellers and end-users of data-based technologies.⁹¹⁸ The code accepts that while the data based technologies in healthcare bring benefits to patients and other stakeholders, there remains an obligation on the government to take advantage of the benefits with due regard the matters, such as: trust; accountability; explicability; equality; and discrimination, all of which are complemented by the EU guidelines.⁹¹⁹

916 Corbett L, Felsner F and Cole M “House of Lords Select Committee publishes report on the future of AI in the UK” <https://www.globalpolicywatch.com/2018/05/house-of-lords-select-committee-publishes-report-on-the-future-of-ai-in-the-uk/> (Date of use: 16 August 2020).

917 NHS Digital “NHS digital, data and technology standards framework” <https://digital.nhs.uk/about-nhs-digital/our-work/nhs-digital-data-and-technology-standards/framework> (Date of use: 16 August 2020).

918 UK Code of Conduct for data-driven health and care technology <https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology> (Date of use: 23 July 2020).

919 UK Code of Conduct for data-driven health and care technology <https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology> (Date of use: 23 July 2020).

4.5.1.5. International organisations

Global organisations, for instance the OECD, have also considered the issue of ethics when it comes to AI. During May 2019, member countries of the OECD implemented the OECD AI Principles, in order to promote “innovative and trustworthy and that respects human rights and democratic values”.⁹²⁰ The OECD principles recognise five related values-centred principles towards realising trustworthy AI:

- (1) inclusive growth, sustainable development and well-being; (2) human-centred values and fairness; (3) transparency and explainability; (4) robustness, security and safety; and (5) accountability.

Notably, these principles are identical to those recognised under the EU guidelines. Despite the fact that the OECD principles are non-binding, they are increasingly persuasive, and as with the EU, it can set the global benchmark and assist governments to create domestic laws with regard to the development of trustworthy AI. For instance, outside of the OECD members, several other countries (such as Costa Rica, Malta, Peru, Argentina, Brazil, Ukraine and Romania), have already followed the AI Principles, with the number of supporters increasing.⁹²¹ Furthermore, the G20 trade ministers and digital economy ministers, approved AI Principles in June 2019, which are human-centred predominantly developed from the OECD Principles on AI.⁹²²

Another leading agency, the WHO, has also begun to examine the way in which the digital revolution for the healthcare industry can advance the condition and accessibility of healthcare globally. In March 2019, the WHO created a public dialogue on the first draft of the “Global strategy on digital health”.⁹²³ The consultation period closed on 3 May 2019. The Strategy considers aspects where the WHO and its global partners should give due regard to the potential application of digital-driven technologies, understanding the power of digital technologies and

920 OECD <http://www.oecd.org/going-digital/ai/principles/> (Date of use: 16 August 2020).

921 OECD <http://www.oecd.org/going-digital/ai/principles/> (Date of use: 16 August 2020).

922 G20 Trade Ministers “G20 ministerial statement on trade and digital economy” <http://www.g20.utoronto.ca/2019/2019-g20-trade.html> (Date of use: 9 July 2019).

923 WHO “Data and innovation: draft global strategy on digital health” https://apps.who.int/gb/ebwha/pdf_files/EB146/B146_26-en.pdf (Date of use: 3 July 2020).

to recognise its complexity. The draft strategy recognises four strategic objectives that aim to:

- (1) Promote global collaboration and advance the transfer of knowledge on digital health, (2) advance the implementation of national digital health strategies, (3) strengthen governance for digital health at global and national levels, and (4) advocate for people-centred health systems that are enabled by digital health.⁹²⁴

The strategy also recommends legalising, standardising and certifying AI based medical devices in order to establish uniform standards, which involves full pre-qualification evaluation for safety and quality control. The strategy encourages the global regulation of healthcare data and prescribes principles for proportionate data sharing in research towards safeguarding patients' interests and enhancing healthcare objectives.⁹²⁵

The IEEE global initiative on ethics of autonomous and intelligent systems (IEEE), published a document for high-level principles presented in the second version of "Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems".⁹²⁶ The document comprises the work of committees established under the IEEE, made up of participants from different constituents, including: academia; business; policy; and governments, associated with technological and human-centred disciplines who would identify imperative ethical and societal concerns relating to the advancement of AI technologies. The main concern of the IEEE is to oversee that engineers in science and technology are trained and empowered so as to prioritise ethics in the development of smart systems.⁹²⁷ In this regard, the IEEE identifies the following ethics principles:

- (1) The protection of human rights in the design and development of AI systems, (2) prioritising human well-being as an outcome in all system designs, using widely accepted, well-being metrics as their reference point, (3) assigning responsibility and accountability on designers, manufacturers, owners, and operators of AI systems, (4)

924 WHO "Data and innovation: draft global strategy on digital health https://apps.who.int/gb/ebwha/pdf_files/EB146/B146_26-en.pdf (Date of use: 3 July 2020).

925 WHO "Data and innovation: draft global strategy on digital health https://apps.who.int/gb/ebwha/pdf_files/EB146/B146_26-en.pdf (Date of use: 3 July 2020).

926 IEEE "Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent system, version 2" https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf (Date of use: 21 August 2020).

927 IEEE "Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent system, version 2" https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf (Date of use: 21 August 2020).

ensuring that AI systems are transparent and understandable, and (5) extending the benefits and minimising the risks of AI systems being misused through education and security awareness.⁹²⁸

The IEEE is also the driving force behind the series of IEEE Working group P7000-P7010 standards, aiming to provide a guide for AI designers during development.⁹²⁹ More recently, in an endeavour to create a universal standard framework towards the AI ethics, UNESCO published a first draft of the Recommendation on the Ethics of AI prepared by its Ad Hoc Expert Group, and circulated in September 2020 to member states for their comments.⁹³⁰

The values and principles recognised in the draft recommendation is to respect, protect and promote basic human rights and to address issues such as safety and security, by recognising social, cultural and economic representation (particularly due to limited resources of LMIC's), privacy and data protection, human oversight, fairness and non-discrimination.⁹³¹ The primary objective is for AI to be created and applied in accordance with principles of equality, multiplicity, and accountability.⁹³² The recommendation will be revised and is intended to be presented for adoption in November 2021 at the UNESCO 41st General Conference.⁹³³

928 IEEE "Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent system, version 2" https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf (Date of use: 21 August 2020).

929 IEEE P7000 – Model process for addressing ethical concerns during system design; IEEE P7001 – Transparency of autonomous systems; IEEE P7002 –Data privacy process; IEEE P7003 – Algorithmic bias considerations; IEEE P7004 – Standard on child and student data governance; IEEE P7005 – Standard for transparent employer data governance; IEEE P7006 – Standard for personal data artificial intelligence (AI) agent; IEEE P7007 – Ontological standard for ethically driven robotics and automation systems; IEEE P7008 – Standard for ethically driven nudging for robotic, intelligent, and automation systems; IEEE P7009 –Standard for fail-safe design of autonomous and semi-autonomous systems; and IEEE P7010 –Wellbeing metrics standard for ethical artificial intelligence and autonomous systems.

930 The first draft of the recommendations are available at <https://en.unesco.org/artificial-intelligence/ethics> (Date of use: 28 December 2020).

931 UNESCO "Principles and values as contained in the draft recommendation" <https://en.unesco.org/artificial-intelligence/ethics> (Date of use: 28 December 2020).

932 UNESCO <https://en.unesco.org/artificial-intelligence/ethics> (Date of use: 28 December 2020).

933 Available at UNESCO <https://en.unesco.org/artificial-intelligence/ethics> (Date of use: 28 December 2020).

4.5.1.6. Summary of the international position on the ethical guidelines

As evident from the previous discussion, several international missions are underway to tackle the ethical and societal concerns of the digital revolution of AI. Prominent examples of such initiatives broadly concern AI in general. The principles identified in most jurisdictions and organisations overlap. They focus on a human centred approach, which strives to ensure regard for human values and basic human rights as fundamental to the approach in how AI systems are to be designed, implemented, applied and controlled.

Common themes include: the importance of ensuring that AI systems are safe and reliable, respect for freedom, upholding fairness and avoiding bias, protection of privacy, and transparency and accountability in the application of AI. Therefore, there is a general belief that these fundamental principles are the foundation for AI's ethical development and deployment, irrespective of sector.

4.5.2. Ethical guidelines for AI in South Africa's healthcare

Akin to the challenges of our socio-economic landscape, the AI terrain in South Africa is unequal and fraught with regulatory issues.⁹³⁴ Unlike other jurisdictions, South Africa lags behind with the rapid pace of AI and is yet to formalise any guiding policies or legislation towards the regulation of AI, particularly for healthcare.

However, through the country's establishment of the Fourth Industrial Commission (4IR Commission), it aims to draft strategies, policies and approaches to hopefully place South Africa as an international participant and contributor in advanced technologies. More recently, various South African stakeholders pledged their commitment to responsible AI. At the "AI dialogue South Africa" (which took place virtually on 5 August 2020) industries, academia, non-profit organisations, and many other stakeholders signed the Expression of Interest (EoI).⁹³⁵ The EoI not only serves as a foundation for collaboration amongst various stakeholders, but seeks to

934 IT Web "SA organisations commit to responsible use of AI" <https://www.itweb.co.za/content/KA3WwqdDj24qrydZ> (Date of use: 26 August 2020).

935 IT-Online "Stakeholders commit to responsible AI" <https://it-online.co.za/2020/08/06/stakeholders-commit-to-responsible-ai/> (Date of use: 26 August 2020).

promote the accountable use of AI and to introduce a framework for ethical AI with an intended regulatory standard.

There are an increasing number of guidelines to address ethical AI. South Africa can certainly benefit from the extensive work already undertaken by the EU's High-Level Expert Group on AI, who has become the front-runner in establishing a comprehensive framework in translating general AI ethical principles into improved tangible guidance to address AI.

Grounding the EU principles on a human rights platform offers credence to their relevance and importance, which can be regarded as universal.⁹³⁶ In healthcare settings, these fundamental rights such as privacy, equality, safety and human autonomy are important ethical concerns. These very principles are entrenched in the observances that for the most part shape the UN Sustainable Development Goals (SDGs),⁹³⁷ such as enhancing health, and that impact development strategies in LMIC's outside the EU.⁹³⁸ Thus, the EU's ethical composition could be expected to exude authority across economic, social and cultural divides. The potential threats and negative effects that can ensue from AI systems, for instance, "on democracy, the rule of law and justice, or on the human mind"⁹³⁹ also become relevant wherever these establishments are positioned or aspired for.⁹⁴⁰

In view of this, in the following section, I thus primarily consider the EU's fundamental ethical principles for AI that can be applied towards South Africa's healthcare sector. Due regard is also given to the "candidate recommendations" and ethics principles of the IEEE Global Initiative, as a practical approach on ethical

936 Baeroc K, Miyata-Sturm A and Henden E "How to achieve trustworthy artificial intelligence for health" 2020 *WHO Bulletin* 3.

937 United Nations Sustainable Development Goals <https://www.un.org/sustainabledevelopment/sustainable-development-goals/> (Date of use: 16 August 2020)

938 Baeroc, Miyata-Sturm and Henden 2020 *WHO Bulletin* 3-4.

939 "European Commission's Ethics guidelines for trustworthy AI" <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

940 Baeroc, Miyata-Sturm and Henden 2020 *WHO Bulletin* 3-4.

operational design standards and principles that can be observed by the developers of AI systems.⁹⁴¹

4.5.2.1. Ethical principles for healthcare

The EU guidelines identifies three elements for trustworthy AI:

- (i) It should be *lawful* in compliance with all applicable laws and regulations,
- (ii) It should be *ethical* in adherence to ethical principles and values; and (iii) It should be *robust*: from a technical⁹⁴² and societal⁹⁴³ perspective.⁹⁴⁴

The EU guidelines do not focus on the lawful component, because there are existing national and international binding rules concerning AI, which consist of primary law,⁹⁴⁵ secondary law,⁹⁴⁶ and domain specific laws.⁹⁴⁷ For instance, the lawful component is applicable to the aspect of liability on the part of the producers and professional end-users where an AI system causes harm to another person, as discussed in the previous sections of this thesis.⁹⁴⁸

The EU guidelines operate on the assumption that all laws that relate to the activities concerning AI, remain legally binding and consequently is to be complied with. The EU guidelines only focus on the ethics and robustness components, centering trustworthy AI around a fundamental rights framework.⁹⁴⁹ These rights appear in

-
- 941 IEEE “Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent system, version 2” https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf (Date of use: 21 August 2020).
- 942 The technical robustness of the system is to ensure that AI is technically reliable, safe and can counter risks, see “European Commission’s Ethics guidelines for trustworthy AI” <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).
- 943 Ensuring the system’s societal robustness with due regard to the landscape in which it functions, see “European Commission’s Ethics guidelines for trustworthy AI” <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).
- 944 “European Commission’s Ethics guidelines for trustworthy AI” <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).
- 945 Primary law would include the EU Charters and Treaties on Basic Rights or the South African Constitution and Bill of Rights.
- 946 Secondary law such as the EU GDPR and South African Protection of Personal Information Act, and product liability directives under both the British and South African CPA.
- 947 Domain-specific laws exist that govern specific AI applications in certain sectors such as the medical device regulations in the healthcare sector.
- 948 Refer to section 4.3 and 4.4.
- 949 Baeroc, Miyata-Sturm and Henden 2020 *WHO Bulletin* 3.

many legal instruments through the observance of human dignity, human autonomy, fairness, privacy, citizen rights and justice.⁹⁵⁰ Respect for these fundamental rights, under the law, offers the basis for developing conceptual ethical standards. The EU guidelines identify four principles as entrenched in relation to such fundamental rights:

(i) respect for human autonomy,⁹⁵¹ (ii) prevention of harm,⁹⁵² (iii) fairness;⁹⁵³ and (iv) explicability⁹⁵⁴ [the report highlights that this is a non-exhaustive list].⁹⁵⁵

The above principles may, however, conflict with each other when identifying, implementing, and evaluating AI systems, for which there may be no ideal solution. This suggests that a value-judgement is required in the context of the various, and at times even opposing, principles.⁹⁵⁶ Conflicting principles are an inherent aspect of an evolving and redesigned value structure of any society.⁹⁵⁷ For instance, the use of AI systems to manage or support medical conditions could prevent serious illnesses, but it may require the observing of activities that infringe on people's freedom and right to privacy, causing the fundamental principles of "prevention of harm", and that of "human autonomy" to conflict.

It would thus be unreasonable to expect of AI producers to find the perfect solution with regard to all of the identified principles. However, they must address the ethical

950 The Bill of Rights enshrines the fundamental rights, that includes the right to dignity, freedoms, privacy, and equality. These fundamental rights are also enshrined in the EU treaties and charters and international human rights law. Other legal instruments also reflect some of these rights such as the GDPR or the POPIA that protect the right to privacy.

951 AI systems should not unreasonably manipulate, or condition humans and instead, they should be designed to enhance and complement human functions and knowledge. This means ensuring that there is human oversight and command relating to the working procedures of AI development.

952 AI systems inclusive of the landscape within which they function should be secure and should not cause harm. This encompasses the safeguarding of human dignity and respect.

953 The AI systems should be fair and free from bias and balance competing interests and objectives.

954 The abilities and intended objective of AI systems should be transparent and its outcomes must as far as possible must be explainable to the people who it affects.

955 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

956 This would also envisage the principle of competition steering consumer and public interests and other fundamental principles found in constitutional principles. See Waltz and Firth-Butterfield 2019 *Duke law & technology review* 228.

957 IEEE "Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent system, version 2" https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf (Date of use: 21 August 2020).

predicaments by identifying the methods which people employ to resolve ethical dilemmas and conflicts and consider it against how they imagine AI systems would resolve comparable ethical struggles. The system's solution to ethical challenges must be clear and transparent i.e, recorded by way of reasoned and evidence-based analysis rather than intuition or arbitrary discretion.⁹⁵⁸

4.5.2.2. Requirements for implementing ethics principles in healthcare

Most of the ethical principles discussed in the EU ethical guidelines are familiar to the healthcare industry. The delivery of healthcare solutions that are safe, harmless and inclusive are fundamental principles that have been entrenched in the application of healthcare technologies for many years.⁹⁵⁹ Therefore, the healthcare profession is possibly much better equipped to assimilate these notions into real world environments than any other sector would be. It is essential to ensure the strict execution of the ethical standards in healthcare settings. According to the sentiments expressed in the EU guidelines, in addition to being lawful:

AI must be grounded in fundamental rights, societal values, and the ethical principles of explicability, prevention of harm, fairness, and human autonomy.⁹⁶⁰

This must be with special focus on circumstances involving vulnerable populations and disproportionateness of data or control.

The EU ethical guidelines resonate with the principles of good “medical ethics”, i.e., “beneficence, non-maleficence, justice, and respect for human autonomy”,⁹⁶¹ with

958 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020); IEEE “Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent system, version 2”

https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf (Date of use: 21 August 2020).

959 Microsoft “Healthcare, artificial intelligence, data and ethics - A 2030 vision” <https://www.digitaleurope.org/wp/wp-content/uploads/2019/02/Healthcare-AI-Data-Ethics-2030-vision.pdf#page=28> (Date of use: 16 August 2020).

960 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

961 Gillon R “Defending the four principles approach as a good basis for good medical practice and therefore for good medical ethics” 2015 *J Med Ethics* 111.

the objective of protecting susceptible patients against social injustice and hierarchies. The first three are recognised principles of biomedical ethics.⁹⁶²

The principle of explicability is a new enabling principle, aimed to acquire an awareness as to how AI generates its outcomes or decisions, which is a significant factor for transparency, challenging outcomes based on AI, and finding appropriate chains of liability.⁹⁶³ These principles must be transformed into specific obligations in order to attain ethical AI in healthcare.⁹⁶⁴ The requirements embodied in the EU guidelines, come to the assistance with such transformation as specific interventions in medical practice.⁹⁶⁵ In terms of their various roles, a divergent group of participants can contribute to making sure that such requirements are observed:

(a) *Developers* should implement and apply the requirements to design and development processes, (b) *Deployers* should ensure that the systems they use and the products and services they offer observe the requirements,⁹⁶⁶ and (c) *End-users* and the broader society should be informed about these requirements and able to request that they are upheld.⁹⁶⁷

As illustrated in the table 2 below, there are requirements that are direct ethical principles such as explicability or autonomy, whilst other requirements are preconditions for fulfilling these principles, for instance human agency and oversight to achieve patients' autonomy.⁹⁶⁸

-
- 962 Beauchamp TL and Childres JF *Principles of biomedical ethics* (New York: Oxford University Press 2001), Gillon 2015 *J Med Ethics* 111.
- 963 Floridi L *et al* "AI4People - An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations" 2018 *Minds and Machines* 699-700.
- 964 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).
- 965 The requirements listed under the EU guidelines although not exhaustive are: "(1) human agency and oversight; (2) technical robustness and safety; (3) privacy and data governance; (4) transparency; (5) diversity; non-discrimination and fairness; (6) societal and environmental wellbeing; and (7) accountability".
- 966 They would apply to healthcare professionals that use AI programs in their clinical recommendations and decision making.
- 967 This would apply to patients who are treated and diagnosed by healthcare professionals with the assistance of AI systems or a broader group of people that may be impacted by this.
- 968 Gillon 2015 *J Med Ethics* 111; Beil M *et al* "Ethical considerations about artificial intelligence for prognostication in intensive care" 2019 *Intensive care medicine experimental* 4-7.

Table 2 – Ethics principles requirements to achieve trustworthy AI in healthcare (founded on the EU guidelines) ⁹⁶⁹	
Leading ethics principle	Requirement for implementation
<i>Traditional Biomedical Principles</i>	
Beneficence	Privacy and Data governance, Accountability
Non-maleficence	Technical robustness
Justice	Fairness, societal wellbeing
Autonomy	Human agency and oversight
<i>New Enabling Principle</i>	
Explicability	Transparency

The principles of doing good and no harm (beneficence and non-maleficence)

The fundamental notion of human self-respect suggests that, medical technology in healthcare should adhere to principles of beneficent (“do good”) and non-maleficent

969 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

“do no harm”)⁹⁷⁰ when patients are concerned.⁹⁷¹ These principles require that, in order to achieve trustworthy AI, there must be (1) accountability, (2) protection of privacy, and (3) technical robustness (safety, security, accuracy, reliability), all of which are intrinsically associated to the EU principle of prevention of harm. The following requirements can be considered in terms of fulfilling the principles of beneficence and non-maleficence:

Accountability: A few human actors are already involved in the design of AI technologies that are already drastically changing how we operate and function in our daily lives, motivating the importance of holding these actors to account for the outputs generated by AI. The requirement of accountability is expressed in the various initiatives to consider the ethical impacts related to AI, which is enhanced by the requirement of transparency and thus closely linked with transparency.⁹⁷² Mechanisms need to be implemented to make sure that there is accountability for AI systems and their outputs over the entire development cycle.⁹⁷³

The requirement of accountability compliments all principles of ethics. To ensure that AI is “beneficent and non-maleficent”, we should be in a position to identify harmful or constructive influences it can impose on society, and this can be done through impact assessments that entails recognising, evaluating, recording, reporting, and mitigating the probable risks during all stages of development of AI.

970 Beil *et al* 2019 *Intensive care medicine experimental*.

971 Floridi *et al* 2018 *Minds and Machines* 697

972 Asilomar “AI principles” <https://futureoflife.org/ai-principles/> (Date of use: 14 August 2020); “European Commission’s Ethics guidelines for trustworthy AI” <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020); OECD <http://www.oecd.org/going-digital/ai/principles/> (Date of use: 16 August 2020); UK House of Lords Select Committee on Artificial Intelligence Report on “AI in the UK: ready, willing and able?” <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf> (Date of use: 16 August 2020); UK’s Independent Review on Artificial Intelligence and Public Standards https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/868284/Web_Version_AI_and_Public_Standards.PDF (Date of use: 20 June 2020); <https://www.partnershiponai.org/#> (Date of use: 6 June 2020); <https://www.microsoft.com/en-us/research/theme/fate/> (Date of use: 15 August 2020); <https://www.Blog.google/technology/ai/ai-principles/> (Date of use: 15 August 2020); IEEE https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf (Date of use: 21 August 2020).

973 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

For AI to be *just*, we must ensure that lawmakers or the judiciary should pronounce on the issues of liability for AI systems so that individuals or corporations involved with the technology are answerable for the harm ensuing therefrom, and that redress is viable when things go wrong, and this in turn requires some knowledge of how an outcome from AI arose.

For AI to foster and not impede on the *autonomy* of people, the “decision about who should decide” must be akin to the awareness of how AI would behave over humans. Lastly, to achieve *explicability* in AI, the relationship between humans and the technology must be based on terms that are transparent and simple to the ordinary person “on the street”, with particular consideration to be given to susceptible populations of society.⁹⁷⁴

Closely associated to accountability, is the absolute right of patient consent. South African law has, over many years, imposed the obligation on healthcare providers to acquire the “informed consent” from a patient pertaining to any healthcare provided to the patient.⁹⁷⁵ Informed consent denotes that, reasonable and adequate information must be provided to patients so that they understand what they are consenting to, as well as the implications thereof. Informed consent is provided for under section 12(2) under the South African Constitution affirming that,

[e]veryone has the right to bodily and psychological integrity, which includes the right to make decisions concerning reproduction; to security in and control over their body; and not to be subjected to medical or scientific experiments without their informed consent.

“Informed consent” is defined in terms of the National Health Act,⁹⁷⁶ as,

consent for the provision of a specified health service given by a person with legal capacity to do so and who has been informed as contemplated in section 6.⁹⁷⁷

In terms of this Act, healthcare providers are obligated to notify patients, in simple and understandable language,⁹⁷⁸ of their health status;⁹⁷⁹ the treatment choices

974 Floridi *et al* 2018 *Minds and Machines* 700.
975 *Stoffberg v Elliot* 1923 CPD 148 148.
976 61 of 2003.
977 Section 7(2) of the National Health Act.
978 Section 6(2) of the National Health Act.
979 Section 6(1)(a) of the National Health Act.

accessible to them;⁹⁸⁰ and the advantages, fees and risks related with the treatment options.⁹⁸¹ The HPCSA's Ethical Guidelines for Good Practice in the Healthcare Professions, prescribes ethical guidelines for healthcare practitioners, which are not only legally binding on them, but also serve as a moral and ethical code of conduct.⁹⁸² According to the HPCSA guidelines:

Successful relationships between health care practitioners and patients depend upon mutual trust. To establish that trust practitioners must respect patients' autonomy – their right to decide whether or not to undergo any medical intervention, even where a refusal may result in harm to themselves or in their own death. Patients must be given sufficient information in a way that they can understand, to enable them to exercise their right to make informed decisions about their care. This is what is meant by an informed consent.⁹⁸³

The POPIA provides guidance as to when a consent would be required in a specific case relating to one's personal information. In the POPIA, consent is defined as: "any voluntary, specific and informed expression of will in terms of which permission is given for the processing of personal information".⁹⁸⁴ Informed consent, as prescribed by the definition, would mean specific and explicit consent for a legitimate purpose, made voluntarily by a data subject or an authorised person (where it pertains to information on a child).⁹⁸⁵

The POPIA further states that, "personal information must be collected for a specific, explicitly defined and lawful purpose related to a function or activity of the responsible party" and that, the affected party is to be informed of this purpose.⁹⁸⁶ When AI systems are used for treatment and diagnosis of patients in healthcare settings, the issue of patient consent from an ethical perspective of "accountability", is therefore a crucial aspect in the application of the bio- medical ethics framework for AI.

980 Section 6(1)(b) of the National Health Act.

981 Section 6(1)(c) of the National Health Act.

982 HPCSA https://www.hpcsa.co.za/Uploads/Professional_Practice/Ethics_Booklet.pdf (Date of use: 19 May 2020).

983 HPCSA https://www.hpcsa.co.za/Uploads/Professional_Practice/Ethics_Booklet.pdf (Date of use: 19 May 2020).

984 Section 1 of the POPIA.

985 Section 11(1)(a) of the POPIA.

986 Section 13(1) & 13(2) of the POPIA.

For this purpose, when AI is applied within medical devices, or as an aid or instrument to physicians in performing surgeries, the patient must be consulted regarding this use in precise detail and advised of any risks associated therewith. For AI to be accountable, it is essential that patients are able to appreciate the advantages and risks of the technology, so that they can partake in an educated decision on whether they wish to consent and proceed with the proposed care or surgery.

Privacy and data governance

Personal privacy, as recognised under the EU principle relating to the prevention of harm, is a fundamental human right that is of particular concern when it comes to AI systems in healthcare settings. Privacy demands acceptable governance to deal with data quality and reliability, its relevance in context of the AI systems that it will be used for, data collection, sharing and storing, access procedures, and the mitigation strategies to safeguard such data.⁹⁸⁷

Privacy and data protection

In healthcare, personal privacy is closely associated to how personal data is accessed, controlled and used. It is vital that when using AI systems, patient privacy and protection of data is ensured at all stages of the system's lifecycle, given the highly sensitive nature of such personal information which could lead to stigmatisation. Privacy regulations such as the EU's General Data Protection Regulation (GDPR)⁹⁸⁸ and the POPIA operate to generally safeguard an individual's information. The protection of privacy is also highly relevant to data used and applied in the creation and application of AI systems.

The POPIA protects the data subject in various ways, such as the rights to consent for sharing or access, rectification and obliteration, and remedies for infringement.⁹⁸⁹ Such rights would include patient information obtained before, during or after the AI

987 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

988 GDPR (EU) 2016/679 in European Union (EU) law regulates data protection for every EU citizen. It also covers dissemination of personal information outside of the EU jurisdiction.

989 Sections 18, 23, 24 and 74 of POPI.

system's use, such as outcomes generated by the system on patient conditions, or patients' progress based on recommendations or treatments undertaken by using the systems.⁹⁹⁰

Both the GDPR and the POPIA impose heavy penalties, premised on the type and extent of the infringement.⁹⁹¹ However, the POPIA imposes a harsher punishment of imprisonment as opposed to the GDPR. In South Africa, the National Health Act also renders it a violation to divulge patients' personal records in the absence of their consent, except in exceptional situations.⁹⁹²

Patient privacy is not only protected in law. The HPCSA's formal guidance: "Confidentiality: Protecting and Providing Information", embodies core tenets of confidentiality as crucial in the relationship between a healthcare practitioner and their patient.⁹⁹³ During 2015, Google's parent company (DeepMind), and the Royal Free London NHS Foundation Trust collaborated in the design of an application that assists in the identification of acute kidney injury (AKI).

During the design of the application, the Trust shared personal information of approximately 1.6 million patients to be employed for the trial and evaluation stages of the application's warning, identification and recognition system. The matter was submitted to the Information Commissioner's Office (ICO) for inquiry, and it delivered a ruling that the Trust had breached the UK Data Protection Act of 1998 in sharing the patients' data with DeepMind.⁹⁹⁴ Doteveryone stated that this matter illustrated what can be regarded as:

[M]any of the major issues at stake: the lack of competence public bodies has in negotiating AI agreements with the private sector; the potential for harm to privacy rights

990 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

991 Under GDPR penalties can include a warning or fines under Articles 83. Under Section 107 of the POPIA at penalties could include a warning, imprisonment and/or fines.

992 Sections 14, 15 and 16 of the National Health Act are significant in relation to patient confidentiality. Section 15 and 16 describes the circumstances under which patients' personal information can be divulged by a healthcare practitioner as "for any legitimate purpose within the ordinary course and scope of his or her duties where such access or disclosure is in the interests of the user".

993 HPCSA https://www.hpcs.co.za/Uploads/Professional_Practice/Ethics_Booklet.pdf (Date of use: 19 May 2020).

994 The Guardian <https://www.theguardian.com/technology/2017/jul/03/google-deepmind-16m-patient-royal-free-deal-data-protection-act> (Date of use: 13 June 2020).

and public trust in data transfers; and the giving away of valuable public data assets to private companies for free.⁹⁹⁵

All stakeholders engaged in the innovation lifecycle of AI systems, should have data security policies that serve to protect personal data. Implementation and enforcement of these policies can be achieved through information officers. The POPIA requires the nomination of an “information officer” charged with the responsibility of, amongst other things: overseeing that organisations conform to the requirements concerning “lawful processing of personal information”; and working together with the Regulator in relation to any investigations executed in terms of the POPIA.⁹⁹⁶ South African Healthcare Institutions must also provide for the nomination of an “information officer”, who would be responsible for safeguarding healthcare data and ensuring that such information is appropriately handled – as is the case with the NHS in UK, where the NHS is required to appoint a “Caldicott Guardian”⁹⁹⁷ for the same purpose.

In furthering the protection of health data, healthcare organisations in South Africa could also consider a national opt-out service, similar to the one implemented by the NHS in the UK. During May 2018, an online service called the “national data opt-out” was implemented in the UK. This permits patients to “opt out” from their information being used towards research and development, and all UK’s healthcare providers are required to adhere to the data opt-out policy. The service also allows patients to review their selection of opting out.⁹⁹⁸

Quality and integrity of data

Digital health data of patients may permit AI to extrapolate not only patient choices, but in addition patients attributes such as their race, age and sex. A way in which to make people comfortable with the manner in which information is collected about

995 Doteveryone “Written evidence (AIC0148)” <http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/artificial-intelligence-committee/artificial-intelligence/written/69653.html> (Date of use: 22 August 2020).

996 Section 1 and 4(1) of the POPIA.

997 UKGC “UK Caldicott Guardian Council” <https://www.ukcgc.uk/> (Date of use: 25 August 2020).

998 <https://digital.nhs.uk/services/national-data-opt-out> (Date of use: 13 June 2020).

them, is to see to it that such information is administered legally, reasonably and in an accountable manner.⁹⁹⁹ Data quality is important when it comes to the issue of how AI will perform. The data collected may be incomplete, inaccurate, misleading and contain errors.¹⁰⁰⁰

This issue must be resolved before training is undertaken on the system with any data sets. It is imperative that data integrity be preserved. Installing compromised data may alter an AI system's performance, particularly when machine-learning applications are used. Data sets must be tested, and processes must be recorded at every stage of the AI system's development lifecycle. This will also be applicable for AI systems that are designed and acquired elsewhere for specific applications in healthcare environments.¹⁰⁰¹

Access to data

The accessibility of data is integral towards the development of AI technologies. Access to data should ultimately be for the benefits generated therefrom for the public and financially for new products created as a result.¹⁰⁰² The healthcare industry is clearly positioned to gain from developments in AI, due to the accessibility of good quality and organised data, addressing a major obstacle in AI advancement.

What is important, is that while availability of patient information is enhanced, there must equally be adequate measures in place to safeguard such data.¹⁰⁰³ Policies and procedures for data access must be implemented by organisations that handle personal information, to avoid mismanagement cases and to define lawful use

999 Article 5 of GDPR and section 5 of the POPIA. The OECD appear to agree with this way of handling personal information but prefers the description "fair means" for processing, OECD "Guidelines on the protection of privacy and transborder flows of personal data" <https://www.oecd.org/sti/ieconomy/oecdguidelinesonthe protectionofprivacyandtransborderflows ofpersonaldata.htm> (Date of use: 22 August 2020).

1000 Article 5(1)(d) of GDPR and section 16(1) of the POPIA.

1001 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

1002 Doteveryone <http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/artificial-intelligence-committee/artificial-intelligence/written/69653.html> (Date of use: 22 August 2020). See also Waltz and Firth-Butterfield 2019 *Duke law technology review* 191.

cases. Only employees who are qualified and who require access to personal information ought to be authorised to do so.¹⁰⁰⁴ Organisations should also execute written agreements with the patients, which clearly set out the terms for sharing and use of data.¹⁰⁰⁵

Technical robustness

AI must be designed in a manner, where they end up behaving as projected while minimising unanticipated injury to others, and preventing of errors. Attaining desirable outputs from AI systems would also depend on addressing the trade-offs between realising its benefits, and diminishing the potential risks.¹⁰⁰⁶

General Safety

As a contingency plan to address potential problem cases, AI systems should have safety mechanisms in place. For instance, AI systems can be designed to switch over from a numerical to rule-based method, or that would need the involvement of a human operator before resuming their action. Furthermore, compliance- and risks migration strategies related to the use of systems, should be developed for various application fields. The degree of safety procedures needed is contingent on the extent of any threat that the AI system presents, which further hinges on its abilities.

The Asilomar Principles echoes this view, alluding to the dangers of systems designed to recursively self-improve, and that is it imperative to “caution” on the “upper limits on future AI capabilities”.¹⁰⁰⁷ Likewise, the Partnership on AI also emphasises the necessity for AI to function “within secure constraints”.¹⁰⁰⁸ Where it is projected that the process of development, or the AI design itself will have

1004 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

1005 IEEE “Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent system, version 2” https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf (Date of use: 21 August 2020).

1006 Floridi *et al* 2018 *Minds and Machines* 694.

1007 Asilomar <https://futureoflife.org/ai-principles/> (Date of use: 14 August 2020).

1008 <https://www.partnershiponai.org/#> (Date of use: 6 June 2020).

inherently high risks, this will necessitate that safety and quality controls be implemented and rigorous testing be done.¹⁰⁰⁹

Security

AI systems should be safe in terms of its developments, data sets and outputs, and it must be designed to be vigorous enough to tackle data abuse and any form of system attack. The system can be designed to include a function for “fail-safe” shutdown (such as for attacks), whilst still allowing for ongoing operation following the shutdown.¹⁰¹⁰ As with any software technologies, AI systems must be safeguarded against susceptibilities (such as cyber-attacks or hacking) that may target sensitive patient data – which is an area of concern in the protection of privacy.¹⁰¹¹

If the event of a system attack, both the data sets and system performance can be modified, causing the system to generate erroneous outcomes that may cause harm to patients. For instance, if somebody changes a blood type in a medical database, it could have ramifications for diagnosis and pose a threat one’s life. Thus, possible unintended uses of the AI system, and the likely misuse of the system by malevolent human actors cannot be ignored, and prevention mitigation measures are required to counter any harmful incidences.¹⁰¹²

The GDPR¹⁰¹³ and the POPIA¹⁰¹⁴ offers remedies to deal with the mismanagement of personal information. It would assist to raise public awareness about the potential misuse of AI systems by offering security training and initiatives, that alert the public to the inherent dangers of abuse of AI systems, for instance by invoking “data

1009 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

1010 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

1011 Doteveryone
<http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/artificial-intelligence-committee/artificial-intelligence/written/69653.html> (Date of use: 22 August 2020).

1012 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

1013 Article 77-84 of the GDPR,

1014 Section 73-99 of the POPIA.

privacy” that alerts the user that their personal data is stored and processed through the system.¹⁰¹⁵

Another form of defence against abuse of AI systems, is to enable as much people as is feasible to have access to it, and to balance the demand for it. If only made available to a few actors, it will render AI systems more powerful over any other individual.¹⁰¹⁶ Open science tools, such as those offered by the GSK, allow for sharing knowledge, data, research and intellectual outputs with the scientific and research community, without any financial, legal or technical constraints so that research or data required for the advancement of AI, is not concentrated in the hands of one actor.¹⁰¹⁷

Accuracy

This is associated with an AI system’s capability to render accurate decisions or outcomes. AI systems’ ability to make medical predictions, diagnosis and recommendations based on patient data or models, has the potential to infringe the principles due to inaccuracy. Conveying probabilities to patients that are generated by AI outcomes would also generate false expectations, anguish, or perpetual doubt.

An incorrect diagnosis generated from inappropriate data sets used for training an AI model, would produce ineffective and/or possibly unsuitable treatment interventions. A heightened degree of precision is particularly vital in circumstances where AI directly impacts the lives of human beings. The issues concerning inaccuracy, will affect all probable intervention methods applied in healthcare and not merely that of AI itself.¹⁰¹⁸

1015 IEEE “Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent system, version 2” https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf (Date of use: 21 August 2020).

1016 This was echoed by Elon Musk at the launch of OpenAI, an AI research and deployment non-profit company based in the US whose objective is to ensure that general AI benefits all of humanity, Boddington *Towards* 17-18.

1017 GSK “Open innovation” <https://uk.gsk.com/en-gb/research/sharing-our-research/open-innovation/> (Date of use: 12 June 2020).

1018 Hwang DY and White DB “Prognostication and ethics” in Shutter L and Molyneaux BJ (eds) *Neurocritical Care* (Oxford University Press 2018).

An informed assessment protocol can minimise and rectify inadvertent risks from erroneous diagnosis. A way to reduce erroneous outcomes is to personalise projections as far as possible, e.g., by having regard to more characteristics that describe the specific issues of patients.¹⁰¹⁹ When erroneous diagnoses is unavoidable, a system should be able to indicate how probable such errors are.¹⁰²⁰

Reliability and reproducibility

It is essential outcomes and decisions of AI systems can be reproduced and are reliable. An AI system is considered to be reliable when it operates correctly in a given range of data inputs and situations. This allows for an AI system to be analysed and to mitigate harm. Reproducibility is concerned with AI that can generate the same outcomes when reproduced under the same circumstances. This allows engineers and lawmakers to correctly explain what the AI systems' capability and how they do it.

The principle of justice

The justice principle imposes the requirement of fairness and societal well-being¹⁰²¹ and is usually applied in the allocation of resources within a population and the equal care of people, such as for new and trial therapies or general access to standard healthcare. For AI ethics, the principle means gaining equal access by way of inclusive design procedures and equitable treatment.¹⁰²²

In the context of AI, justice as an ethical principle allows for diversity, inclusiveness and equality through the entire life cycle of an AI system by enabling all individuals to gain access to AI products and services – notwithstanding their individualities,

1019 Hwang and White *Prognostication and ethics*.

1020 "European Commission's Ethics guidelines for trustworthy AI" <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

1021 OECD "Guidelines on Measuring Subjective Well-being states": "Being able to measure people's quality of life is fundamental when assessing the progress of societies. There is now widespread acknowledgement that measuring subjective well-being is an essential part of measuring quality of life alongside other social and economic dimensions", available at <http://www.oecd.org/statistics/oecd-guidelines-on-measuring-subjective-well-being-9789264191655-en.htm> (Date of use: 22 August 2020).

1022 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

capabilities or qualities.¹⁰²³ Across jurisdictions, justice concerns: (a) applying AI to rectify previous injustices, for instance, inequality; (b) guaranteeing that any benefits ensuing from AI's become available and accessible to all people; and (c) mitigating any new harms or threats (for instance the undervaluing of prevailing social systems).¹⁰²⁴

Fairness

Because AI technologies are data-driven, discrimination against a particular group of people will manifest due to type of data sets selected for training the systems, leading to increased bias and marginalisation.¹⁰²⁵ Unintended bias and poor governance protocols can unintentionally proliferate from various populations, by ignoring inherent rules rooted in the social structure of a particular ecosystem.¹⁰²⁶

Algorithms are trained using data sets of a particular patient and by using the same algorithm for other patients, it can generate inaccurate outcomes as the system did not learn the particulars of a broader range of patient categories. For instance, algorithms may assign a high mortality rate to formerly disadvantaged populations in rural settings, whose social standing was associated with a biased indicator, e.g., skin colour.

Unfair discrimination and bias should be eliminated during the data collection stage if possible. It is therefore imperative to use data-sets for training AI systems, which would represent the patient group that the trained application system will be used for. The manner in which the systems are designed (e.g., software coding), are also likely to generate prejudice. This can be addressed with suitable strategies for oversight processes and testing models. Continuously auditing outcomes to

1023 The Asilomar Principles prescribes both “shared benefit” and “shared prosperity” from AI, see Asilomar <https://futureoflife.org/ai-principles/> (Date of use: 14 August 2020).

1024 Floridi *et al* 2018 *Minds and Machines* 699.

1025 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

1026 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

evaluate and focus on the system's objective, limitations, and outcomes in a transparent and understandable way, will also reduce that risk.¹⁰²⁷

In enabling inclusion and diversity, AI systems must be designed on a user-centric module that reflects universal usage, and contributes to global justice, regardless of individualities, capabilities or characteristics. A universal design approach for AI must be considered to reach a comprehensive array of users, which would include future generations, following appropriate accessibility guidelines.¹⁰²⁸ This will allow for fair access to, and inclusiveness, of all individuals in present and evolving healthcare technological activities, particularly the most vulnerable and poor who are normally deprived from the benefits that costly new technology offers.¹⁰²⁹

Consultation with stakeholders would be recommended, as they could be directly or indirectly impacted by the technology through its entire development. Regular feedback is also vital, even after deployment, for example by securing customers and employees input and involvement within an organisation.¹⁰³⁰

Societal well-being

From an ethical standpoint, AI systems can be restrictively designed where they may be lawful and harmless with regards to their use, but fail to effectively contribute to societal well-being. This could lead to serious adverse effects on individuals' emotions, autonomy and other dimensions of well-being.¹⁰³¹ Ubiquitous exposure and growing reliance on AI systems, could also negatively impact on social relationships.¹⁰³²

1027 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

1028 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

1029 This connects to the United Nations Convention on the Rights of Persons with disabilities.

1030 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020); IEEE "Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent system, version 2" https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf (Date of use: 21 August 2020).

1032 Waltz and Firth-Butterfield 2019 *Duke law & technology review* 190.

In healthcare settings, patients may be at risk of isolation as AI systems continue to replace healthcare workers in certain tasks, such as healthcare robots in hospitals. This may have a bearing on a patient's psychological wellbeing.¹⁰³³ The impact of AI systems requires meticulous monitoring from a societal perspective, using the most effective and universally recognised well-being system of assessment as the benchmark.¹⁰³⁴ Organisations such as the OECD and IEEE are amongst the few that have issued guidelines for measuring the impact of social well-being, which may be a useful resource for assessing social well-being outputs in system designs.¹⁰³⁵

The principle of patients' autonomy

The principle is concerned with the requirement that AI systems must serve as a tool to enhance an individual's choices and judgements – requiring that the systems support an independent and autonomous society, fostering basic rights of choice and allowing for human oversight.¹⁰³⁶ It pronounces the necessity for achieving a balance amongst human- and AI-based judgements.¹⁰³⁷

Human autonomy

An acceptance of the patients' autonomy is an acceptance of their basic right of decision-making, as well as to make decisions premised on their own values and principles and regarding the treatment they wish to accept.¹⁰³⁸ From a medical perspective, autonomy can be limited when patients do not have the required mental capacity to make the best choices in regard to their own affairs, where their independence is then relinquished involuntarily.

1033 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

1034 IEEE "Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent system, version 2" https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf (Date of use: 21 August 2020).

1035 <http://www.oecd.org/statistics/oecd-guidelines-on-measuring-subjective-well-being-9789264191655-en.htm> (Date of use: 22 August 2020), IEEE P7010 Well-being metric for autonomous and intelligent systems <https://standards.ieee.org/standard/7010-2020.html> (Date of use: 22 August 2020).

1036 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

1037 Floridi *et al* 2018 *Minds and Machines* 698.

1038 Beauchamp and Childres *Principles of biomedical ethics*.

When it comes to AI, the position is very similar and it becomes more complicated. By accepting AI as a smart agent, we are accepting that some of our decision-making capability is being relinquished to these machines. Therefore, in upholding the principle of autonomy where AI is concerned, it would require achieving a balance between human beings' decision-making authority, which we choose to keep, and that which we elect to assign to smart agents.¹⁰³⁹ The overreliance on technologies by humans also impacts on their autonomy, depending on the extent of reliance for assistance in the performance of simple tasks, such as checking a patient's blood pressure. Because of the increased dependence on AI assistive technologies, fundamental cognitive abilities of humans could deteriorate – referred to as “digital dementia”.¹⁰⁴⁰

As a whole, human autonomy must be a key aspect of the system's design and application. In this regard, autonomy of human beings must be encouraged concomitantly with the limitation of AI autonomy, and any delegation must remain over-ridable in principle if individuals' autonomy needs to be reclaimed.¹⁰⁴¹ Parallel to this, is the users' right to make informed autonomous decisions, to decide which decisions to take, and not to be dependent on outcomes premised merely on AI automation that critically impacts them.¹⁰⁴² To uphold the autonomy of patients, the application of new medical procedures and practices involving AI must necessitate that the patient be clearly apprised of risks that may be involved with the treatment.¹⁰⁴³

Human oversight

It is important to ensure that the autonomy of humans is not overridden by AI systems, to the extent that it leads to adverse harm. The relationship of trust between patients and doctors is still integral to assisting decision-making in

1039 Floridi *et al* 2018 *Minds and Machines* 698.

1040 Waltz and Firth-Butterfield 2019 *Duke law & technology review* 192.

1041 Floridi *et al* 2018 *Minds and Machines* 698.

1042 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

1043 *Bundesgerichtshof* [BGH] [Federal Court of Justice] June 13, 2006, case no. VI ZR 323/04, NJW 2006, 2477 (Ger.); confirmed by decision March 27, 2007, case no. VI ZR 55/05, NJW 2007, 2767, 2769 (Ger.).

healthcare. Conventionally, the responsibility of principled decision-making is that of the doctor who has an obligation to ensure that the patient is apprised of his or her recommendations to enable the patient to decide thereon.

However, when the recommendations or decision-making is delegated to AI systems, the doctor must serve as the gatekeeper of a patient's data and as the guardian to their autonomy. A doctor's intervention could comprise of a choice to not rely on AI in a given scenario, to identify the degrees of the doctor's discretion at any stage when applying the AI system, or to safeguard that there are mechanisms in place to overrule the decision-making made by a system.¹⁰⁴⁴

The lower the oversight that a healthcare practitioner has in respect of an AI system, the greater the requirement for extensive testing and stricter governance throughout the entire lifecycle of the system.¹⁰⁴⁵

The principle of explicability

Over and above the other four medical ethics principles, new guidelines dealing with AI ethics also encompass explicability, i.e., transparency of aspects related to AI systems: the data sets, the design, and the patterns applied for generating outputs premised on particular inputs.¹⁰⁴⁶ The requirement of transparency designates technical approaches that can be integrated in the design phase of an AI system.

Without any clarification in terms of reasoning for decisions in specific cases, relinquishing medical judgments to "black box" AI systems, can be viewed as violating the inherent ethical duties of healthcare professionals. Transparency in the advancement of AI will foster confidence and trust towards the integration of smart machine outputs into healthcare. A key area of concern regarding AI, is that its

1044 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

1045 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

1046 London AJ "Artificial intelligence and black-box medical decisions: Accuracy versus explainability" 2019 *Hastings Cent Rep* 15-21.

functionality should be clear and transparent to the different stakeholders for distinct purposes.

Notably, the extent of transparency will essentially vary for the different stakeholders.¹⁰⁴⁷ This opacity of an AI system, combined with the decentralised way in which it is often designed, presents challenges to determine and assign responsibility when the system malfunctions. Consequently, the absence of transparency heightens the risk and extent of damage (i.e., end-users not grasping the technology they rely on), and simultaneously augments the obstacle of safeguarding accountability.¹⁰⁴⁸

Traceability

“Black-box” software systems should only be deployed with due care, as they usually generate outcomes that are not capable of being fully scrutinised, supported, or explained by conventional methods, and thus can lead to unidentified or unexpected faults, biases, and injuries.¹⁰⁴⁹ Traceability allows for transparency. The data sets and methods that generate AI’s outputs, inclusive of BD and cataloguing, and the algorithms used, requires a high standard of recording so that there can be an objective evaluation of the AI systems. Degrees of conformity can be established to allow for traceability and improving transparency.¹⁰⁵⁰

Traceability to its fullest degree may not be achievable, and other methods to assess outputs must be considered to ensure that the medical ethics principles are

-
- 1047 For users, transparency offers an easy way for them to understand how that system functions and the reasons why, for accreditation and certification as transparency reveals the system’s procedures and input of data, if an AI system malfunctions it will need to be transparent to a claims investigator or a court and transparency to society at large is necessary to foster public trust and in AI technology, see <https://standards.ieee.org/standard/7010-2020.html> (Date of use: 22 August 2020).
- 1048 IEEE “Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent system, version 2” https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf (Date of use: 21 August 2020).
- 1049 IEEE “Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent system, version 2” https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf (Date of use: 21 August 2020).
- 1050 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

observed.¹⁰⁵¹ Transparency could also be encouraged, either through laws requiring adoption of a sector code and criteria for AI systems that are marketed for commerce, or through indirect approaches such as tax rebates or standards that reduce the liability of producers who develop more transparent systems.¹⁰⁵² This approach should also extend to the decision-making of the AI system.

This allows for detection as to why an AI-outcome was flawed, which would avert impending errors and also facilitate clarification on issues of ethical and legal liability in situation of errors.¹⁰⁵³

Explainability

This standard involves the capability to clarify the AI system's technological methods and associated human judgments (e.g. in its operational fields) through transparent and traceable standards.¹⁰⁵⁴ These standards could entail preferential implementation of operative design practices for constructing "explainable AI (XAI)" systems¹⁰⁵⁵ that is able to offer validating reasons or some other dependable "explanatory" information illustrating the reasoning methods leading to, as well as prominent sources for, their deductions.¹⁰⁵⁶

Furthermore, trade-offs can be effected between augmenting AI's explainability (which could decrease its precision or power), or improving its precision or power (at the expense of explainability).¹⁰⁵⁷ It is important that in domains where AI has a major bearing on human beings, such as in healthcare, it must be possible to request reasonable clarification for AI's thought processes. As one of the

1051 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

1052 Scherer 2016 *Harvard journal of law and technology* 374.

1053 London 2019 *Hastings Cent Rep* 15-21.

1054 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

1055 Explainable AI (XAI) is an area of research that tries to address this concern in an attempt to more effectively understand the inherent mechanisms of a system.

1056 IEEE "Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent system, version 2" https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf (Date of use: 21 August 2020).

1057 London 2019 *Hastings Cent Rep.* 15-21.

stakeholders concerned, it can be demanded of clinicians to motivate their decisions in the absence of assurance or trust.

Therefore, incessant observation of AI behaviour and patient impact, must become compulsory to standardise these processes alongside ethical norms and diverting from them will require justification. Furthermore, justifications of the extent to which AI systems impact and influence the organisational decision-making methods, model selections for the system, inclusive of any underlying principle for implementing same, ought to be accessible – thereby guaranteeing “business model transparency”.¹⁰⁵⁸

Communication

Patients are entitled to be notified by healthcare practitioners when they are interfacing with an AI system. This goes to the issue of the absolute right of patient consent for all medical treatment, as was discussed under the requirement of accountability – relating to the principles of beneficence and non-maleficence. This right includes the right to decline any interfacing with AI in favour of only human contact, to ensure conformity with basic rights. Furthermore, an AI system’s abilities and shortcomings must be explained to AI practitioners or operators under circumstances where it is applicable to the use case in the relevant scenario.¹⁰⁵⁹

4.5.2.3. AI assessment list

The EU guidelines considered a limitless “trustworthy AI assessment list”, to put into operation the identified ethical principles of AI, to help assess the type of threats that an AI system could produce, and how to mitigate same while augmenting the

1058 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

1059 This accords with the strict (no-fault) product liability of South Africa, UK and US which passes to the seller or distributor or persons deemed to be part of the distribution chain for failing to provide adequate “instructions or warnings [...] is defective because of inadequate instructions or warnings and when the foreseeable risks of harm posed by the product could have been reduced or avoided by the provision of reasonable instructions or warnings by the seller or other distributor”. Refer to sections 4.3.7, 4.4.1.7 and 4.4.2.8 on the discussions relating to product liability.

opportunities offered by the system.¹⁰⁶⁰ Following a feedback process on the pilot version of this assessment list, a revised version was issued by the EU in July 2020.¹⁰⁶¹

This assessment list is designed for adaptable application and allows corporations to consider, or add aspects appropriate to the specific AI system as they deem necessary, with due regard to the industry in which they function.¹⁰⁶² Thus, the flexibility of the assessment list serves as a useful application for the healthcare sector. It should be noted that adherence to this assessment list should not be construed as compliance with the law, nor does it serve as guidance to ensure adherence to relevant laws.

The assessment list relates to AI systems which end-users directly interact with, and is mainly directed to data scientists, producers and implementers of AI systems. It would therefore also be applicable to healthcare professionals that use AI systems in their daily clinical practices. By way of vigorous interaction with the questions that the assessment list raises, it encourages insightful deliberation to elicit the correct engagement, and to cultivate a corporate ethos in creating and promoting ethical AI.¹⁰⁶³ Drawing from the EU assessment list, the below assessment list can be considered for AI driven technologies that are designed, deployed and applied for healthcare practices in South Africa:

ASSESSMENT LIST FOR ETHICAL AND TRUSTWORTHY AI¹⁰⁶⁴

1. Privacy, Data Governance and Accountability

*Privacy*¹⁰⁶⁵

1060 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

1061 European Commission “The assessment list for trustworthy artificial intelligence (ALTAI)” <https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment> (Date of use: 25 August 2020).

1062 <https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment> (Date of use: 25 August 2020).

1063 <https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment> (Date of use: 25 August 2020).

1064 <https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment> (Date of use: 25 August 2020)

1065 <https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment> (Date of use: 25 August 2020).

- ✓ Have you considered the effect of the AI system with regard to the following rights: privacy, physical, mental or moral integrity and data protection?
- ✓ For the specific case, have you considered processes that allow for flagging concerns on data collection and administering for the AI system's training and operation?

*Data Governance*¹⁰⁶⁶

- ✓ Is your AI system being trained, or was it developed, by using or processing personal information?
- ✓ Were the following measures implemented, some of which are mandatory under the POPIA?¹⁰⁶⁷
 - Processing limitations placing stringent restrictions on lawfully processing data;¹⁰⁶⁸
 - Purpose specification assessing the nature and extent of data in the applied data sets (for instance to determine if they include personal data);¹⁰⁶⁹
 - Information quality controls for data that is gathered and processed to be correct and complete;¹⁰⁷⁰
 - Data subject participation allowing access to the data subject's personal records and to call for rectifications of their record;¹⁰⁷¹
 - Designate an Information Officer and include such Information Officer at the earliest possible time of the design, procurement or application stage of the AI system;¹⁰⁷²
 - Measures to achieve security safeguards on the integrity and confidentiality of personal information (e.g., encryption, pseudonymisation, aggregation, anonymisation).¹⁰⁷³
- ✓ Data Access:¹⁰⁷⁴
 - What strategies were implemented to administer and enforce appropriate data governance?
 - Have you identified who would be authorised to access a data subject's information, and the conditions under which they would be authorised to do so?
 - Have you determined if the authorised persons are eligible and should have permission to obtain the data, and if they possess the required proficiency to appreciate the particulars of data privacy and protection policies?
 - Have you implemented a system to monitor and record the date, time, manner of access and person who accessed the data (including limiting access to qualified personnel, mechanisms for logging data access and making modifications)?

1066 <https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-atai-self-assessment> (Date of use: 25 August 2020).

1067 Protection of Personal Information Act No 4 of 2013.

1068 Condition 1- 4 of the POPIA.

1069 Condition 3 of the POPIA.

1070 Condition 5 of the POPIA.

1071 Condition 8 of the POPIA.

1072 Section 39 of the POPIA.

1073 Condition 7 of the POPIA.

1074 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

*Accountability*¹⁰⁷⁵

- ✓ Audibility
- Have you implemented methods which enable the AI system's auditability, for instance transparency of the development process, the sourcing of training data and the recording of procedures, outputs and consequences of the AI system?
- With due regard to the AI system impacting basic rights, have you ensured that it can be independently audited?

*Risk Management*¹⁰⁷⁶

- ✓ Have you performed any kind of risk evaluation relevant to the AI system, to oversee ethical concerns and accountability measures?
- ✓ Has auditing procedures been implemented to manage ethical conduct and liability? Does the involvement of these third parties go beyond the development phase?
- ✓ Have you arranged for risk training and education to assist with developing accountability practices and, if so, does this also inform about the possible regulatory approach relevant to the AI system?
- ✓ Have you considered the establishment of an AI ethical review panel or committee to deliberate on the general accountability and ethical protocols, including any possible areas of uncertainty?
- ✓ Have you established a process to discuss and continuously monitor and assess the AI system's adherence to this assessment list? Does this process include identification and documentation of conflicts between the different ethical principles and explanation of the 'trade-off' decisions made?
- ✓ Have you considered a mechanism for various stakeholders (e.g., sellers, end-users, subjects, manufactures or employees) to report possible susceptibilities, threats or prejudices in the AI system?
- ✓ For applications that can adversely affect individuals have you provided suitable mechanisms to facilitate possible redress in respect of any harm or undesirable impact ensuing from the AI system?

2. Technical Robustness

*Resilience to Attack and Security*¹⁰⁷⁷

- ✓ Could the AI system have adversarial, critical or damaging effects (e.g., to human or societal safety) in case of risks or threats such as design or technical faults, defects, outages, attacks, misuse, inappropriate or malicious use?

1075 <https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment> (Date of use: 25 August 2020).

1076 <https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment> (Date of use: 25 August 2020).

1077 <https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment> (Date of use: 25 August 2020).

- ✓ Is the AI system certified for cybersecurity or anti-virus attacks or is it compliant with specific security standards?
- ✓ Have you identified any possible areas of attacks or vulnerabilities to which the AI system could be exposed to? Have you evaluated the various modes of vulnerabilities and potential entry points for attacks such as manipulation of training data?
- ✓ Are there adequate measures to safeguard and test the reliability, resilience and general security against possible attacks of the AI system over its lifecycle, including verifying how the system performs in unanticipated conditions and settings?
- ✓ Have you informed end-users of the duration of security coverage and updates and expected timeframe within which you provide security updates for the AI system?

*Safety*¹⁰⁷⁸

- ✓ Have you defined risks, risk metrics and risk levels of the AI system in each specific use case?
- ✓ Have you identified and assessed the possible risks or threats to the AI system (design faults, technical faults, environmental threats) and is there an adequate contingency plan if the system experiences unanticipated adversarial conditions (for example technical automatic converting actions or requesting human intervention before continuing)?
- ✓ Have you informed end-users of existing or potential risks or threats?
- ✓ Have you assessed the risk of possible malicious use, misuse or inappropriate use of the AI system and is there a mitigation plan to handle possible risks?
- ✓ Have you implemented a process to continuously measure and assess risks?
- ✓ Have you reviewed liability and consumer safety regulations,¹⁰⁷⁹ and implemented them where required?
- ✓ Have you projected the probable damage or consequences due to a breakdown of the AI system where it generates inaccurate outcomes or decisions, becomes inaccessible, or generates outcomes that adversely impact on societal values (e.g bias)?
- ✓ To counter any possible liability for loss or injury ensuing from the AI system, has an insurance policy been considered?

*Accuracy*¹⁰⁸⁰

- ✓ Could a minimal state of precision of the AI system result in critical, adversarial or damaging consequences and is more data required, for instance to augment accuracy and avoid harm such as discrimination?
- ✓ Have measures been implemented to safeguard that data (including training data) applied to develop the AI system is comprehensive and up-to-date, of high quality, and representative of the environment the system will be deployed in?

1078 <https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment> (Date of use: 25 August 2020).

1079 For instance, the CPA 68 of 2008.

1080 <https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment> (Date of use: 25 August 2020).

- ✓ Have you implemented measures to oversee, as well as document the AI system's accuracy?
- ✓ Have you considered whether the AI system's functionality can invalidate the data or assumptions it was trained on, and how this might lead to harmful effects?
- ✓ Have you implemented procedures to ascertain that the level of precision of the AI system to be expected by end-users is properly communicated?

*Reliability and Reproducibility*¹⁰⁸¹

- ✓ Could the AI system cause significant or damaging consequences (e.g., pertaining to human safety) in case of low reliability or duplicability?
- ✓ Have you implemented a well-defined process to oversee whether the AI system is achieving the intended targets, objectives, and functionalities?¹⁰⁸²
- ✓ Have you determined if specific environments or situations must be considered to ensure reliability and duplicability?
- ✓ Have you implemented authentication and validation procedures and documentation (e.g., logging) to assess and incorporate the AI system's different features for reliability and duplicability?
- ✓ Have you clearly logged and implemented procedures for the assessment and authentication of reliability and duplicability of the AI system?
- ✓ Have you defined tested failsafe fallback plans to address AI system errors of whatever origin and put governance procedures in place to trigger them?
- ✓ Have you developed systems of reporting to reassure third parties regarding the system's reliability?

3. Fairness and Societal Well-Being

*Fairness*¹⁰⁸³

- ✓ Have you established mitigation strategies to counter generating or strengthening of unwarranted bias from the AI system, both from the data that is applied and the construct of the algorithms?
 - Have you tested for particular population groups or challenging use cases?
 - Have you explored and used publicly accessible state-of-the-art technological systems, to enhance your knowledge of the data, patterns and applications?
 - Have you assessed and implemented procedures to review and observe for probable biases during in the entire lifespan of the AI system (e.g., biases due

1081 <https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment> (Date of use: 25 August 2020).

1082 Monitoring by a supervisory mechanism is a way to verify that the system operates as intended as opposed to performance metrics that are merely an abstraction of the actual system behaviour.

1083 <https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment> (Date of use: 25 August 2020).

to probable restrictions ensuing from the configuration of the data sets that are applied (absence of diversity, marginalisation)?

- Where relevant, have you taken in account diversity and non-marginalisation of end-users and or data subjects in the data that is applied?
- ✓ Have you implemented educational and training awareness initiatives to help AI designers and developers be more aware of the possible bias they can inject in designing and developing the AI system?
- ✓ Have you ensured a mechanism that allows for the flagging of matters associated with prejudice, unfairness, or inadequate behaviour of the AI system?
- Have you provided a system or platform for communicating or raising such issues?
- Have you identified the subjects that could potentially be impacted due to the operation of an AI system, over and above any end-users?
- ✓ Is your definition of fairness commonly used and is it an acceptable and workable definition of “fairness” when it comes to the designing AI systems?
- Have you considered alternative definitions of fairness prior to selecting the one you use?
- Have you consulted with the impacted communities about the correct definition of fairness, i.e., representatives of elderly persons or persons with disabilities?
- Have you implemented a system of measurement to evaluate and assess the fairness definition that you use?
- Have you created procedures to ascertain fairness in the AI system that you use?
- ✓ To allow for fair access and inclusiveness of all individuals in current and developing technologies, have you determined that the AI system allows for a broad selection of inclinations and proficiencies in communities irrespective of their individualities, capabilities or characteristics?
- ✓ Have you engaged or consulted with end-users during the development and construction stage of the AI system, to assess whether its interface is accessible and usable particularly by people with special requirements, physical incapacities or that are at risk of marginalisation?
- ✓ Have you ensured that universal design standards to address a broad spectrum of users are considered during every step of the planning and development process, if applicable?
- ✓ Have you considered the possible impact of the AI system on end-users?
- Have you assessed whether certain populations or communities who could be unduly impacted by the outcomes of the AI system?
- Have you assessed the risk of the possible unfairness of the system onto the end-user's communities?

*Societal Well-Being*¹⁰⁸⁴

1084 <https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-atai-self-assessment> (Date of use: 25 August 2020).

- ✓ Have you evaluated the social implications of the AI system's application over and above the end-user, for instance stakeholders that may be indirectly impacted or society at large?
- ✓ Have you ascertained that the social implications of the AI system on society at large are fairly appreciated? For instance, have you considered if there is a risk of reduction or deskilling of the labour force? What measures have been implemented to offset such risks?
- ✓ Have you taken action to minimise potential societal harm of the AI system?
- ✓ Have you taken measures to make sure that the AI system does not negatively impact democracy?
- ✓ Have you made sure that the AI system patently conveys that any interaction with it is virtual and that it is incapable of expressing any emotion?

4. Human Agency and Autonomy

Human Agency and Autonomy¹⁰⁸⁵

- ✓ Does the AI system interact with humans, recommend or take decisions that affect humans or society?
- ✓ For pre-programmed chatbots or other communicative systems, are the end-users sufficiently informed that the AI system and not a human agent is interacting with them?
- ✓ Are end-users adequately informed that any recommendation, conclusion, or output is the result of the system's determination?
- ✓ Could the AI system impact on human autonomy by generating over-reliance on the system?
- ✓ Have you implemented procedures to circumvent over-reliance on the AI system?
- ✓ Could human autonomy be impacted by the AI system intervening in the end-user's choice or selection process involving an inadvertent approach?
- ✓ Have you implemented any procedure to avoid that the AI system unintentionally affects human autonomy?
- ✓ Have you implemented procedures to deal with possible negative consequences for end-users in case they develop an imbalanced attachment to the AI system?
- ✓ Has the AI system been deployed into the workforce systems? If so, has due been given with regard to the work distribution between the system and human workforce and suitable level of human supervision and control?

Human Oversight¹⁰⁸⁶

1085 <https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment> (Date of use: 25 August 2020).

1086 <https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment> (Date of use: 25 August 2020).

- ✓ Assess whether the AI system:
 - Is a self-learning or autonomous system;
 - Is overseen by a Human-in-the-Loop;¹⁰⁸⁷
 - Is overseen by a Human-on-the-Loop;¹⁰⁸⁸
 - Is overseen by a Human-in-Command.¹⁰⁸⁹
- ✓ Have the humans (“human-in-the-loop, human-on-the-loop, human-in-command”) been given specific guidance and training on how to exercise oversight?
- ✓ Have you established any detection and response mechanisms for undesirable adverse effects of the AI system for the end-user or subject?
- ✓ Have you implemented a “fail-safe” device or other mechanism to securely abandon the AI system’s operation when needed?
- ✓ Have you implemented any specific mechanisms of detection, oversight and control signifying the self-learning or autonomous nature of the AI system?

1. Transparency

*Traceability*¹⁰⁹⁰

- ✓ Have you implemented measures that address the AI system’s traceability in its lifespan?
 - Have you implemented processes to constantly review the suitability and accuracy of data sets applied to train the system?¹⁰⁹¹
 - Can you conduct an audit as to which data sets were applied by the AI system to produce an outcome?
 - Can you conduct an audit on which AI procedures produced the outcome of the AI system?
 - Have you implemented methods to continuously test and validate the type of the outcomes produced?¹⁰⁹²

1087 Human-in-the-loop refers to the capability for human involvement in each decision lifecycle of an AI system.

1088 Human-on-the-loop refers to the capability for human involvement throughout the design stage of an AI system.

1089 Human-in-command refers to the capability for oversight on the entire activities of an AI system (this would include wider financial, social and ethico-legal issues arising) and the capability to decide the manner and when the AI system should be applied in any given scenario.

1090 <https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment> (Date of use: 25 August 2020).

1091 “This could take the form of a standard automated quality assessment of data input: quantifying missing values and gaps in the data; exploring breaks in the data supply; detecting when data is insufficient for a task; detecting when the input data is erroneous, incorrect, inaccurate or mismatched in format”– for instance, health records are incorrectly recorded, <https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment> (Date of use: 25 August 2020).

1092 “This could take the form of a standard automated quality assessment of AI output: e.g., predicted scores are within expected ranges; anomalies in output are detected and input data leading to the anomaly detected and corrected”, see <https://ec.europa.eu/digital->

- Are there acceptable recording procedures employed to record the outcomes?

*Explainability*¹⁰⁹³

- ✓ Have you explained and clarified the decision or recommendations generated by the AI system to end-users?¹⁰⁹⁴
- ✓ Are surveys regularly conducted with end-users to establish if they appreciate the outcomes of the AI system?
- ✓ Have you evaluated why this specific system was implemented in this particular region or domain?¹⁰⁹⁵
- ✓ Have you designed the AI system with due consideration of the explainability thereof from the outset?
- ✓ Have you attempted to make use of the easiest and most explicable method for the AI system's application in the given situation?
- ✓ Have you determined whether you are able to assess interpretability after the system's training and development?

*Communication*¹⁰⁹⁶

- ✓ For AI systems such as assistive robots, have you adequately informed end-users when they are not communicating and engaging with another person?
- ✓ Have specific use cases been identified for the AI system and clearly conveyed this to ensure that it is explicable and appropriate in respect of the targeted market?
- ✓ Have you implemented adequate procedures to notify end-users regarding the objectives, gains and shortcomings of the outcomes produced by the AI system?
- Have you communicated possible threats (e.g bias) that the AI system can pose?¹⁰⁹⁷
- Has sufficient education material, instructions and warnings been provided to end-users regarding the AI system?
- ✓ Have you implemented structures that take into account users' views and comments and applied this to modify the system?

It should be stated that the above assessment list is not an exhaustive list. The assessment of AI's ethical impacts are firmly premised on the applicable social and

[single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment](https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment) (Date of use: 25 August 2020).

1093 <https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment> (Date of use: 25 August 2020).

1094 This requirement is for the developers. Users must understand the AI system and any misunderstandings must be addressed by the developing team.

1095 This requirement also relates to healthcare professionals who elect to use a specific AI program for an intended purpose, e.g., treatment, diagnosis, recommendations.

1096 <https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment> (Date of use: 25 August 2020).

1097 With regard to the system's development to those that are involved in its deployment and with regard to the system's application to the end-user.

financial framework requirements.¹⁰⁹⁸ The reason for this is because a users' appetite for risk differs, on account of deviations in the value structures and socio-economic experiences at ground level.¹⁰⁹⁹ An ethical assessment could thus differ from the viewpoint of LMIC's, such as South Africa, to address their challenging socio-economic needs and opportunities, where increasing access to healthcare is a good way to narrow the gap between developed and under-developed nations.¹¹⁰⁰

Ethical AI is not achieved by ticking off boxes, but rather by endlessly determining requirements within cultural and economic frameworks, assessing and fostering better outcomes throughout AI's development, and consulting stakeholders that are impacted directly or indirectly.

4.5.2.4. Technical methods: Ethics-by-design

It becomes necessary to determine how AI systems can be technically designed to allow for the systems to perform in an ethical way, as a minimum in certain perilous use cases scenarios ("ethics-by-design").¹¹⁰¹ Therefore, this section considers the approach to designing and operating AI systems to perform in an ethical way.

Architectures

There needs to be transparency in terms of the following:

- The algorithm's functions or capabilities;
- The algorithm's learning protocols;
- The algorithm's readiness for use;

1098 Waltz and Firth-Butterfield 2019 *Duke law & technology review* 225.

1099 Mialhe <https://cpr.unu.edu/ai-global-governance-why-we-need-an-intergovernmental-panel-for-artificial-intelligence.html> (Date of use: 29 August 2020).

1100 A PwC study conducted in 2017 revealed that in Nigeria, Turkey, and South Africa, nations whose advancement and access to basic services are of a priority, its citizens were twice as desirous as people in the UK, Germany, and Belgium to undergo significant surgery undertaken by AI robotics, notwithstanding current shortcomings with the technology, see PwC <https://www.pwc.com/gx/en/industries/healthcare/publications/ai-robotics-new-health/ai-robotics-new-health.pdf> (Date of use: 29 August 2020).

1101 Waltz and Firth-Butterfield 2019 *Duke law & technology review* 198.

- How the decision was arrived at for acceptance of the algorithm in relation to its use (for instance, was such decision taken by a panel, and what evidence was considered to reach this decision?); and
- The possible resource impacts.¹¹⁰²

Protocols and/or any restrictions on protocols must be embedded into an AI system's structural design. Same can be achieved by way of established "white-list" instructions (actions or conditions) which a system is required to execute, "black-list" limitations on actions or conditions that it must not contravene, and combinations of these instructions or securities concerning the system's performance.¹¹⁰³

ML AI systems have the capability to exhibit unanticipated conduct. Usually, they are studied in terms of a theoretical approach of a "sense-plan-act"-step cycle. Acclimating this structural design requires incorporation during all three steps: (i) during the "sense"-step, the system must be designed to enable it to recognise all elements in its settings needed to achieve compliance regarding the requirements; (ii) during the "plan"-step, the system must exclusively recognise and accept strategies which meet the requirements; and (iii) during the "act"-step, the system's performance must be constrained to actions which meet the requirements.¹¹⁰⁴

Casuistic versus dogmatic methods

Ethics can further be incorporated into AI systems based either on a casuistic or a dogmatic method. The casuistic method entails that AI systems are to be designed based on how to respond in a particular way in each scenario, by anticipating all possible scenarios in their decision-making process, where they may have to take

1102 UK Code of Conduct for data-driven health and care technology
<https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology>
 (Date of use: 23 July 2020).

1103 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

1104 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

an ethical decision. This can be achieved by applying the “sense-plan-act” cycle described above.¹¹⁰⁵

For instance, a robot for healthcare could be designed to regard the patient’s desire before it acts. If the patient does not convey his or her desire in each scenario, the robot should request the patient’s approval before acting. When no prior instruction is available, difficulties will ensue, and the patient will no longer be able to convey its desire.

With due regard to patient safety, a fail-safe mechanism of the AI system can be implemented in such instance that allows for the best prospect of preventing injury or loss to the patient.¹¹⁰⁶ Designers can enhance the structural design of their systems with fail-safe mechanisms that manage unexpected ethical breaches. For example, a “guardian AI” that would technically intervene in AI systems and rectify decisions that are contrary to law or ethical principles.¹¹⁰⁷

Rather than the casuistic method, AI systems could be designed on the basis of a dogmatic method that follows a specific ethical school of thought in its decision-making process – such the “Three Laws of Robots” by Asimov,¹¹⁰⁸ or the primary rule of many philosophies or religions that ascribes to not treating others in a way that one would not themselves want to be treated.¹¹⁰⁹

This method is premised on the concept that observance to ethics could be incorporated into the architecture of AI systems.¹¹¹⁰ The main concern with this concept however, is that it would instinctively only follow that school of thought, causing it to be a very rigid method, and could ensue in an outcome or decision that

1105 Waltz and Firth-Butterfield 2019 *Duke law & technology review* 199.

1106 Waltz and Firth-Butterfield 2019 *Duke law & technology review* 200.

1107 Waltz and Firth-Butterfield 2019 *Duke law & technology review* 184

1108 Asimov’s three laws are: “1. A robot may not injure a human being or, through inaction, allow a human being to come to harm. 2. A robot must obey the orders given to it by human beings, except where such orders would conflict with the first law. 3. A robot must protect its own existence as long as such protection does not conflict with the first or second laws”, See Asimov *I, Robot* “Runaround”.

1109 Waltz and Firth-Butterfield 2019 *Duke law & technology review* 200.

1110 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

could be unethical in a given situation. Many ethical scholars follow different principles of thought to address a particular ethical concern so that balanced decisions are taken, as opposed to merely following a specific principle of thought. Moreover, it may not be possible for AI systems to be designed to merely follow one principle of thought.

For these reasons, it may be best to consider the technical casuistic approach for incorporating ethical principles into the structural design of AI systems, by relying on specifically programmed decision-making structures.

Testing and validation

As a result of the non-predictive and content-driven landscape of AI systems, conventional testing methods for these systems may not be sufficient, as difficulties with the models applied by the system may only become exposed when an algorithm is used for appropriately accurate data.

Therefore, to confirm and authenticate the administering of data, the primary design application should be carefully scrutinised throughout the training and deployment of the system to ensure its reliability, and functionality within logical and foreseeable limits, and, that conclusions are reached in a manner permitting authentication of the causal method.¹¹¹¹

Evaluating and authentication should be effected at the earliest stage possible and throughout the AI system's lifecycle, so that it can be established whether the system is performing as expected, particularly after implementation. It must incorporate all elements of an AI system, including data, training patterns and protocols, settings and the performance of the system as a whole.¹¹¹²

Service instructions should be demarcated for AI systems, which could comprise of procedures to assess the testing and training of systems, inclusive of conventional software assessment of operability, behaviour, serviceability, reliability, safety and

1111 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

1112 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

maintainability.¹¹¹³ Evaluation procedures must be made available to all parties concerned, from the producers and users to lawmakers, and which must also incorporate procedures to solve inconsistent evaluation outcomes.¹¹¹⁴

Mitigation strategies

Because designers are not able to foresee every conceivable functioning condition and possible failure of AI systems, methods to counter the possibility and extent of damage must be implemented. Designers should consider various stringent laws (i.e specific ethical values) that should never be violated, and can define standards and metrics for assessment, which would enable the discovery and alleviation of failures.¹¹¹⁵

These metrics for assessment would encompass technical-based elements, for instance traceability and verification; user-based elements such as dependability and consistency, understandability, and receptiveness to end-user reaction; and community-based elements such as confidence and trust.¹¹¹⁶ A methodical risk assessment and oversight method would be beneficial. This method attempts to foresee possible areas of failure (e.g., privacy or fairness abuses) and, where feasible, creates certain ways to lessen or eliminate the impact of such failures.¹¹¹⁷

While the technical approaches, described above, may be a prospective method to deal with ethical concerns in specific instances, they may not be adequate to ensure that AI systems have due regard to ethical considerations for their decision-making process. It will still remain the AI system producers' decision to select the design

1113 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

1114 IEEE "Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent system, version 2" https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf (Date of use: 21 August 2020).

1115 IEEE "Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent system, version 2" https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf (Date of use: 21 August 2020).

1116 IEEE "Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent system, version 2" https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf (Date of use: 21 August 2020).

1117 Waltz and Firth-Butterfield 2019 *Duke law & technology review* 201.

philosophy they desire for AI structures, which may only be aligned to what is beneficial for them, financially or otherwise.

It is therefore necessary to consider a variety of non-technical methods, to establish a shared appreciation of the ethical principles and to foster an ethically aligned AI system design.

4.5.2.5. Non-technical Methods for effective application and enforcement of ethical principles

In addition to the technical ethics-by-design method, regulatory frameworks will need to employ an intelligent and articulate blend of “soft” and “hard” instruments to tackle the ethics of AI.¹¹¹⁸ Specific regulatory instruments can steer the reasoning for AI systems’ lifecycle, changing as society develops. Many of the ethical principles, are for the most part already echoed in current regulatory instruments, which are legally binding – such as the POPIA or the GDPR,¹¹¹⁹ which protects privacy and information security, or the Consumers Protection Regulations in both South Africa and the United Kingdom.¹¹²⁰

In the event that voluntary ethical rules may not be enough to deal with the ethical concerns ensuing from AI, tougher intervention will be required from regulators to promote the execution of these rules. A code of conduct for ethical and professional standards is essential for those engaged in the design and application of AI, together with a concrete, legislative framework calling for both transparency and accountability that recognises human participation when it comes to AI’s substructures.¹¹²¹

In support of this sentiment, the EC’s President, Ursula von der Leyen, declared her intention to offer legislative suggestions such as a new Digital Services Act on

1118 Mialhe <https://cpr.unu.edu/ai-global-governance-why-we-need-an-intergovernmental-panel-for-artificial-intelligence.html> (Date of use: 29 August 2020).

1119 GDPR (EU) 2016/679.

1120 South African CPA 68 of 2008 and British CPA of 1987.

1121 Doteveryone <http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/artificial-intelligence-committee/artificial-intelligence/written/69653.html> (Date of use: 22 August 2020).

accountability and security guidelines for digital platforms, towards a harmonised European initiative on AI's ethical impacts.¹¹²² Ultimately, there is a need for promoting the implementation of ethical guidelines and lawfully binding instruments, to impose a collective set of rules on requirements for accountability and rights-based impact assessments for the entire lifecycle of AI.¹¹²³ This section describes various non-technical methods which could have a significant role in accomplishing and sustaining ethics for AI.¹¹²⁴

Legislation

Legislation is regarded as the conventional regulatory method to execute standards that embody ethics – for example, privacy rights, liberty, equality, and the obligation not to harm other persons. The benefit of legislation, is that it is an instrument for established mandatory and enforceable rules and democratically adopted under a transparent system. In certain circumstances, legislation as a minimum offers a degree of legal certainty and social recognition.

As mentioned above, legislation that supports ethical principles, for instance the right of privacy,¹¹²⁵ dignity, equality and freedom,¹¹²⁶ already exists today, inclusive of consumer legislation for product safety¹¹²⁷ and legal accountability frameworks¹¹²⁸ to make sure that AI remains accountable and provides adequate redress when it causes harm.¹¹²⁹

1122 Von der Leyen U “Political Guidelines for the Next European Commission 2019-2024” <https://www.europarl.europa.eu/resources/library/media/20190716RES57231/20190716RES57231.pdf> (Date of use: 16 August 2020).

1123 The EU Ethical Guidelines sets outs a non-exhaustive Trustworthy AI assessment list to operationalise trustworthy AI, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

1124 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

1125 In South Africa the Protection of Personal Information Act 4 of 2013 and section 14 the Bill of Rights under the Constitution protects privacy.

1126 Section 7,9,10, and 12 of the Bill of Rights in the Constitution protect these rights.

1127 South African CPA 68 of 2008.

1128 South African law of delict.

1129 The Constitution of South Africa is considered as one of the more progressive constitutions worldwide and is the highest law of the country that protects basic rights on which ethical values are grounded.

However, even in democracies such as ours, legislation may have some shortcomings, and new legislation may be required to address certain grey areas. To the extent that existing legislation may have to be reviewed, modified or newly promulgated, both as a safety measure and as an enabler, this can be introduced by way of policy and investment recommendations developed through stakeholder involvement and engagement.¹¹³⁰

Stakeholder participation

This encompasses all stakeholders, e.g., those involved in designing, manufacturing, or implementing the AI systems (the producers), professional users of the system (such as companies or individuals who deploy the system in their business or practices), and end-users (those who participate in the buying or use of AI systems, but who are subjected or impacted by the outcomes or decisions made by the system), and society at large.

Various organisations have already engaged in stakeholder initiatives such as collaborative committees, or panels to examine and review the usage of AI systems and data analysis.¹¹³¹ These initiatives comprise of various participants, such as legal specialists, computer experts, technical and scientific experts, ethicians, employees and consumers from industry.¹¹³² Aggressively pursuing involvement and engagement on the application and implications of AI systems, encourages the assessment of outcomes and methods employed, which could be very beneficial in complicated scenarios.

It promotes objectivity and acceptance of different viewpoints, requirements and goals that fosters and establishes an ethical structure with due regard to regulation. Communication platforms should thus be implemented with stakeholders in the

1130 Saxenian *et al* Artificial Intelligence and Life in 2030 49.

1131 IT-Online “Stakeholders commit to responsible AI” <https://it-online.co.za/2020/08/06/stakeholders-commit-to-responsible-ai/> (Date of use: 26 August 2020)

1132 Stakeholders’ participation in an online conference of the “AI Dialogue South Africa” resulted in the execution of an expression of interest (EOI) that promotes accountable AI, IT-Online “Stakeholders commit to responsible AI” <https://it-online.co.za/2020/08/06/stakeholders-commit-to-responsible-ai/> (Date of use: 26 August 2020). Also see <https://www.partnershiponai.org/#> (Date of use: 6 June 2020).

technology industry, exchanging best practices, deliberating, and informing on developing ethical issues.

Codes of conduct and governance structures

Relevant industries can accept the proposed ethical rules, and acclimate their corporate charters of accountability, key performance assessments (“KPA’s”), and codes of good conduct or policies, to strive in the direction of ethical AI. Organisations that deal or trade with AI systems can record their objectives and endorse them with principles of some important values, such as basic human rights, accountability, transparency and safety.¹¹³³

Organisations should implement governance structures, both inside and outside, promoting responsibility for ethical aspects of decisions-making and recommendations connected with AI systems. It can be achieved by the nomination of an ethics officer that is certified, and oversees ethical aspects concerning AI systems, or by establishing an independent ethics advisory committee or panel.¹¹³⁴

Amongst the possible functions of such an officer, committee, or panel, is to observe and provide guidance. Such structures can augment, but cannot substitute legal compliance, e.g., the requirement to employ an information officer as provided for in the POPIA. A proper governance framework for AI, calls for the correct governing instrument to be selected to address every ethical concern. Legislators must have due regard to the varied landscape of ethical issues and apply a rated governance framework for these issues in AI, to establish appropriate structures and methods for regulation.¹¹³⁵ Given the range of ethical principles, it should be accepted that a

1133 Many major technology firms have already developed their own ethical guidelines or policies. See <https://www.blog.google/technology/ai/ai-principles/> (Date of use: 15 August 2020); <https://www.microsoft.com/en-us/research/theme/fate/> (Date of use: 15 August 2020).

1134 Microsoft has its own AI advisory board overseeing AI Ethics, see <https://www.microsoft.com/en-us/research/theme/fate/> (Date of use: 15 August 2020).

1135 Waltz and Firth-Butterfield 2019 *Duke law & technology review* 226.

“one size fits all” resolution is not feasible in respect of the different environments in which the AI systems are deployed.

Standardisation

For AI systems, industry standards can be introduced that employ technological methods which support ethical acceptable behaviour by these systems.¹¹³⁶ Technology standards, such as design, industrial and commercial practices, could be effective as a quality control mechanism for AI stakeholders, and advancing the ability to acknowledge and promote ethical behaviour through their procuring choices.

Technology design standards such as transparency, can deal with accountability concerns by ensuring that AI systems log their outcomes and recommendations which they rely on.¹¹³⁷ In addition, to address controllability of AI’s autonomous nature, the introduction of “kill-switch” technology could be considered,¹¹³⁸ or, adopting an identity tag standard where only systems that bear an identity tag can be authorised for release – to establish a pertinent pathway for liability.¹¹³⁹

The IEEE’s “Ethically Aligned Design” principles, offer an appropriate example of industry guidelines intended to achieve secure and ethical AI approaches.¹¹⁴⁰ Outside of traditional regulation, there are co-regulatory structures, such as certification schemes and professional rules. Current examples are ISO Standards series,¹¹⁴¹ reinforcing technical guidelines by way of referencing that a system, for example, prescribes to reliable technical procedures. Healthcare professionals are

1136 Waltz and Firth-Butterfield 2019 *Duke law & technology review* 215

1137 Refer to the IEEE P7001 Transparency of Autonomous Systems, available at <https://standards.ieee.org/project/7001.html> (Date of use: 27 August 2020).

1138 Refer to the IEEE P7009 Standard for Fail-Safe Design of Autonomous and Semi-Autonomous Systems, <https://standards.ieee.org/project/7009.html> (Date of use: 27 August 2020).

1139 Waltz and Firth-Butterfield 2019 *Duke law & technology review* 196, 216.

1140 IEEE “Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent system, version 2” https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf (Date of use: 21 August 2020).

1141 See ISO 9000 and 9001 standards for quality management systems' available at <https://www.iso.org/standards.html> (Date of use: 29 July 2020).

also required to adhere to a professional code of ethics¹¹⁴² when it comes to issues of patient healthcare, such as patient consent and patient privacy that is impacted when they use AI systems as part of their services.

Certification

The very notion of AI implies a substitution for human skill and expertise. Not many people will in a position to entirely appreciate how AI systems functions or its impact. The producers of the systems must be able to assure the wider public that they are reliable and safe.¹¹⁴³

In many contexts, such as healthcare, professionals must possess some certification or permit before they can perform a task that requires specific expertise. Similarly, when it comes to AI, certifications would invoke standards created for diverse AI systems, with due regard to the trade and social norms of various settings, such as healthcare. Certification should, however, not become a substitute for accountability and it should instead be supplemented by accountability structures – involving disclaimers, assessment and remedial processes.¹¹⁴⁴

Thus, akin to technology guidelines, conformity with AI principles of ethics could be accomplished by way of implementing licensing or accreditation structures.¹¹⁴⁵ However, law and policy will have to contend with the method of establishing competency for an AI system.¹¹⁴⁶ For instance, if a company designs a medical program that can autonomously remove an appendix, it is uncertain in this scenario who in law would be expected to pass the medical assessments to obtain the certification.

1142 HPCSA https://www.hpcsa.co.za/Uploads/Professional_Practice/Ethics_Booklet.pdf (Date of use: 19 May 2020).

1143 As advocated by e.g., IEEE “Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent system, version 2” https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf (Date of use: 21 August 2020).

1144 “European Commission’s Ethics guidelines for trustworthy AI” <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

1145 Waltz and Firth-Butterfield 2019 *Duke law & technology review* 216.

1146 Saxenian *et al* Artificial Intelligence and Life in 2030 47.

Certification should also encompass an ethics-based certification framework for accountable organisations (e.g., healthcare providers, government, businesses), akin to accreditations in biomedical ethics imposed as a professional prerequisite (e.g., a Hippocratic Oath), that must incorporate practical and situation training concerning principles of patient consent, ethics accreditation assessment, and even a notarised declaration of acceptance to honour principles of ethics.¹¹⁴⁷

4.5.2.6. Key guidance for implementing ethics principles in healthcare

The preceding discussions examined the basic rights, related ethical principles and listed five key requirements that can be considered to realise both ethical and trustworthy AI. An assessment list for trustworthy AI was proposed that may assist with implementation of the five key requirements. It was proposed that non-technical methods can assist with the implementation of AI systems in an ethical manner. The below is a summary of the key guidance derived from the discussions, when applying the identified ethical principles for AI systems:¹¹⁴⁸

- ✓ The AI system's full lifecycle must adhere to the five key requirements of the ethical principles: (1) privacy and data governance and accountability; (2) technical robustness; (3) fairness and societal well-being; (4) human agency and oversight; and (5) transparency.
- ✓ For transparency, users must be notified when they are engaging with or subjected to an AI interface. Information regarding the AI system's proficiencies and shortcomings, must be conveyed to all stakeholders in a simple and understandable way, allowing for pragmatic expectations. Accelerate AI's explainability, traceability, and validation, especially in crucial conditions and settings such as healthcare.
- ✓ Allow for stakeholder participation for the entire duration of the AI system's lifecycle. Encourage training to ensure that stakeholders develop and enhance their knowledge and awareness of ethical AI.

1147 IEEE "Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent system, version 2" https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf (Date of use: 21 August 2020).

1148 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

- ✓ Be aware of the possible conflicts between the various ethical principles and their associated requirements that may require trade-offs. One should constantly verify, assess, record, and report these compromises as well as their resolutions.
- ✓ Consider assessment methods for the operationalisation of the ethical-principles requirements (e.g., refer to the assessment list for AI).
- ✓ Consider the technical methods that can be introduced into, and during the entire lifespan of an AI system.
- ✓ Consider non-technical methods (regulation, codes of conduct and governance structures, standardisation and certification) for effective application and enforcement of ethical principles.

4.5.3. Conclusion

The growing acceptance and applications of AI technologies, will have ground-breaking implications for society. Notwithstanding the various advantages and benefits it offers in healthcare, AI systems encompass considerable risks and ethical concerns that need to be managed on a regulatory basis. Minimising these risks will emphasise the respective benefits as it ensures trust, while simultaneously safeguarding the ethical values outlined by basic constitutional rights, thus upholding a “human-centric” society.

As AI systems are employed more frequently by the healthcare industry, closely linked to patients, it becomes increasingly crucial to establish which of the human actors in the supply chain ought to be held accountable for ensuing injuries or losses resulting from the use of AI systems. This significant is heightened, as a failure in autonomous systems could have proliferating consequences. The issue on how to address liability and accountability when AI causes harm which must be answered with due regard to existing legal regimes. AI has not been accorded legal personhood, as it has not progressed to the stage where it would be considered acceptable under the law to ordinarily assign AI systems rights and responsibilities

(innate to the understanding of personhood, as legally defined in the present day).¹¹⁴⁹

The responsibility thus shifts to human actors employed in the construction, advancement, deployment and application of the system. Legal accountability is generally not a given in the law of delict of South Africa (nor the jurisdictions reviewed in this Chapter), if autonomous actions or decisions ensue in damages (intervening causation), unless there are strict liability regimes such as product liability law in South Africa. Liability regimes solely based on fault may, consequently, subject injured parties to insufficient protection under the law, where AI caused harm persuading future regulators to consider ways to implement future AI systems so that the best possible legal protections are available.¹¹⁵⁰

Technically, there are several already existing options to choose from – amongst producers and service providers. Thus, the combination of negligent (fault) based, and strict product (no-fault) based regimes for liability, are suitable for the possible harm resulting from AI systems of today.

The liability question is not difficult technically, it is more of an ethical and moral dilemma.¹¹⁵¹ The notions relating to “responsibility”, “accountability” and “liability”, therefore relate to a few of the central issues relating to ethics which require extensive dialogue with regard to emerging AI technologies. A critical assessment of the notion of autonomy is essential. Present-day AI systems are still designed by human actors, and operate within the parameters of the particular human-imposed coding and training, suggesting that they execute tasks set by human actors, and communicate the objectives of humans.

-
- 1149 IEEE “Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent system, version 2” https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf (Date of use: 21 August 2020).
- 1150 IEEE “Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent system, version 2” https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf (Date of use: 21 August 2020).
- 1151 IEEE “Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent system, version 2” https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf (Date of use: 21 August 2020).

Consequently, it may not be appropriate to regard an AI decision as a legitimately autonomous or independent decision, so as to absolve the human actor from liability who designed, manufactured or used the AI technology. The fundamental ethical consideration, is if human actors ought to be accountable for harms initiated by AI systems (where, after applying the training data, the system has made an independent decision which ensues in damages to person or property).

It may be argued that persons involved in any stage of the AI system's development (to meet owner expectations), is accountable. However, this issue will become progressively more critical to address as AI systems decisions grow to be more autonomous. This may however, change, taking into account the possibilities of general/strong AI in the future.¹¹⁵² As AI technology advances, the existing legal regime on liability may require further adaption, where separate legal subjectivity will be given to AI systems similar to a company, with rights and responsibilities of its own.

From the long-term perspective, this is the most reasonable solution. New laws can be considered, introducing the concept of an "e-person", which would hold an AI system directly accountable.¹¹⁵³ In this regard, a suitable funding or insurance model may be imperative to warrant that AI systems have been adequately capitalised, and is not capable of being misused as a means to avoid accountability.¹¹⁵⁴

In consideration of AI's ground-breaking effect, it is imperative to create awareness of the ethical considerations that arise from the application of AI platforms, not merely from the regulators side, but also in respect of all AI stakeholders involved in any stage of AI's lifecycle, and that are affected by it. Any possible regulation of ethics ought to consider the multiplicity of individuals' values.

Projects such as those of The IEEE Global Initiative, and the EU guidelines produce by the High-Level Expert Group on Artificial Intelligence, are amongst the first feasible and practical global attempts towards the consideration of an ethics model

1152 Defined as "an intelligent machine that has the capacity to understand or learn any intellectual task that a human being can."

1153 Waltz and Firth-Butterfield 2019 *Duke law & technology review* 197.

1154 Waltz and Firth-Butterfield 2019 *Duke law & technology review* 197.

concerning AI. In considering the proposed model for South Africa's healthcare sector, the research considered the core principles of the EU guidelines, which focus on a "human-centric" method for AI – that remains mindful of fundamental rights.

This can be regarded as particularly fitting for the cultural and social framework of South Africa, whose core ethical values such as social justice, sustainability, democracy, equality, and privacy are enshrined in the Constitution, intended to protect the most vulnerable in society. These ethics guidelines are non-binding, and regulatory oversight mechanisms are required to support their implementation.

Both technical methods (included all the lifecycle stages of AI systems), and non-technical methods (existing binding legislation, codes of conduct and governance structures, standardisation and certification), can support enforcement and conformity. In the absence of such proposed frameworks, there is not much motivation to obey these principles of ethics.

With regard to emerging and complex AI technologies, finding the right solution for regulation that can offer benefits and alleviate the threats of AI, poses a certain difficulty because the technology transforms at a swift pace – rendering it problematic to predict the future of the technology in the succeeding five years. Furthermore, innovation lifecycles are usually exceptionally short in the technology sector, inclusive of AI, making it difficult for regulation to keep up in the sector.

A collaborative stakeholder initiative would be the most useful approach to make sure that AI will benefit and address the requirements of society, in allowing producers, operators and regulators to discuss and collaborate from the onset. The regulatory approach proposed in this thesis, should thus be seen as a working framework that will evolve in the realm of the universal race on AI. There are various forms of ethical standards and codes for the healthcare industry of South Africa, which comprise of codes for ethical and good practice by healthcare professional bodies applicable to their members,¹¹⁵⁵ health and safety rules produced by industry

¹¹⁵⁵ HPCSA https://www.hpcsa.co.za/Uploads/Professional_Practice/Ethics_Booklet.pdf (Date of use: 19 May 2020).

or government agencies,¹¹⁵⁶ or ethical codes for research and directives issued by organisations that are financing or undertaking AI research and development.¹¹⁵⁷

As demonstrated in Chapter three, given the tempestuous context within which AI is evolving in healthcare, for these codes of ethics to be effective in tackling the ethical and legal challenges confronting society, they must be regulated and enforceable by law and would benefit from the assistance and support of recognised and established authorities.

In the next Chapter, I summarise the discussions of the previous chapters and offer conclusions and recommendations on the research.

¹¹⁵⁶ National Healthcare Act 61 of 2003 of the Department of Health establishes a framework towards a standardised health system in South Africa.

¹¹⁵⁷ National Health Research Committee created in terms of the National Health Act 61 of 2003 provides guidance on ethical matters concerning health and establishes guidelines on research relating to humans and animals.

5. CHAPTER FIVE: SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

5.1. Summary

The previous Chapters demonstrate that, AI is the capability of systems that are designed to imitate human performance in an intellectual way, which it has begun to increasingly transform healthcare delivery in progressive markets and shows promise to promote life-changing advancements for vulnerable societies in universal healthcare.

William Kissick's explanation of the "iron triangle" in healthcare, encompassing "access, quality and cost containment" is succinct: "all societies confront the equal tensions among access to health services, quality of healthcare, and cost containment" and "tradeoffs are inevitable regardless of the size of the triangle."¹¹⁵⁸

The above offers a superlative rationale as to why healthcare remains an arduous challenge. From allowing rural-health workers to serve vulnerable patients more resourcefully in inaccessible areas, to aiding governments in LMIC's to manage pandemics through early detection and intervention, there is mounting acknowledgement of the noteworthy capabilities of AI systems, to disrupt the fundamental trade-offs in health "access, quality, and cost" depicted by Kissick.

As with most LMIC's, South Africa's health system faces impediments, including calamitous shortages of healthcare professionals, insufficient medical equipment and infrastructure, and a lack of other capabilities that require innovative methods to tackle these challenges. AI technologies promise not only to augment resources that are available, and to aid with overcoming other resource barriers, but can also considerably advance healthcare availability and outcomes for patients in low-income situations in unimagined ways.

Although this enthusiasm and hype around the deliverance of AI for healthcare has been welcomed, detailed assessment as to how to effectively position and scale such innovative technologies in healthcare systems, remains wanting. This is largely

1158 Kissick W *Medicine's dilemmas: Infinite needs versus finite resources* 1st ed (Yale University Press 1994) 2.

due to the fact that, technology issues are often reduced to computer life sciences involving binaries and codes.

On the contrary, the deployment of AI technology is complex. The manner in which health data are collected, stored and used, distinct from conventional healthcare settings; the challenge of defining what a “medical device” is in a world overtaken by user-driven technologies; and the multi-faceted affiliation between healthcare stakeholders and the emerging technologies which they are compelled to navigate, are some of the factors that contribute to the complexities.

It is also a complicated task to deploy disruptive technological improvements from HICs, and to adapt them to the distinctive requirements of, and to offer benefits to communities in low-income settings. The issue of a regulation for AI becomes increasingly significant to address these challenges, particularly as healthcare integrates not only user-interfacing tools (such as medical applications), but also futuristic technologies in the form of robotics and AI algorithms (that enables hardware to perform independently, and software to make critical decisions and recommendations for patient care).

The notion of identifying broader regulatory principles and how they are to be applied, has provoked numerous themes in medical legal literature, and policy. Scholars have considered the appropriate regulatory method to be adopted to new healthcare technologies and have studied the challenges of offsetting innovation and regulation, with due regard to the negative social and economic externalities of emerging AI technologies.

This thesis endeavours to provide answers to some of the extremely fundamental questions regarding the issue of regulation of AI for healthcare: whether AI should indeed be regulated, and if so, why. To answer these questions, Chapter one of the thesis began by adopting the commonly used scientific definition of AI,¹¹⁵⁹ and considered what its benefits and challenges are for the healthcare sector.

1159 The scientific definition as referred to in Chapter one of this thesis.

Chapter two conducts an overview of AI use cases in healthcare that strengthen healthcare systems across LMICs and HICs, demonstrating that these real-world examples of use cases appeared to have greatest potential for application in LMICs such as South Africa where investment in AI specific technologies or platforms are needed. The Chapter examines present issues that are encountered when scaling use cases for AI in LMICs healthcare which included: data availability; data privacy; absence of trust; fairness; accountability; and inadequate regulatory guidelines for evaluating the ethical implications of AI technologies.

Chapter three scrutinises regulation and what it means. It is concluded that, a control system must be developed based on a top-down approach, and the state must intervene to regulate AI as there are compelling reasons to do so. In assessing the arguments in support of and against regulating AI, it is established that a regulation of the new and evolving technology is not only unavoidable, but that AI stands to benefit from decisive and well-timed regulation. Moreover, a regulation should not be considered as a “necessary evil”. It can present various advantages to the AI sector itself by promoting the development of trustworthy and safe AI systems, and to society in general by improving access to affordable and quality healthcare.

Chapter four navigates from the viewpoint that a regulation should fulfil an objective, and therefore, the AI as a field only stands to be regulated if it is able to accomplish its goals of offering an advantage to the healthcare industry, and to the public, and what it takes for it to be successful.

A good regulation through government should embody a convergence of both horizontal and vertical regulatory approaches, where the government is involved in issuing rules and guidelines enforceable against other actors, through coordinating its efforts with the cooperation of those actors. The research establishes in what way a “good” regulation may be established through effective and appropriate governance structures, and oversight in the design, deployment and operation of sophisticated socio-technical technologies such as AI.

Thus, for South Africa the application of existing negligent liability regimes found under law of delict and product liability frameworks contained in the CPA, were considered to assess whether this could be successfully applied to ensure that AI

remains accountable and offers redress in instances where it fails and generates an unsafe outcome for patients.¹¹⁶⁰

The existing frameworks are well-placed to acclimate and tackle concerns on AI liability. The research also analyses the tort liability regimes of foreign jurisdictions in the UK and US concluded that the position is not fundamentally different when compared with South Africa, but, that unlike the US, South Africa appeared to be lagging behind in terms of developing a proper framework dealing with approval and regulation of autonomous and adaptive AI technologies (that must meet the definition for a “medical device” to deliver technology that is effective, but also ensures the safety of end-users – being patients and healthcare workers).

This Chapter also serves to close a lacuna in South Africa’s existing ethico-legal frameworks towards addressing the ethical challenges, by analysing international and national ethical guidelines and policies as a guidance for AI in healthcare, and to ascertain which norms and standards from these documents could be useful when developing ethico-legal safeguards for South Africa.

The ongoing initiatives of several countries were considered. The IEEE Global Initiative and the EU guidelines established by the High-Level Expert Group on AI is identified as amongst the first concrete and practical global initiatives towards the consideration of an ethics framework into AI. Based on the guidelines adopted through the above-mentioned initiatives, requirements for medical ethical principles premised on a human-centric¹¹⁶¹ point of view, and assessment methods in order to achieve trustworthy AI, were considered for the construction, deployment and application of AI for the healthcare industry in South Africa.

Chapter five concludes by summarizing and concluding on the argument made in the preceding sections, and making further recommendations. The conclusion is that, whilst AI offers many benefits towards the advancement of healthcare, it is also necessary to foster social and ethical norms in such an advanced digital age. It

1160 This thesis only dealt with unintentional failure giving rise to liability for negligence in delict. This is where AI systems fail by their own accord due to a design or development error.

1161 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

further ensues from the responsibility of government to safeguard human rights. This, entails a duty to implement legislation and regulatory frameworks that will ensure that both historic and future accountability are duly allocated to deal with the grave risks, injuries and wrongdoing on account of the use of unconventional technologies. The conclusion is that accountability for AI is achievable without having to stifle innovation.

5.2. *Conclusions and recommendations*

Despite its numerous benefits, AI systems encompass many risks, including adverse societal concerns that need to be managed. While some of the risks may be harmless, there will be instances of serious damage or injury caused to an individual or property, which could threaten to unsettle the institutions upon which society's moral structure exists.

The main initiator for a global conversation on the topic of AI risks, will be an AI-triggered event that causes extensive losses or harms, as has happened with the COVID-19 pandemic. Sensitising one to possible risks should not be misconstrued as being an anti-innovative endeavour. Rather, it is essential to have due regard to risks so as to ensure that novel technologies like AI, are developed and operate based on an appropriate AI governance regime that secures the trust of individual users, and serves the public interest with emphasis on the safeguarding of a human-centric culture.¹¹⁶²

In alignment with the broader legal and ethical principles for AI technologies that have been earlier identified in this thesis, the following recommendations serve to further develop a visionary inclusive AI governance regime, to fast-track the appropriate development and application of AI, to advance healthcare within LMIC settings such as South Africa.

1162 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

5.2.1. Assessment methods

Risk assessments and prevention mechanisms during the design and deployment stage of AI technologies, serve not only to protect people and their basic rights but promotes widespread approval of such technologies. Risk prevention means mechanisms should be employed in assessing the risks associated with AI. A good starting point could be the people, processes and data. The total risk being evaluated will be a function of the talent of the people building the AI systems, the processes embedded within their workflows, and the quantity and quality of the input data.

People

AI systems are the product of work by talented domain experts, knowledge engineers and programmers – often with PhD's in complex research areas. There are many different factors that could go into assessing the development team. Factors such as whether a PhD is really better than an MSc may become relevant, or the quality of the institutions that awarded those certifications may be considered. Perhaps even more relevant is the question of experience over certificates. For instance, the statistics could indicate that a team made up of engineers that previously worked at Google, develops AI that generates fewer errors than a team made up of engineers from another start-up technology company.

Process

Development teams work within environments that have existing workflows, checks and sign-off processes. Domain experts develop algorithms, knowledge engineers help interpret large or varied data sets, and software engineers build the algorithm into working software. If a company invests in its staff, and embeds layers of additional checks in its workflow process before it goes to production, it could be at a lower risk than an identical company that does not. The way these experts design the AI systems versus how it is constructed in production by the manufacturers, is another source of risk that could be mitigated by its own effective workflow process.

Data

ML models are merely as effective as the training data entered into them that enables them to learn and adapt, and therefore a noteworthy difficulty to promoting the application of AI technologies in LMICs such as South Africa, concerns the quality and volume of health data (because of a poor digital health infrastructure in these markets).

Given the existing endeavours by global healthcare researchers to capture and populate digital healthcare data, it is recommended that participants augment management of their contributions to generate an adaptable and open science platform for sharing of data sets germane to AI in universal healthcare, and to also then make it publicly available among interested AI innovators, researchers and data contributors.

AI models require volumes of healthcare data sets to enable algorithms to learn, adapt and provide correct outputs representative of the relevant populations, and such wider healthcare data tend to be lacking in LMICs. Data sets can be aggregated in the open science platform, so that the entered data speaks to a range of contexts such as gender, ethnicity, class, geographical location, etc., which will aid in developing accurate predictions or diagnoses for patients in LMICs.

An absence or limitation of healthcare data does not only impede the precision in outcomes of AI technologies, but it also generates a bias emanating within these technologies, and could later lead to morally questionable or unlawful decisions being made concerning patients.

Oversight structures

Given the complex technological landscape in which AI is developing, specialised AI courts could also be initiated, comprising of judges who are knowledgeable and conversant of the field with a reduced likelihood of being deceived by theoretical propositions of expert witnesses. If incapacity is established, new institutional structures should be formed to offer redress to the public.

It would also be important to assess the ability of prevailing institutional structures, such as the judiciary, to provide reparation for oversights or losses exacted by AI. The effective discovery and interrogative procedures of civil litigation, are powerful

mechanisms for securing and divulging pertinent information associated with the design aspects and safety features of AI technology causing harm, and securing such comprehensive information is effective when having to prove causal aspects. National legal systems must be assessed to ensure that it is capacitated so that liability for injury or damage caused by AI technologies can be correctly imputed, and imposing legal reform where possible gaps have been established. As AI progresses from narrow/weak to general/strong AI, the implementation of human-designed algorithms will no longer be required.

An “AI ombudsperson” could be appointed to investigate complaints against institutions, a mechanism for lodging a complaint on AI matters similar to the process for a Promotion of Access to Information Act (PAIA) application, or the establishment of an insurance regime for AI as a mandatory requirement for specific AI offerings in the South African market typically covering losses resulting from unintentional errors akin to medical malpractice insurance.

A market-based approach could be considered that calls upon those who are part of the supply value chain of AI systems to procure public liability insurance from appropriate brokers, thus leaving the open market to manage the risk of injury or damage generated by AI systems or establishing a national insurance scheme, funded by the technology industry.

There could also be an assessment of which responsibilities and decision-making functions should or should not be assigned to AI systems, with due regard to prevailing legislation and engaging all participants (government, industry and society) through participatory schemes, to position same with societal norms and public opinion.

The existing legal structures for liability would need to be assessed and adapted, to create a legal basis through the conferring of legal personhood on AI systems. This will allow “AI doctors” to integrate into a working environment as a non-human subject that is regarded to be a person under the law, and that would be capable of being sued in court. In such an instance, the burden of liability should attach to the AI system, and those involved in the supply value chain would not be subjected to lawsuits for negligence except if the AI system was inadequately capitalised due to

the absence of AI insurance or other financial mechanisms (or if the court decided to “pierce the AI veil”).

5.2.2. Development approaches

Developing of auditing functions for AI systems to detect and recognise undesirable outcomes can be considered. With the design and deployment of AI medical technologies or devices, the proposed auditing framework must be developed in cooperation with the SAHPRA (who as regulatory oversight), the healthcare industry and insurance sector as a coherent approach to mitigate severe risks in the healthcare sector. This would include developing a series of system of measurement for methodically detecting and assessing the extent and gravity of possible risks and challenges to human values and rights posed by potential AI systems.

Developing a dedicated “watchdog” bureau that will safeguard public interests through the technical review and observation of AI systems is also important. The researcher supports the adoption of new legislation proposed by Matthew Scherer,¹¹⁶³ the Artificial Intelligence Development Act (AIDA), to safeguard the interest of humans and to adopt safe-AI, by overseeing and preventing the development of AI that do not address these issues and by promoting beneficial AI that do exemplify such issues. AIDA could create an independent agency operated by AI experts, which have the experience in the field for endorsing and certifying any AI system as safe by imposing rules for pre-accreditation, but with limited powers to interfere in development.

AIDA would confer authority on the agency to implement a certification process for AI systems that are commercially traded in the open market to be accredited as safe. Thus, instead of assigning the powers to the SAHPRA to assess whether AI systems are safe, as in the context of medical devices, the Medicines Act¹¹⁶⁴ can be amended to give the AIDA agency the powers to make recommendations to the

1163 Scherer 2016 *Harvard journal of law and technology* 393.

1164 101 of 1965 as amended.

SAHPRA regarding the safety of AI systems as medical devices (for registration and to oversee ongoing reassessments of these systems).

The agency would be more suitably placed than the courts to evaluate the efficacy and reliability of AI systems, mainly as a result of the misplaced interest of the courts that would adopt a clinical approach by solely focusing on the technology that caused harm in a specific case instead of the collective public efficacy concerning the technology. This will hinder the development of AI, as companies may be deterred from an AI impacted industry in fear of being exposed to liability. AIDA can also impose legal duties on producers to demonstrate satisfactory testing and accreditation prior to the release of AI systems in the open market, and with every reconfiguration or modification of the system.

In conjunction with stakeholders an AI Corporate Governance Code of best practices can be developed to assume accountability for the ethical impact of AI systems within organisations, akin to the King Report on Corporate Governance. The AI Corporate Governance Code can consider some of the principles of King IV¹¹⁶⁵ that focus on ethical leadership, organisation behaviour in society, organisational ethics, responsible corporate citizenship, stakeholder inclusivity and integrated reporting that can be applied towards the design and application of AI. For instance, specialised training could be given for board of executives and the establishment of a mandatory internal ethics review committee with auditing functions to be implemented by enterprises who engage in the creation, deployment or application of AI systems with oversight on AI projects.

Developing an African focused observatory initiative for AI is also an approach to consider for the African context. The observatory would monitor and report advances in AI, particularly with regard to critical sector-based perspectives such as healthcare, offer a forum that would cultivate dialogue and harmonised approaches,

1165 King IV https://cdn.ymaws.com/www.iodsa.co.za/resource/collection/684B68A7-B768-465C-8214-E3A007F15A5A/loDSA_King_IV_Report_-_WebVersion.pdf (Date of use: 8 September 2020).

offer an open source for AI research in healthcare, and issue recommendations and guidelines for action.

5.2.3. Incentives

Over and above the implementation of targeted regulatory frameworks, governments can create financial incentives to encourage the design and deployment of AI systems complementary to specific regulatory objectives. Governments could also impose the need to address identified ethical requirements as a condition to the approval of research funding towards the advancement of AI.

AI's implementation in the African continent that are ethically established and favourable to the environment can be incentivised financially. This will include the public awareness of AI and introducing organised public consultation structures such as surveys and focus groups or simulated experiments of the ethical dilemmas presented by AI systems, to design ethical guidelines and rules related to AI.

5.2.4. Support Initiatives

It is worthwhile to support the adoption of a codes of conduct specific to AI through a professional accreditation scheme, with specific ethical standards of conduct. This would apply to appropriately qualified AI technical experts as a class of professionals trained in authenticating the development and programming of software codes and algorithms. These professionals would require an attendant certification for "ethical AI" so that they understand the virtues of ethical AI, and which they declare that they will bear responsibility with regard to the technologies they develop.

Supporting the development of an educational program on AI's ethico-legal, challenges is another factor to consider. This could include: training and educating employees in organisations involved with AI systems on the societal and ethico-legal implications of working with AI particularly for employees that are tasked with every aspect relating to the management of the systems, or an African-focused approach to include fundamental human-centred rights in the study of computer science and engineering curricula dealing with AI systems. Communities should

also be educated and adequately engaged with, to ensure that their perceptions around the use of AI is considered.

Support must comprise of capacitating LMIC governments as well as all other stakeholders within the healthcare environment, to drive the success of AI. Capacity building at a domestic level further reduces the possibility of bias when development of AI occurs in a HIC, and is then implemented in an LMIC. There should also be technical assistance for LMIC's to facilitate the acquisition and adoption of AI systems with due regard to their long-term concerns and resource constraints.

5.2.5. Considering the Africanisation of AI

Ethical frameworks on AI have already been developed by several organisations all over the world. However, the recommendation is to also consider a regulating instrument that does not only focus on the vocalisation of Western values, but one that should be enriched from an African perspective on deeper, theoretical levels based on South Africa's local, cultural and ethical heterogeneity, value systems and the demands of global fairness to deal with the positive and negative impacts of AI technologies.¹¹⁶⁶

Many researchers and academics have already made vast contributions towards the application of ICTs to improve African's lives.¹¹⁶⁷ However, the reality is that most academic pursuits have thus far accepted that the extent of the universe's insight and intelligence, is derived from Western countries, particularly in the field of AI where more of the technology is being developed and used in countries such as Europe and America. This is also evidenced in most areas of research and thinking, where less research has been undertaken on the Africanisation of software design and development approaches.¹¹⁶⁸

1166 The draft UNESCO recommendation of the ethics of Artificial Intelligence recognise the need to pay specific attention to social, ethical and cultural diversity, available at <https://en.unesco.org/artificial-intelligence/ethics> (Date of use: 9 July 2019).

1167 Kroeze JH *A Framework for the Africanisation of the information systems discipline* (College of Science, Engineering and Technology University of South Africa 2019) 38-39.

1168 Kroeze *A Framework* 39.

Since end-user technology has become universally ubiquitous, research is necessary to develop scholarship, capacity and leadership to permeate ICT's with African content, in order to promote an Africanised technological environment for the continent.¹¹⁶⁹ It is the intention that this thesis would create awareness for the need to explore new avenues of examination and research and open a new dialogue for approaches to the ethical application of AI that have for most part been advanced from a Western academic and scientific perspective, towards an African theological model.

There is a need to consider the African dogmatic perspective, which can advance the academic debate on the values and principles that would apply to the application of AI for the African notion. The question is what makes a particular notion African? As a result of the multiplicity of populations and ethnicities that encompass the population who inhabit South Africa, it becomes impossible to adopt a universal approach that could be termed “an African philosophy”.

If scholars are able to accept descriptive notions such as Western, or Oriental, associated with factors such as locality, ethnicity and philosophical disparities, then equally this should accommodate the acceptance of “African thought” in equivalent or similar fields of research.¹¹⁷⁰ In this regard, the African idea of personhood is often associated with the notion of “ubuntu”, and the participation of community as central elements of African theology.

The notion of ubuntu is loosely interpreted as “humanity”, “humanness”, or even “humaneness” and articulated into a general respect and kindness in relation to others.¹¹⁷¹ Considered as a societal philosophy, ubuntu embraces the ideologies of collaboration, cooperation and community, and has been an important influence towards the transformation of public service delivery after the end of apartheid era in South Africa, where the new government enacted the concept of ubuntu in its

1169 Kroeze *A Framework* 40.

1170 Makgoba MW “Patterns of African thought: A critical analysis” in Du Toit CW (ed) *Faith, science and African culture – African cosmology and Africa's contribution to science* (Research institute for Theology and Religion University of South Africa) 99-106.

1171 Bolden R “Ubuntu” in Coghlan D and Brydon-Miller M (eds) *Encyclopedia of action research* (London Sage Publications 2014) 1.

social welfare program.¹¹⁷² Whilst the concept of ubuntu is recognised as uniquely African, similar notions have been observed in other communities, which include the Chinese concept known as Jen, the Filipino concept referred to as Loob and the Russian notion known as Obschina.¹¹⁷³ Many of these philosophies therefore commonly recognise the core principles of humanity, dignity and respect – which signifies that when it comes to technological development, whether it be tangible or intangible, it should be subject to the ubuntu values of humanness.

In addition, for many communities in South Africa, ancestors and spirit rituals are key elements when it comes to the governance of the inherent order and social matters to restore balance, or to influence or change a situation (e.g., to seek healing for illness, or prosperity, request blessings or guidance).¹¹⁷⁴ Thus, no technology or associated laws that are developed should exploit or disregard the existing beliefs and practices of diverse communities in South Africa for individual enrichment. In practice, the concept of ubuntu focuses on the collective as opposed to privileging one over the other.¹¹⁷⁵

Whilst the exact implications of the African theological ethics for AI has not been fully developed at this stage, it is important that principles from national and international levels are applied within the South African context, for the foreseeable regulation of AI technologies in the country.¹¹⁷⁶ Both Africans and non-Africans must advocate for the ubuntu concept as it would aid AI scholars and scientists, practitioners and regulators to interact with greater social structures and more recognition of indigenous knowledge in laws and stakeholder practice.¹¹⁷⁷

A greater appreciation of the social and cultural framework and how it affects the idiosyncratic realities of diverse communities should therefore be an essential element of the AI knowledge creation and implementation process.¹¹⁷⁸ The notion of Africanisation can be applied, for example, by integrating African traditions to

1172 Bolden *Ubuntu* 1.

1173 Bolden *Ubuntu* 1.

1174 Mbiti JS *African religions & philosophy* 2nd ed (Heinemann Oxford 1990) 30-104.

1175 Bolden *Ubuntu* 3-4.

1176 Kroeze *A Framework* 40.

1177 Bolden *Ubuntu* 3-4.

1178 Bolden *Ubuntu* 4.

inform the theory and practice of information systems.¹¹⁷⁹ AI systems analysis and design could be enriched and adapted by integrating the principles and customs of conventional African decision-making processes.¹¹⁸⁰

The objective should be to strive towards building indigenous knowledge and ubuntu values into AI technology, and perfecting an ethical framework for Western technological advances in areas such as AI healthcare.¹¹⁸¹ The values of ubuntu can also be embraced by healthcare practitioners through capitalising on the efficiency offered by AI, and focusing more on the humane aspect of healthcare, which entails establishing trust and interacting with patients. Only then, will the fundamental values of empathy, kindness and trust in AI as a patient-centred model of healthcare be achieved amongst the diverse communities of South Africa.

In conclusion, digital technologies in the field of AI are still novel but are also rapidly evolving. The participants involved in the supply chain for AI, have an essential role to play in shaping the market towards supporting technologies that are properly and ethically implemented, and that are scaled in healthcare systems such as South Africa to address the unique challenges of low-income populations.

Those who deploy and stand to profit from AI in the provision of healthcare services, must be accountable for their adverse risks and consequences. With the recommendations offered here, a case is made for an inclusive and harmonised approach of imperative policy adoption and technological transformation, which is believed will underwrite the benefits, and mitigate risks resulting from AI deployed in the healthcare industry, for the benefit of all people, and more particularly for the most vulnerable populations in South Africa.

1179 Kroeze *A Framework* 45.

1180 Kroeze *A Framework* 45.

1181 Kroeze *A Framework* 43.

ANNEXURE 1 – ETHICAL CLEARANCE



UNISA CLAW ETHICS REVIEW COMMITTEE

Date 20191105

Dear AA Jogi

Reference: STE 134 of
2019

Applicant: AA Jogi

Decision: ETHICS APPROVAL

FROM 01 November 2019

TO 01 November 2022

Researcher: Ms Anisha Amarat Jogi

Supervisor: Prof M Labuschaigne

Artificial Intelligence And Healthcare in South Africa: Ethical and Legal Challenges

Qualification: LLD

Thank you for the application for research ethics clearance by the Unisa CLAW Ethics Review Committee for the above mentioned research. Ethics approval is granted for 3 years.

The CLAW Ethics Review Committee reviewed the Negligible risk application on 1

November 2019 in compliance with the Unisa Policy on Research Ethics and the Standard Operating Procedure on Research Ethics Risk Assessment. The decision was ratified by the committee.

The proposed research may now commence with the provisions that:

1. The researcher(s) will ensure that the research project adheres to the values and principles expressed in the UNISA Policy on Research Ethics.

University of South Africa

Preiler Street. Muckleneuk Ridge City of Tshwane

PO Box 392 UNtSA 0003
South Africa Telephone: +27 1 2 429 Facsimile:
+27 1 2 429 41 50 www.unisa.ac.za

Yours sincerely,



PROF T BUDHRAM PROF M BASDPR F M BASDEO

Chair of CLAW ERC

Executive Dean : CLAW

E-mail: budhrt@unisa .ac.za

E-mail: MBasdeo@unisa.ac.za

2. Any adverse circumstance arising in the undertaking of the research project that is relevant to the ethicality of the study should be communicated in writing to the CLAW Committee.
3. The researcher(s) will conduct the study according to the methods and procedures set out in the approved application.
4. Any changes that can affect the study-related risks for the research participants, particularly in terms of assurances made with regards to the protection of participants' privacy and the confidentiality of the data, should be reported to the Committee in writing, accompanied by a progress report.
5. The researcher will ensure that the research project adheres to any applicable national legislation, professional codes of conduct, institutional guidelines and scientific standards relevant to the specific field of study. Adherence to the following South African legislation is important, if applicable: Protection of Personal Information Act, no 4 of 2013; Children's act no 38 of 2005 and the National Health Act, no 61 of 2003.
6. Only de-identified research data may be used for secondary research purposes in future on condition that the research objectives are similar to those of the original research. Secondary use of identifiable human research data require additional ethics clearance.
7. No research activities may continue after the expiry date 1 November 2022. Submission of a completed research ethics progress report will constitute an application for renewal of Ethics Research Committee approval.

Note:

The reference number STF 134 of 2019 should be clearly indicated on all forms of communication with the intended research participants, as well as with the Committee.

Tel: (012) 433-9462

Tel: (012) 429-8603

URERC 25.04.17 - D

University of South Africa

City of Tshwane

Decision template (V2) - ApprovePreUer Street, Mucklenel.ik Ridge

PO Box 392 City Centre
South Africa Telephone. +27 1 2 429
+27 1 2 429 50 www.unisa.ac.za

BIBLIOGRAPHY

Abbreviations

1. AI Artificial Intelligence
2. CPA Consumer Protection Act
3. GDPR General Data Protection Regulation
4. HIC High Income Countries
5. HPCSA Health Professions Council of South Africa
6. LMIC Low and Middle Income Countries
7. NHS National Health Service
8. POPIA Protection of Personal Information
9. SAHPRA South African Health Products Regulatory Authority
10. WHO World Health Organisation

Journal Articles

1. Alheit 2006 *The comparative and international law journal of Southern Africa*

Alheit K “Delictual liability arising from the use of defective software: comparative notes on the positions of parties in English law and South African law” 2006 *The Comparative and international law journal of Southern Africa* 266-297

2. Asay CD “Artificial Stupidity” 2020 *William and Mary Law Review*

Asay CD “Artificial Stupidity” 2020 *William and Mary law review* 1187-1257

3. Arkfeld 2020 *Judges' Journal*

Arkfeld M “Litigating and judging artificial intelligence cases” 2020 *Judges' journal* 1-7

4. Bathae 2018 *Harvard journal of law and technology*

Bathae Y “The artificial intelligence black box and the failure of intent and causation” 2018 *Harvard journal of law & technology* 890-938

5. Beil et al 2019 *Intensive care medicine experimental*

Beil M *et al* "Ethical considerations about artificial intelligence for prognostication in intensive care" 2019 *Intensive care medicine experimental* 1-13

6. Beutel 1985 *The computer lawyer*

Beutel RA "Government regulation of diagnostic software: a threat to artificial intelligence software developers" 1985 *The computer lawyer* 22

7. Black 2002 *Australian journal of legal philosophy*

Black J "Critical reflections on regulation" 2002 *Australian journal of legal philosophy* 1-35

8. Bostrom 2002 *Futures*

Bostrom N "When machines outsmart humans" 2003 *Futures* 759-764

9. Brundage *et al* 2018 *Computer Science*

Brundage *et al* "The malicious use of artificial intelligence: forecasting, prevention, and mitigation" 2018 *Computer Science* 1-99

10. Buch 2018 *British journal of general practice*

Buch VH, Ahmed I and Maruthappu M "Artificial intelligence in medicine: current trends and future possibilities" 2018 *British journal of general practice* 143-144

11. Callaghan and O'Sullivan 2005 *Computer law & security report*

Callaghan D and O'Sullivan C "Liability for software errors" 2005 *Computer law & security report* 1-57

12. Calo 2017 *University of California*

Calo R “Artificial intelligence policy: a primer and roadmap” 2017 *University of California* 404-435

13. Calo 2015 *California law review*

Calo R “Robotics and the lessons of cyberlaw” 2015 *California law review* 513-563

14. Cath *et al* 2017 *Science and engineering ethics*

Cath C *et al* “Artificial intelligence and the ‘good society’: the US, EU and UK approach” 2018 *Science and engineering ethics* 505-528

15. Chung and Zink 2018 *Asia Pacific journal of health law & ethics*

Chung J and Zink A “Hey Watson – Can I sue you for malpractice? Examining the liability of artificial intelligence in medicine” 2018 *Asia Pacific journal of health law & ethics* 51-80

16. Chung and Kim 2016 *KSII transactions on internet and information systems*

Chung M and Kim J “The internet information and technology research directions based on the fourth industrial revolution” 2016 *KSII transactions on internet and information systems* 1311-1318

17. Compton and Jansen 1988 *Proceedings of the Australian joint artificial intelligence conference*

Compton P and Jansen B “Knowledge in context: a strategy for expert system maintenance” 1988 *Proceedings of the Australian joint artificial intelligence conference* 1-14

18. Davenport and Kalakota 2019 *Future healthcare journal*

Davenport T and Kalakota R “The potential for artificial intelligence in healthcare” 2019 *Future healthcare journal* 94-98

19. Erixon, Ferracane and Van Der Marel 2015 *European centre for international political economy*

Erixon F, Ferracane MF and Van der Marel E "The health of nations: a transatlantic trade and investment agenda for better healthcare" 2015 *European centre for international political economy* 1-30

20. Felländer-Tsai 2020 *Acta orthopaedica*

Felländer-Tsai L "AI ethics, accountability, and sustainability: revisiting the Hippocratic oath" 2020 *Acta orthopaedica* 1-2

21. Floridi *et al* 2018 *Minds and Machines*

Floridi L *et al* "AI4People - an ethical framework for a good AI society: opportunities, risks, principles, and recommendations" 2018 *Minds and machines* 590-706

22. Gerke *et al* 2020 *NPJ digital medicine*

Gerke S *et al* "The need for a system view to regulate artificial intelligence/machine learning-based software as medical device" 2020 *NPJ Digital medicine* 1-4

23. Gerstner 1993 *Santa Clara law review*

Gerstner ME "Liability issues with artificial intelligence software" 1993 *Santa Clara law review* 239-269

24. Gill 1995 *MIS quarterly*

Gill GT "Early expert systems: where are they now" 1995 *MIS quarterly* 51-81

25. Gillon 2015 *J Med Ethics*

Gillon R “Defending the four principles approach as a good basis for good medical practice and therefore for good medical ethics” 2015 *J med ethics* 111-116

26. Guo and Li 2018 *Health equity*

Guo J and Li B “The application of medical artificial intelligence technology in rural areas of developing countries” 2018 *Health equity* 174-181

27. Habli, Lawton and Porter 2020 *WHO*

Habli I, Lawton T and Porter Z “Artificial intelligence in health care: accountability and safety” 2020 *WHO* 251-256

28. Ho, Ali and Caals 2020 *WHO*

Ho CLW, Ali J and Caals K “Ensuring trustworthy use of artificial intelligence and big data analytics in health insurance” 2020 *WHO* 263-269

29. Hudson and Rosenbloom 2013 *Hastings Law Journal*

Hudson B and Rosenbloom JD “Uncommon approaches to commons problems: nested governance commons and climate change” 2013 *Hastings law journal* 1274-1342

30. Hutter and Legg 2007 *Minds and Machines*

Hutter M and Legg S “Universal intelligence: a definition of machine intelligence” 2007 *Minds and machines* 391-444

31. Hylton 2014 *Encyclopedia of law and economics*

Hylton KN “Nuisance” 2014 *Encyclopedia of law and economics* 1-15

32. Iszaid, Hafizan and Muhamad 2018 *International journal of public health and clinical sciences*

Iszaid I, Hafizan A and Muhamad HJ "Market failure in health care: A review" 2018 *International journal of public health and clinical sciences* 16-25

33. Kerasidou 2020 *WHO*

Kerasidou M "Artificial intelligence and the ongoing need for empathy, compassion and trust in healthcare" 2020 *WHO* 245-250

34. Kroeze A *Framework*

Kroeze JH "A framework for the Africanisation of the information systems discipline" 2019 *College of science, engineering and technology University of South Africa* 38-45

35. London 2019 *Hastings Cent Rep*

London AJ "Artificial intelligence and black-box medical decisions: accuracy versus explainability" 2019 *Hastings cent rep* 15–21

36. Mariano 2020 *WHO*

Mariano B "Towards a global strategy on digital health" 2020 *WHO* 231-232

37. Mascharka *et al* 2018 *Cornell University*

Mascharka D *et al* "Transparency by design: closing the gap between performance and interpretability in visual reasoning" 2018 *Cornell University* 4942-4943

38. Matsuzaki 2018 *California Western law review*

Matsuzaki K "Ethical issues of artificial intelligence in medicine" 2018
California western law review 255-273

39. Maule 1992 *Tulsa law review*

Maule RM "Applying strict products liability to computer software" 1992
Tulsa law review 735-756

40. Maxwell and Bourreau 2014 *Computer and telecommunications law review*

Maxwell WJ and Bourreau M "Technology neutrality in internet, telecoms
and data protection regulation" 2014 *Computer and telecommunications
law review* 1-6

41. Miron and Zwiebel 1995 *The journal of economic perspectives*

Miron JA and Zwiebel J "The economic case against drug prohibition"
1995 *The journal of economic perspectives* 175-192

42. Moses 2013 *Law, innovation and technology*

Moses LB "How to think about law, regulation and technology: problems
with 'technology' as a regulatory target" 2013 *Law, innovation and
technology* 1-19

43. Nguyen 1994 *Santa Clara law review*

Nguyen FD "Regulation of medical expert systems: a necessary evil"
1994 *Santa Clara law review* 1187-1233

44. Orbach 2012 *Yale journal on regulation online*

Orbach B "What is regulation?" 2012 *Yale journal on regulation online* 1-
10

45. Pasquale 2007 *Minnesota journal of law science and technology*

Pasquale F "Technology, competition, and values" 2007 *Minnesota journal of law science and technology* 607-622

46. Payne *et al* 2017 *Journal of Medical Internet Research*

Payne P *et al* "Enabling open science for health research: collaborative informatics environment for learning on health outcomes (CIELO)" 2017 *Journal of medical internet research* 1-5

47. Peek *et al* 2015 *Artificial intelligence in medicine*

Peek N *et al* "Thirty years of artificial intelligence in medicine (AIME) conferences: a review of research themes" 2015 *Artificial intelligence in medicine* 61-73

48. Powles and Hodson 2017 *Health Technology*

Powles J and Hodson H "Google DeepMind and healthcare in an age of algorithms" 2017 *Health technology* 351–367

49. Price 2017 *The SciTech lawyer*

Price WN II "Artificial intelligence in health care: applications and legal implications" 2017 *The SciTech lawyer* 10-13

50. Reimann 2003 *The American journal of comparative law*

Reimann M "Liability for defective products at the beginning of the twenty-first century: emergence of a worldwide standard" 2003 *The American journal of comparative law* 751–838

51. Rowland and Rowland 1993 *Law, computers and artificial intelligence*

Rowland D and Rowland JJ "Competence and legal liability in the development of software for safety-related applications" 1993 *Law, computers and artificial intelligence* 229–243

52. Saxenian *et al* *Artificial intelligence and life in 2030*

Saxenian A *et al* "Artificial intelligence and life in 2030: One-hundred-year study on artificial intelligence" 2015-2016 *Study Panel Stanford University* 4–52

53. Scherer 2016 *Harvard journal of law and technology*

Scherer UM "Regulating artificial intelligence systems: risks, challenges, competencies and strategies" 2016 *Harvard journal of law & technology* 354-400

54. Scott 2008 *Maryland law review*

Scott MD "Tort liability for vendors of insecure software: has the time finally come?" 2008 *Maryland law review* 426-484

55. Smith *et al* 2006 *University of Washington*

Smith C *et al* "The history of artificial intelligence" 2006 *University of Washington* 1-27

56. Smith *et al* 2020 *WHO*

Smith MJ *et al* "Four equity considerations for the use of artificial intelligence in public health" 2020 *WHO* 291-292

57. Staunton *et al* 2020 *Oxford University press*

Staunton C *et al* "Protection of Personal Information Act 2013 and data protection for health research in South Africa" 2020 *Oxford University press* 1-20

58. Stigler 1971 *The Bell journal*

Stigler JG "The theory of economic regulation" 1971 *The bell journal of economics and management science* 287-299

59. Terry 2018 *Journal of health care law & policy*

Terry NP "Appification, AI, and healthcare's new iron triangle" 2018 *Journal of health care law & policy* 118–182

60. Terry 2019 *Yale journal of law and technology*

Terry NP "Of regulating healthcare AI and robots" 2019 *Yale journal of law and technology* 2-53

61. Thompson and Morgan 2020 *WHO*

Thompson CL and Morgan HM "Ethical barriers to artificial intelligence in the national health service, United Kingdom of Great Britain and Northern Ireland" 2020 *WHO* 293-295

62. Tiažkijus 2012 *Jurisprudencija*

Tiažkijus V "Gaps in labour law and their influence on flexibility and stability of the labour law system" 2012 *Jurisprudencija* 1551-1566

63. Turing 1950 *Mind*

Turing AM "Computing machinery and intelligence" 1950 *Mind* 433-460

64. Turley 1988 *Computer law journal*

Turley TM "Expert software systems: the legal implications" 1988 *Computer law journal* 455-477

65. Tuthill 1991 *AI expert*

Tuthill GS "Legal liabilities and expert systems" 1991 *AI expert* 45–51

66. Tutt 2017 *Administrative law review*

Tutt A "An FDA for algorithms" 2017 *Administrative law review* 84-123

67. Vladeck 2014 *Washington law review*

Vladeck DC “Machines without principles: liability rules and artificial intelligence” 2014 *Washington law review* 117-150

68. Waltz and Firth-Butterfield 2019 *Duke law & technology review*

Waltz A and Firth-Butterfield K “Implementing ethics into artificial intelligence: a contribution, from a legal perspective to the development of an AI governance regime” 2019 *Duke law & technology review* 180-231

69. Wang *et al* 2017 *WHO*

Wang Z *et al* “Records of medical malpractice litigation: a potential indicator of health-care quality in China” 2017 *WHO* 430-436

70. Willick 1983 *The AI Magazine*

Willick MS “Artificial intelligence: some legal approaches and implications” 1983 *The AI magazine* 5-16

71. Wu, Huang and Gong 2020 *Science direct*

Wu W, Huang T and Gong K “Ethical principles and governance technology development of AI in China” 2020 *Science direct* 302-309

72. Yadin 2019 *Cardozo arts & entertainment law journal*

Yadin S “E-Regulation” 2019 *Cardozo arts & entertainment law journal* 101-152

Books

1. Asimov I, *Robot "Runaround"*

Asimov I I, *Robot "Runaround"* The Isaac Asimov Collection ed
(Doubleday New York City 1950)

2. Avis , Pauw and Van Der Spuy *Psychological*

Avis P, Pauw A and Van Der Spuy I *Psychological perspectives, an
introductory workbook* 1st ed (Pearson Education South Africa 1999)

3. Baldwin, Cave and Lodge *Understanding regulation*

Baldwin R, Cave M and Lodge M *Understanding regulation: theory,
strategy, and practice* 2nd ed (Oxford University Press New York 2012)

4. Beauchamp and Childres *Principles of biomedical ethics*

Beauchamp TL and Childres JF *Principles of biomedical ethics* (Oxford
University Press New York 2001)

5. Boberg *Aquilian Liability*

Boberg PQR *The law of delict: Aquilian liability Vol 1* (Juta Cape Town
1984)

6. Boddington *Towards*

Boddington P *Towards a code of ethics for artificial intelligence* (Springer
International Publishing Cham 2017)

7. Bolden *Ubuntu*

Bolden R "Ubuntu" in Coghlan D and Brydon-Miller M (eds)
Encyclopedia of action research (Sage Publications London 2014)

8. Brenner *Law in an era*

Brenner SW *Law in an era of smart technology* (Oxford University Press New York 2007)

9. Brooks *The mythical man-month*

Brooks FP Jr. *The mythical man-month: essays on software engineering* (Addison-Wesley Crawfordsville 1995)

10. Brownsword *Rights, regulation and the technological revolution*

Brownsword R *Rights, regulation and the technological revolution* (Oxford University Press New York 2008)

11. Carpenter *Confidence games*

Carpenter D "Confidence games: how does regulation constitute markets?" in Balleisen EJ and Moss DA (eds) *Government and markets: toward a new theory of regulation* (Cambridge University Press New York 2009) 164–190

12. Chinen *Autonomous*

Chinen M *Law and autonomous machines: the co-evolution of legal responsibility and technology* (Edward Elgar Publishing Cheltenham UK and Massachusetts USA 2019)

13. Collingridge *The social control of technology*

Collingridge D *The social control of technology* (St. Martin's Press New York 1980)

14. Davel and Jordaan *Law of persons*

Davel CJ and Jordaan RA *Law of persons* 4th ed (Juta Cape Town 2005)

15. Feinman *Everything*

Feinman JM *Law: everything you need to know about American law* 4th ed (Oxford University Press New York 2014)

16. Fleming *The law of torts*

Fleming JG *The law of torts* 8th ed (Law Book Co Sydney 1992)

17. Hayek *The Road to Serfdom*

Hayek FA *The Road to Serfdom* (University of Chicago Press Chicago 1944)

18. Hildebrandt *Smart technologies*

Hildebrandt M *Smart technologies and the end(s) of law: novel entanglements of law and technology* (Edward Elgar Publishing Cheltenham UK and Massachusetts USA 2015)

19. Hwang and White *Prognostication and ethics*

Hwang DY and White DB "Prognostication and ethics" in Shutter L and Molyneaux BJ (eds) *Neurocritical care* (Oxford University Press New York 2018)

20. Keeton *et al Products Liability*

Keeton WL *et al Products liability and safety* 2nd ed (Foundation Press Inc. Westbury, N. Y. 1989)

21. Keeton *et al Law of Torts*

Keeton WL *et al Prosser and Keeton on the law of torts* 5th ed (West Publishing Co. Saint Paul, Minnesota 1984)

22. Khan *Medicine*

Khan F “Medicine in the age of smart machines: legal liability challenges” in Olleros FX and Zhegu M (eds) *Research handbook on digital transformations* (Edward Elgar Publishing UK and Massachusetts USA 2016) 59-84

23. Kissick *Medicine's dilemmas*

Kissick W *Medicine's dilemmas: infinite needs versus finite resources* 1st ed (Yale University Press 1994)

24. Koops BJ *Ten dimensions of technology regulation: finding your bearings in the research space of emerging technologies*

Koops BJ “Ten dimensions of technology regulation: finding your bearings in the research space of emerging technologies” in Goodwin MEA, Koops BJ and Leenes RE (eds) *Dimensions of technology regulation* (Wolf Legal Publishers Netherlands 2010) 309-310

25. Koops *Should ICT regulation be technology-neutral?*

Koops BJ “Should ICT regulation be technology-neutral?” in Koops BJ *et al* (eds) *Starting points for ICT regulation: Deconstructing prevalent policy one-liners* (IT & Law Series, T.M.C. Asser Press The Hague 2006) 77-108

26. Lambert *How to regulate*

Lambert TA *How to regulate: a guide for policymakers* (Cambridge University Press New York 2017)

27. Leveson *Safeware*

Leveson NG *Safeware: system safety and computers* (Reading, Mass Addison-Wesley Crawfordsville 1995)

28. Loubser *et al* *The law of delict in South Africa*

Loubser MM *et al* *The law of delict in South Africa* (Oxford University Press New York 2009)

29. Makgoba MW *Patterns of African thought: a critical analysis*

Makgoba MW "Patterns of African thought: a critical analysis" in CW Du Toit (ed) *Faith, science and African culture – African cosmology and Africa's contribution to science* (Research Institute for Theology and Religion University of South Africa) 99-106

30. Mbiti *African religions*

Mbiti JS *African religions & philosophy* 2nd ed (Heinemann Oxford 1990)

31. Midgley and Van der Walt *Principles of delict*

Midgley JR and Van der Walt JC *Principles of delict* 3rd ed (Lexis Nexis Butterworths Durban 2005)

32. Mill *Principles of Political Economy*

Mill JS *Principles of Political Economy with some of their applications to social philosophy* 1st ed (John W. Parker London 1848)

33. Moss & Cisternino *New perspectives*

Moss D and Cisternino J *New perspectives on regulation* 1st ed (The Tobin Project Cambridge 2009)

34. Neethling, Potgieter and Visser *Law of delict*

Neethling J, Potgieter JM and Visser PJ *Law of delict* 7th ed (Lexis Nexis Butterworths Durban 2015)

35. Neethling, Potgieter and Visser *law of personality*

Neethling J, Potgieter JM and Visser PJ *Neethling's law of personality*
2nd ed (Lexis Nexis Butterworths Durban 2005)

36. Poole, Mackworth and Goebel *Computational Intelligence*

Poole DL, Mackworth AK and Goebel R *Computational intelligence: a logical approach* (Oxford University Press New York 1998)

37. Reed and Angel *Computer law*

Reed C and Angel J *Computer law* 5th ed (Oxford University Press New York 2004)

38. Reimann *Product liability*

Reimann M "Product liability" in Bussani M and Sebok AJ (eds) *Comparative tort law: global perspectives* (Edward Elgar Publishing Cheltenham UK and Massachusetts USA 2015) 250 – 278

39. Rich *Artificial intelligence*

Rich E *Artificial intelligence* (McGraw-Hill New York 1983)

40. Russell and Norvig *Artificial Intelligence* 3rd ed

Russell SJ and Norvig P *Artificial intelligence a modern approach* 3rd ed
(Pearson Prentice Hall New Jersey 2010)

41. Russell and Norvig *Artificial Intelligence: A Modern Approach*

Russell SJ and Norvig P *Artificial intelligence: a modern approach*
(Pearson Prentice Hall New Jersey 1995)

42. Russell and Norvig *Instructor's manual*

Russell SJ and Norvig P *Instructor's manual: exercise solutions for artificial intelligence a modern approach 2nd ed* (Pearson Prentice Hall New Jersey 2003)

43. Strous *Should artificial intelligence*

Strous L "Should artificial intelligence be more regulated?" in Strous L and Cerf VG (eds) *Internet of things: information processing in an increasingly connected world* (Springer International Publishing Cham 2019)

44. Waterman *A guide to expert systems*

Waterman DA *A guide to expert systems 1st ed* (Addison-Wesley Crawfordsville 1986)

45. Weaver *Regulation of artificial intelligence*

Weaver JF "Regulation of artificial intelligence in the United States" in Woodrow Barfield W and Pagallo U *Research handbook on the law of artificial intelligence* (Edward Elgar Publishing 2019) 2-671

46. Yudkowsky *Artificial Intelligence as a positive and negative factor in global risk*

Yudkowsky E "Artificial intelligence as a positive and negative factor in global risk" in Bostrom N and Ćirković MM (ed) *Global catastrophic risks* (Oxford University Press New York 2008) 1-518

Internet Sources

1. Abernethy J *et al* "ActiveRemediation: the search for lead pipes in Flint, Michigan"
<https://arxiv.org/pdf/1806.10692.pdf> (Date of use: 16 May 2020)
2. Accenture "Artificial intelligence: healthcare's new nervous system"
https://www.accenture.com/t20171215T032059Z_w_us-en_acnmedia/PDF-49/Accenture-Health-Artificial-Intelligence.pdf (Date of use: 16 May 2020)
3. Accenture "Why artificial intelligence is the future of growth"
<https://www.accenture.com/us-en/insight-artificial-intelligence-future-growth> (Date of use: 16 May 2020)
4. ACM "Code of ethics and professional conduct" <https://www.acm.org/code-of-ethics>
(Date of use: 15 August 2020)
5. AI Now "AI Now Report 2018" https://ainowinstitute.org/AI_Now_2018_Report.pdf
(Date of use: 15 August 2020)
6. American Medical Association "Augmented intelligence in health care H-480.940"
<https://policysearch.ama-assn.org/policyfinder/detail/augmented%20intelligence?uri=%2FAMADoc%2FHOD.xml-I-H-480.940.xml> (Date of use: 16 August 2020)
7. Asilomar "AI principles" <https://futureoflife.org/ai-principles/> (Date of use: 14 August 2020)
8. Association of American Medical Colleges "Will artificial intelligence replace doctors?"
<https://www.aamc.org/news-insights/will-artificial-intelligence-replace-doctors> (Date of use: 20 May 2020)
9. Australian Institute of Health and Welfare "Social isolation and loneliness"
<https://www.aihw.gov.au/reports/australias-welfare/social-isolation-and-loneliness>
(Date of use: 20 May 2020)
10. Anyoha R <http://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/> (Date of use: 8 April 2020)
11. Baker A "Von der Leyen, new commission take aim at AI legislation"
<https://www.atlanticcouncil.org/commentary/blog-post/von-der-leyen-new-commission-take-aim-at-ai-legislation/> (Date of use: 15 August 2020)

12. Basu M “Robot dispenses drugs to HIV patients in South Africa”
<https://govinsider.asia/smart-gov/robot-dispenses-drugs-to-hiv-patients-in-south-africa/>
(Date of use: 3 July 2020)
13. Baweja S and Singh S “Beginning of artificial intelligence, end of human rights”
<https://blogs.lse.ac.uk/humanrights/2020/07/16/beginning-of-artificial-intelligence-end-of-human-rights/> (Date of use: 20 July 2020)
14. Beales H *et al* “Government regulation: the good, the bad, & the ugly”
<https://regproject.org/wp-content/uploads/RTP-Regulatory-Process-Working-Group-Paper.pdf> (Date of use: 14 July 2020)
15. Becker’s Health IT “IBM lobbies congress to ease FDA regulations on Watson”
<https://www.beckershospitalreview.com/healthcare-information-technology/ibm-lobbies-congress-to-ease-fda-regulation-on-watson.html> (Date of use: 5 July 2020)
16. Becker’s Health IT “15 things to know about Amazon's healthcare strategy heading into 2020”
<https://www.beckershospitalreview.com/healthcare-information-technology/15-things-to-know-about-amazon-s-healthcare-strategy-heading-into-2020.html> (Date of use: 5 July 2020)
17. Benjamin P *et al* “What a people-centred response to covid-19 would look like”
<https://www.dailymaverick.co.za/article/2020-07-14-what-a-people-centred-response-to-covid-19-would-look-like/> (Date of use: 23 July 2020)
18. Brodwin D “Healthcare is a market failure”
<https://www.usnews.com/opinion/economic-intelligence/articles/2017-06-23/senate-obamacare-repeal-plan-ignores-market-failure-of-us-health-care> (Date of use: 5 June 2020)
19. Bullock L “The top 6 ways that artificial intelligence will affect your business in the near future”
<https://www.forbes.com/sites/lilachbullock/2019/02/25/the-top-6-ways-that-artificial-intelligence-will-affect-your-business-in-the-near-future/#4eb1811a1966> (Date of use: 15 May 2020)
20. BusinessTech “How encryption can help protect your sensitive data”
<https://businesstech.co.za/news/industry-news/406191/how-encryption-can-help-protect-your-sensitive-data/> (Date of use: 16 June 2020)
21. Business Advocacy Network “Principles of good regulation: OECD and UK better regulation task force”
<http://www.businessadvocacy.net/dloads/fsPrinciplesGoodRegulation.pdf> (Date of use: 14 July 2020)

22. Chatterjee A “Use of artificial intelligence to reduce medical errors”
<https://blog.myhealthvectors.com/index.php/use-of-artificial-intelligence-to-reduce-medical-errors/> (Date of use: 13 May 2020)
23. Cliff Dekker Hofmeyer “AI Regulation in South Africa and the global regulatory trends”
<https://www.cliffedekkerhofmeyr.com/export/sites/cdh/en/news/publications/2019/technology/downloads/Technology-Media-and-Telecommunications-Alert-1-July-2019.pdf>
(Date of use: 30 May 2020)
24. Cohen Healthcare Law “FDA policy on when software that uses artificial intelligence and machine learning qualifies as a medical device”
<https://cohenhealthcarelaw.com/2020/05/fda-policy-on-when-software-that-uses-artificial-intelligence-and-machine-learning-qualifies-as-a-medical-device/> (Date of use: 12 August 2020)
25. Corbett L, Felsner F and Cole M “House of lords select committee publishes report on the future of AI in the UK” <https://www.globalpolicywatch.com/2018/05/house-of-lords-select-committee-publishes-report-on-the-future-of-ai-in-the-uk/> (Date of use: 16 August 2020)
26. Council for Medical Schemes “Fraud in the medical schemes industry”
<https://www.medicalschemes.com/files/CMS%20News/CMSNews1of2018.pdf> (Date of use: 22 June 2020)
27. Creighton J “OpenAI wants to make safe AI, but that may be an impossible task”
<https://futurism.com/openai-safe-ai-michael-page> (Date of use: 6 June 2020).
28. Cyberstoep “Discovery breaks new ground with genetic testing”
<http://www.cyberstoep.co.za/discovery-breaks-new-ground-with-genetic-testing/> (Date of use: 11 June 2020)
29. Danaher J “Is effective regulation of AI possible? Eight potential regulatory problems”
<https://philosophicaldisquisitions.blogspot.com/2015/07/is-effective-regulation-of-ai-possible.html> (Date of use: 7 April 2020)
30. DeepLearning.AI “A visionary doctor prescribes AI”
<https://blog.deeplearning.ai/blog/the-batch-ai-for-medicine-special-eric-topols-planetary-health-system-discovering-drugs-diagnosing-heart-disease-predicting-infections-alexandra-for-doctors> (Date of use: 31 March 2019)

31. Deloitte “2019 Global health care outlook”
<https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Life-Sciences-Health-Care/gx-lshc-hc-outlook-2019.pdf> (Date of use: 18 May 2020)
32. DispatchLive “Google submits to €50m fine by French authority”
<https://www.dispatchlive.co.za/news/2020-06-20-google-submits-to-50m-fine-by-french-authority/> (Date of use: 21 June 2020)
33. Doteveryone “Written evidence (AIC0148)”
<http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/artificial-intelligence-committee/artificial-intelligence/written/69653.html> (Date of use: 22 August 2020)
34. <https://dha.gov.ae/en/RashidHospital/Pages/Pharmacy.aspx> (Date of use: 16 June 2020)
35. Ducharme J “Artificial intelligence could help solve America’s impending mental health crisis” <https://time.com/5727535/artificial-intelligence-psychiatry/> (Date of use: 16 May 2020)
36. Eggers WD and Turley M “The future of regulation: principles for regulating emerging technologies” <https://www2.deloitte.com/content/dam/Deloitte/lu/Documents/risk/lu-future-of-regulation.pdf> (Date of use: 29 June 2020)
37. EIT Health pilot of the AI & ethics guidelines of the European Commission’s (EC) High-Level Group “AI and ethics in the health innovation community”
<https://eithealth.eu/wp-content/uploads/2020/01/AI-and-Ethics-in-the-Health-Innovation-Community.pdf> (Date of use: 14 August 2020)
38. Etzioni A and Etzioni O “Why regulating AI is a mistake”
<https://www.forbes.com/sites/ciocentral/2017/01/09/why-regulating-ai-is-a-mistake/#201e8c9f2be3> (Date of use: 24 June 2020)
39. EU Better Regulation Guidelines “Guidelines on stakeholder consultation”
<https://ec.europa.eu/info/sites/info/files/better-regulation-guidelines-stakeholder-consultation.pdf> (Date of use: 14 July 2020)
40. “European Commission’s Ethics guidelines for trustworthy AI” <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Date of use: 20 June 2020).

41. European Commission “The assessment list for trustworthy artificial intelligence (ALTAI)” <https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment> (Date of use: 25 August 2020)
42. EU “Regulations for Medical Devices” <https://www.gov.uk/guidance/medical-devices-eu-regulations-for-mdr-and-ivdr> (Date of use: 13 August 2020)
43. Fin24 “Digital technology could create, rather than destroy, jobs – expert” <https://www.fin24.com/Tech/digital-technology-could-create-rather-than-destroy-jobs-expert-20170730> (Date of use: 11 May 2020)
44. Forbes “AI and healthcare: a giant opportunity” <https://www.forbes.com/sites/insights-intelai/2019/02/11/ai-and-healthcare-a-giant-opportunity/#56d362cd4c68> (Date of use: 23 March 2019)
45. Future of Life Institute “Benefits and risks of artificial intelligence” <https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/?cn-reloaded=1> (Date of use: 11 April 2020)
46. GDPR (EU) 2016/679 <https://gdpr.eu/article-1-subject-matter-and-objectives-overview/> (Date of use: 20 June 2020)
47. Goertzel KM “Legal liability for bad software” https://zlmonroe.com/CSE566/Readings/47.Legal_Liability_for_Bad_Software.pdf (Date of use: 10 August 2020)
48. GOV.UK. “New code of conduct for artificial intelligence (AI) systems used by the NHS” <https://www.gov.uk/government/news/new-code-of-conduct-for-artificial-intelligence-ai-systems-used-by-the-nhs> (Date of use: 20 May 2019)
49. G20 Trade Ministers “G20 ministerial statement on trade and digital economy” <http://www.g20.utoronto.ca/2019/2019-g20-trade.html> (Date of use: 9 July 2019)
50. *Government Gazette* 42388 of 4 April 2019: Terms of reference for the presidential commission on the fourth industrial revolution <https://www.gov.za/documents/presidential-commission-fourth-industrial-revolution-members-and-terms-reference-9-apr> (Date of use: 3 July 2020)
51. “Google AI principles” <https://www.blog.google/technology/ai/ai-principles/> (Date of use: 15 August 2020)
52. Graetz G and Michaels G “Robots at work” https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2589780## (Date of use: 11 May 2020)

53. Gretton C “The dangers of AI in health care: risk homeostasis and automation bias” <https://towardsdatascience.com/the-dangers-of-ai-in-health-care-risk-homeostasis-and-automation-bias-148477a9080f> (Date of use: 23 July 2020)
54. GSK “Open innovation” <https://uk.gsk.com/en-gb/research/sharing-our-research/open-innovation/> (Date of use: 12 June 2020)
55. Haffajee F “Dlamini Zuma turns cigarettes into illicit drugs as the underground economy takes over” <https://www.dailymaverick.co.za/article/2020-06-12-dlamini-zuma-turns-cigarettes-into-illicit-drugs-as-the-underground-economy-takes-over/> (Date of use: 29 June 2020)
56. Hager GD *et al* “Artificial intelligence for social good” <https://cra.org/ccc/wp-content/uploads/sites/2/2016/04/AI-for-Social-Good-Workshop-Report.pdf> (Date of use: 14 May 2020)
57. Harvard University “The history of artificial intelligence” <http://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/> (Date of use: 8 April 2020)
58. Hawkins A “How elderly, sickly farmers are quenching China's thirst for data” <https://www.wired.co.uk/article/china-ai-healthcare> (Date of use: 18 May 2020)
59. <https://www.hhs.gov/hipaa/for-professionals/security/laws-regulations/index.html> (Date of use: 19 May 2020)
60. Health Capital Consultants “Artificial intelligence in healthcare – competition” https://www.healthcapital.com/hcc/newsletter/07_17/HTML/AI/10.7_formatted_esp_hc_topics_ai_competition_draft_cite_check_7.28.php (Date of use: 25 June 2020)
61. Howard A “The regulation of AI — should organizations be worried?” <https://sloanreview.mit.edu/article/the-regulation-of-ai-should-organizations-be-worried/> (Date of use: 29 June 2020)
62. Health Professions Council of South Africa “Guidelines for good practice in the health care professions” https://www.hpcs.co.za/Uploads/Professional_Practice/Ethics_Booklet.pdf (Date of use: 19 May 2020)
63. Huffpost “Regulating artificial intelligence: a look at the pros and cons” https://www.huffpost.com/entry/regulating-artificial-intelligence-a-look-at-the-pros_b_59eadb0ae4b034105edd4ec4 (Date of use: 18 April 2020)

64. IBM “Conversational AI” <https://www.ibm.com/watson/advantage-reports/future-of-artificial-intelligence/ai-conversation.html> (Date of use: 8 April 2020)
65. IEC “Standards” <https://www.iec.ch/standardsdev/publications/is.htm> (Date of use: 7 August 2020)
66. IEEE “Ethically aligned design: a vision for prioritizing human well-being with autonomous and intelligent system, version 2”
https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf (Date of use: 21 August 2020)
67. IEEE “P7001 transparency of autonomous systems”
<https://standards.ieee.org/project/7001.html> (Date of use: 27 August 2020)
68. IEEE “P7009 Standard for fail-safe design of autonomous and semi-autonomous systems” <https://standards.ieee.org/project/7009.html> (Date of use: 27 August 2020)
69. IEEE “P7010 Well-being metric for autonomous and intelligent systems”
<https://standards.ieee.org/standard/7010-2020.html> (Date of use: 22 August 2020)
70. IEEE-USA <https://ieeeusa.org/volunteers/committees/lrc/> (Date of use: 12 October 2020)
71. I-lawsuit “Comparative and contributory negligence by state” <https://www.i-lawsuit.com/comparative-and-contributory-negligence-by-state/> (Date of use: 12 October 2020)
72. IOL “What 2019 holds in store for pharmaceutical sector”
<https://www.iol.co.za/busindeloitess-report/economy/what-2019-holds-in-store-for-pharmaceutical-sector-19127833> (Date of use: 15 June 2020)
73. International Diabetes Federation “IDF Diabetes Atlas”
<https://www.diabetesatlas.org/en/sections/demographic-and-geographic-outline.html>
(Date of use: 13 June 2020)
74. ISO 13485:2016 “Medical devices - quality management systems - requirements for regulatory purposes” <https://www.iso.org/obp/ui#iso:std:iso:13485:ed-3:v1:en> (Date of use: 12 October 2020)
75. IT web “Artificial intelligence projects now prolific across SA firms”
<https://www.itweb.co.za/content/KA3WwMdIL4EMrydZ> (Date of use: 22 May 2020)
76. IT Web “SA organisations commit to responsible use of AI”
<https://www.itweb.co.za/content/KA3WwqdDj24grydZ> (Date of use: 26 August 2020)

77. IT-Online "Stakeholders commit to responsible AI" <https://it-online.co.za/2020/08/06/stakeholders-commit-to-responsible-ai/> (Date of use: 26 August 2020)
78. International Telecommunication Union "Focus group on artificial intelligence for health" <https://www.itu.int/en/ITU-T/focusgroups/ai4h/Pages/default.aspx> (Date of use: 3 July 2020)
79. Jacobs W, Stoop PN and Van Niekerk R "Fundamental consumer rights under the Consumer Protection Act 68 of 2008: a critical overview and analysis" <http://www.scielo.org.za/pdf/pej/v13n3/v13n3a09.pdf> (Date of use: 5 August 2020)
80. King IV https://cdn.ymaws.com/www.iodsa.co.za/resource/collection/684B68A7-B768-465C-8214-E3A007F15A5A/loDSA_King_IV_Report_-_WebVersion.pdf (Date of use: 8 September 2020)
81. Larson A "Negligence and tort law" https://www.expertlaw.com/library/personal_injury/negligence.html (Date of use: 11 August 2020)
82. Lawyer Monthly "Where artificial intelligence will take legislation" <https://www.lawyer-monthly.com/2017/04/why-regulators-should-take-the-drivers-seat-of-autonomous-vehicles/> (Date of use: 27 June 2020)
83. Lea G "Why we need a legal definition of artificial intelligence" <https://theconversation.com/why-we-need-a-legal-definition-of-artificial-intelligence-46796> (Date of use: 6 April 2020)
84. Legal Dictionary <https://legal-dictionary.thefreedictionary.com/Government+regulation> (Date of use: 30 May 2020)
85. Life Healthcare "Life Healthcare announces cyber incident" <https://www.lifehealthcare.co.za/news-and-info-hub/latest-news/life-healthcare-announces-cyber-incident/> (Date of use: 20 June 2020)
86. Liu X *et al* "A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis" <https://www.thelancet.com/action/showPdf?pii=S2589-7500%2819%2930123-2> (Date of use: 16 June 2020)
87. London School of Economics and Political Science "Electronic health privacy and security in developing countries and humanitarian operations" <http://personal.lse.ac.uk/martinak/eHealth.pdf> (Date of use: 19 May 2020)

88. Mack E “Why Elon Musk put down \$10million to keep artificial intelligence friendly”
<https://www.forbes.com/sites/ericmack/2015/01/15/elon-musk-puts-down-10-million-to-fight-skynet/?sh=1fe213452e5b> (Date of use: 6 June 2020).
89. Marcus G “DeepMind's losses and the future of artificial intelligence”
<https://www.wired.com/story/deepminds-losses-future-artificial-intelligence/> (Date of use: 5 July 2020)
90. Marr B “Coronavirus: how artificial intelligence, data science and technology is used to fight the pandemic” <https://www.forbes.com/sites/bernardmarr/2020/03/13/coronavirus-how-artificial-intelligence-data-science-and-technology-is-used-to-fight-the-pandemic/#5552252b5f5f> (Date of use: 15 April 2020)
91. Marr B “What is the Impact of artificial intelligence (AI) on society?”
<https://bernardmarr.com/default.asp?contentID=1828> (Date of use: 11 April 2020)
92. Marr B “The incredible ways artificial intelligence is now used in mental health”
<https://www.forbes.com/sites/bernardmarr/2019/05/03/the-incredible-ways-artificial-intelligence-is-now-used-in-mental-health/#bf6b0b1d02e4> (Date of use: 16 May 2020)
93. Marr B “How Is AI used in healthcare”
<https://www.forbes.com/sites/bernardmarr/2018/07/27/how-is-ai-used-in-healthcare-5-powerful-real-world-examples-that-show-the-latest-advances/#3756ee895dfb> (Date of use: 13 May 2020)
94. Marr B “The key definitions of artificial intelligence (AI) that explain its importance”
<https://www.forbes.com/sites/bernardmarr/2018/02/14/the-key-definitions-of-artificial-intelligence-ai-that-explain-its-importance/#fe0a3744f5d8> (Date of use: 10 June 2020)
95. Markets and Markets “Pharmaceutical robots market”
<https://www.marketsandmarkets.com/Market-Reports/pharmaceutical-robot-market-210985096.html> (Date of use: 15 June 2020)
96. Martinovs D “10 Organizations leading the way in ethical AI”
<https://ocean.sagepub.com/blog/10-organizations-leading-the-way-in-ethical-ai> (Date of use: 16 August 2020)
97. McCarthy J “What is artificial intelligence?” <http://www-formal.stanford.edu/jmc/whatisai.pdf> (Date of use: 16 April 2020)
98. McGuire B *et al* “The history of artificial intelligence”
<https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf> (Date of use: 29 November 2018)

99. Mckinsey Global Institute “Applying AI for social good”
<https://www.mckinsey.com/~media/mckinsey/featured%20insights/artificial%20intelligence/applying%20artificial%20intelligence%20for%20social%20good/mgi-applying-ai-for-social-good-discussion-paper-dec-2018.ashx> (Date of use: 16 May 2020)
100. Merriam-Webster <https://www.merriam-webster.com/dictionary/regulation#legal>
Dictionary (Date of use 30 May 2020)
101. Medical Academic “R22 billion lost to fraud waste and abuse”
<https://www.medicalacademic.co.za/news/r22-billion-lost-to-fraud-waste-and-abuse/>
(Date of use: 22 June 2020)
102. Miailhe N “AI & global governance: why we need an intergovernmental panel for artificial intelligence” <https://cpr.unu.edu/ai-global-governance-why-we-need-an-intergovernmental-panel-for-artificial-intelligence.html> (Date of use: 29 August 2020)
103. “Microsoft FATE: fairness, accountability, transparency, and ethics in AI”
<https://www.microsoft.com/en-us/research/theme/fate/> (Date of use: 15 August 2020)
104. Mimecast “State of email security report 2019 - South Africa”
https://info.mimecast.com/sa-the-state-of-email-security.html?utm_medium=SEMPPC&utm_source=GooglePPC&utm_campaign=7013l000000TmCuAAK&utm_term=email%20security&gclid=EAlaIQobChMI_K_v5oW46QlVwu3tCh2jzgWiEAAAYASAAEgJL7PD_BwE (Date of use: 15 May 2020)
105. McKinsey and Company Report March 2020 “Transforming healthcare with AI: the impact on the workforce and organization” https://eithealth.eu/wp-content/uploads/2020/03/EIT-Health-and-McKinsey_Transforming-Healthcare-with-AI.pdf (Date of use: 30 July 2020)
106. MHRC “Guidance on medical device stand-alone software including apps (including IVDMDs)”
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/890025/Software_flow_chart_Ed_1-06_FINAL.pdf (Date of use: 13 August 2020)
107. Microsoft “AI in Middle Eastern and African markets – South Africa outlook for 2019 and beyond” <https://info.microsoft.com/rs/157-GQE-382/images/MicrosoftSouthAfricanreportSRGCM1070.pdf> (Date of use: 15 June 2020)

108. Microsoft “Healthcare, artificial intelligence, data and ethics - a 2030 vision”
<https://www.digitaleurope.org/wp/wp-content/uploads/2019/02/Healthcare-AI-Data-Ethics-2030-vision.pdf#page=28> (Date of use: 16 August 2020)
109. MIT “Centennial symposium: One-one-one with Elon Musk”
<https://aeroastro.mit.edu/videos/centennial-symposium-one-one-one-elon-musk> (Date of use: 31 March 2019)
110. Morley S and Lawrence D “Written evidence (AIC0036)”
<http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/artificial-intelligence-committee/artificial-intelligence/written/69435.html> (Date of use: 30 July 2020)
111. Musariri D “John L Hennessy: profiling the man at the helm of Google’s parent company alphabet” <https://www.ns-businesshub.com/business/john-l-hennessy-alphabet/> (Date of use: 2 July 2020)
112. Mutter J “Opportunity from crisis, who really benefits from post-disaster rebuilding efforts” <https://www.foreignaffairs.com/articles/2016-04-18/opportunity-crisis> (Date of use: 25 July 2020)
113. National Institute of Justice “Using artificial intelligence to address criminal justice needs” <https://www.ncjrs.gov/pdffiles1/nij/252038.pdf> (Date of use: 14 May 2020)
114. National Geographic “Five years on, the Flint water crisis is nowhere near over”
<https://www.nationalgeographic.com/environment/2019/04/flint-water-crisis-fifth-anniversary-flint-river-pollution/> (Date of use: 16 May 2020)
115. National Conference of State Legislatures “Autonomous Vehicles - Self Driving Vehicles Enacted Legislation”
<https://www.ncsl.org/research/transportation/autonomous-vehicles-self-driving-vehicles-enacted-legislation.aspx> (Date of use: 9 June 2020)
116. National Health Care Anti-Fraud Association “The challenge of health care fraud”
https://www.nhcaa.org/media/127538/nhcaa_ushealthcaresystem_2017.pdf (Date of use: 22 June 2020)
117. New America “Translation: Chinese AI alliance drafts self-discipline 'joint pledge'”
<https://www.newamerica.org/cybersecurity-initiative/digichina/blog/translation-chinese-ai-alliance-drafts-self-discipline-joint-pledge/> (Date of use: 16 August 2020)
118. New America “Translation: Chinese expert group offers 'governance principles' for 'responsible AI'” <https://www.newamerica.org/cybersecurity->

- [initiative/digichina/blog/translation-chinese-expert-group-offers-governance-principles-responsible-ai/](https://www.digichina.com/blog/translation-chinese-expert-group-offers-governance-principles-responsible-ai/) (Date of use: 16 August 2020)
119. News 24 “Prevalence of diabetes in South Africa”
<https://www.health24.com/Medical/Diabetes/About-diabetes/Diabetes-tsunami-hits-South-Africa-20130210> (Date of use: 13 June 2020)
120. New Scientist “The NHS is setting up a lab for medical artificial intelligence”
<https://www.newscientist.com/article/2212928-the-nhs-is-setting-up-a-lab-for-medical-artificial-intelligence/> (Date of use: 20 May 2020)
121. NHCAA “The challenge of health care fraud”
<https://www.nhcaa.org/resources/health-care-anti-fraud-resources/the-challenge-of-health-care-fraud/> (Date of use: 16 May 2020)
122. NHS Digital “National data opt-out” <https://digital.nhs.uk/services/national-data-opt-out> (Date of use: 13 June 2020)
123. NHS Counter Fraud Authority “Annual report & accounts 2018-19”
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/859362/nhscfa-annual-report-and-accounts-2018-2019.pdf (Date of use: 22 June 2020)
124. NHS Digital “NHS digital, data and technology standards framework”
<https://digital.nhs.uk/about-nhs-digital/our-work/nhs-digital-data-and-technology-standards/framework> (Date of use: 16 August 2020)
125. Nielsen M “Who owns big data?”
<https://www.technologyreview.com/2015/01/05/169719/who-owns-big-data/> (Date of use: 20 May 2020)
126. Nuffield Council on Bioethics “Artificial intelligence (AI) in healthcare and research”
<http://nuffieldbioethics.org/wp-content/uploads/Artificial-Intelligence-AI-in-healthcare-and-research.pdf> (Date of use: 31 March 2019)
127. Obermeyer Z *et al* “Dissecting racial bias in an algorithm used to manage the health of populations” <https://science.sciencemag.org/content/366/6464/447.full> (Date of use: 13 June 2020)
128. OECD “Principles on AI” <http://www.oecd.org/going-digital/ai/principles/> (Date of use: 9 July 2019)

129. OECD “Data governance: enhancing access to and sharing of data”
<https://www.oecd.org/sti/ieconomy/enhanced-data-access.htm> (Date of use: 20 May 2020)
130. OECD “Recommendation of the council on health data governance”
<https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0433> (Date of use: 20 May 2020)
131. OECD “Guidelines on the protection of privacy and transborder flows of personal data”
<https://www.oecd.org/sti/ieconomy/oecdguidelinesonthe protectionofprivacyandtransborderflowsofpersonaldata.htm> (Date of use: 22 August 2020)
132. OECD “Guidelines on measuring subjective well-being”
<http://www.oecd.org/statistics/oecd-guidelines-on-measuring-subjective-well-being-9789264191655-en.htm> (Date of use: 22 August 2020)
133. OpenAI “Charter” <https://openai.com/charter/> (Date of use: 6 June 2020)
134. Oxford Dictionary https://www.lexico.com/en/definition/artificial_intelligence (Date of use: 23 March 2019)
135. Oxford Dictionary https://www.lexico.com/definition/big_data (Date of use: 15 May 2020)
136. Oxford Dictionary <https://www.lexico.com/en/definition/regulation> (Date of use: 30 May 2020)
137. “Partnership on AI” <https://www.partnershiponai.org/#> (Date of use: 6 June 2020)
138. Patel BN *et al* “Human– machine partnership with artificial intelligence for chest radiograph diagnosis” <https://www.nature.com/articles/s41746-019-0189-7> (Date of use: 16 June 2020)
139. Pley C, Dhatt R and Keeling A “Gender bias in health AI - prejudicing health outcomes (or getting it right!)” <https://www.womeningh.org/single-post/2019/09/11/Gender-bias-in-Health-AI---prejudicing-health-outcomes-or-getting-it-right> (Date of use: 13 June 2020)
140. Popper N and Rattman J “The robots are coming for Wallstreet”
<https://www.nytimes.com/2016/02/28/magazine/the-robots-are-coming-for-wall-street.html> (Date of use: 11 May 2020)

141. Pressman A “How AI is aiding the coronavirus fight”
<https://fortune.com/2020/03/16/ai-coronavirus-health-technology-pandemic-prediction/>
(Date of use: 13 May 2020)
142. PWC “No longer science fiction, AI and robotics are transforming healthcare”
<https://www.pwc.com/gx/en/industries/healthcare/publications/ai-robotics-new-health/transforming-healthcare.html> (Date of use: 31 March 2019)
143. PWC “What doctor? Why AI and robotics will define new health”
<https://www.pwc.com/gx/en/industries/healthcare/publications/ai-robotics-new-health/ai-robotics-new-health.pdf> (Date of use: 29 August 2020)
144. Raths D “Digital health dilemma: regulators struggle to keep pace with health-care technology innovation” <https://www.govtech.com/health/Digital-Health-Dilemma-Regulators-Struggle-to-Keep-Pace-with-Health-Care-Technology-Innovation.html>
(Date of use: 29 June 2020)
145. Regulation (EU) 2017/745 of the European Parliament and of the Council <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32017R0745&from=EN> (Date of use: 27 August 2020)
146. Right ePharmacy <https://rightepharma.co.za/solutions/tele-pharmacy/> (Date of use: 3 July 2020)
147. Rowe J “Cerner partners with Dubai hospital in AI research center”
<https://ai.healthcareitnews.com/ai-powered-healthcare/cerner-partners-dubai-hospital-ai-research-center-0> (Date of use: 16 June 2020)
148. Rowland D “Negligence, professional competence and computer systems”
https://warwick.ac.uk/fac/soc/law/elj/jilt/1999_2/rowland/#Carr (Date of use: 7 August 2020)
149. Rural Health Advocacy Project “Protecting rural healthcare in times of economic crisis there is another way – An overview”
https://www.groundup.org.za/media/uploads/documents/Austerity%20report_FINAL.pdf
f (Date of use: 18 May 2020)
150. Russell S “Provably beneficial artificial intelligence”
<https://people.eecs.berkeley.edu/~russell/papers/russell-bbvabook17-pbai.pdf> (Date of use: 6 June 2020)
151. Sallstrom L, Morris O and Mehta H “Artificial intelligence in Africa’s healthcare: ethical considerations” <https://www.orfonline.org/wp->

- [content/uploads/2019/09/ORF_Issue_Brief_312_AI-Health-Africa.pdf](#) (Date of use: 19 May 2020)
152. SACAP “The shocking state of mental health in South Africa in 2019”
<https://www.sacap.edu.za/blog/management-leadership/mental-health-south-africa/>
(Date of use: 15 May 2020)
153. Salathé M, Wenzel M and Kishnamurthy R “Focus group on artificial intelligence for health” https://www.itu.int/en/ITU-T/focusgroups/ai4h/Documents/FG-AI4H_Whitepaper.pdf (Date of use: 13 June 2020)
154. SAHPRA “Medical devices” <https://www.sahpra.org.za/medical-devices/> (Date of use: 13 August 2020)
155. SAPHRA “Medical device quality manual” https://www.sahpra.org.za/wp-content/uploads/2020/01/Medical_Device_Quality_Manual_Nov19_v2.pdf (Date of use: 12 October 2020)
156. SAHPRA “Guideline for a licence to manufacture, import, export or distribute medical devices and IVD’s” https://www.sahpra.org.za/wp-content/uploads/2020/01/Licence_Medical_Devices_IVDs_Nov19_v3.pdf (Date of use: 12 October 2020)
157. Schwarb K “The fourth Industrial Revolution: what it means, how to respond”
<https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/> (Date of use: 30 March 2019)
158. <https://www.sciencedaily.com/releases/2020/03/200331092704.htm> (Date of use: 16 May 2020)
159. Sebulela G “Covid-19: opportunity for investment in South Africa’s healthcare sector”
<https://www.iol.co.za/business-report/opinion/covid-19-opportunity-for-investment-in-south-africas-healthcare-sector-46372334> (Date of use: 20 July 2020)
160. Shetty S and Tse D “Using AI to improve breast cancer screening”
<https://www.blog.google/technology/health/improving-breast-cancer-screening/> (Date of use: 17 May 2020)
161. Silver N “Rage against the machines” <https://fivethirtyeight.com/features/rage-against-the-machines/> (Date of use: 25 June 2020)
162. Smart Dubai “Artificial intelligence principles & ethics”
<https://www.smartdubai.ae/initiatives/ai-principles-ethics> (Date of use: 16 June 2020)

163. Smialowski B “Should artificial Intelligence be regulated?”
<https://www.forbes.com/sites/quora/2017/08/31/should-artificial-intelligence-be-regulated/#26674c3d331d> (Date of use: 24 June 2020)
164. Software Engineer Insider “The movement to license software engineers”
<https://www.softwareengineerinsider.com/articles/software-engineer-licensure.html>
(Date of use: 12 October 2020)
165. Society of Actuaries “Predictive analytics in healthcare trend forecast”
<https://www.soa.org/globalassets/assets/Files/programs/predictive-analytics/2017-health-care-trend.pdf> (Date of use: 15 May 2020)
166. South African Pharmacy Council “Good pharmacy practice manual and associated SAPC rules” https://www.pharmcouncil.co.za/Legislation_Rules (Date of use: 3 July 2020)
167. Stinson L “Remembering the legend behind 'form follows function”
<https://www.wired.com/2015/09/man-coined-form-follows-function-born-today/> (Date of use: 20 July 2020)
168. Straub J “Does regulating artificial intelligence save humanity or just stifle innovation?” <https://techxplore.com/news/2017-10-artificial-intelligence-humanity-stifle.html> (Date of use: 31 March 2019)
169. Tangcharoensathien V, Boonperm J and Jongudomsuk P “Sharing health data: developing country perspectives” <https://www.who.int/bulletin/volumes/88/6/10-079129.pdf> (Date of use: 19 May 2020)
170. Tengai “The interview robot – how it works” <https://www.tengai-unbiased.com/how-robot-job-interview/> (Date of use: 27 June 2020)
171. The Alan Turing Institute “Understanding artificial intelligence ethics and safety”
https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf (Date of use: 21 May 2020)
172. The BMJ “What does successful regulation look like?”
<https://www.bmj.com/content/350/bmj.h1641> (Date of use: 14 July 2020)
173. The Digital Stream “How AI in video games is furthering our knowledge of programming AIs in general”
<https://andrewdouglasblog.wordpress.com/2016/09/28/how-ai-in-video-games-is-furthering-our-knowledge-of-programming-ais-in-general/> (Date of use: 8 April 2020)

174. The Foundation for Law and International Affairs “Next generation artificial intelligence development plan” <https://flia.org/wp-content/uploads/2017/07/A-New-Generation-of-Artificial-Intelligence-Development-Plan-1.pdf> (Date of use: 16 August 2020)
175. The Guardian “Elon Musk: artificial intelligence is our biggest existential threat” <https://www.theguardian.com/technology/2014/oct/27/elon-musk-artificial-intelligence-ai-biggest-existential-threat> (Date of use: 5 June 2020)
176. The Guardian “Royal Free breached UK data law in 1.6m patient deal with Google's DeepMind” <https://www.theguardian.com/technology/2017/jul/03/google-deepmind-16m-patient-royal-free-deal-data-protection-act> (Date of use: 13 June 2020)
177. The Urology Hospital “What is Robotic Surgery” <https://urology.co.za/health-professionals/robotic-surgeons/> (Date of use: 9 July 2019)
178. The Vaccine Reaction “Medical Errors Kill Five People Per Minute, 2.6 Million People Every Year” <https://thevaccinereaction.org/2019/11/medical-errors-kill-five-people-per-minute-2-6-million-people-every-year/> (Date of use: 13 May 2020)
179. The World Bank Group “Helping countries combat corruption: the role of the world bank” <http://www1.worldbank.org/publicsector/anticorrupt/corruptn/cor02.htm> (Date of use: 19 May 2020)
180. Thomadakis SB “What makes good regulation?” https://www.ifac.org/system/files/downloads/30th_anniversary_Thomadakis_Pres_Nov_07.pdf (Date of use: 15 July 2020)
181. Thomson Reuters Foundation and TrustLaw Connect “Patient privacy in a mobile world: a framework to address privacy laws in mobile health” <http://www.trust.org/contentAsset/raw-data/03172beb-0f11-438e-94be-e02978de3036/file> (Date of use: 19 May 2020)
182. Trading Economics <https://tradingeconomics.com/south-africa/rural-population-percent-of-total-population-wb-data.html> (Date of use: 18 May 2020)
183. Turpitka D “The challenges of applying AI to healthcare and how to address them” <https://www.forbes.com/sites/forbestechcouncil/2020/03/31/the-challenges-of-applying-ai-to-healthcare-and-how-to-address-them/?sh=17d9d19e5557> (Date of use: 25 July 2020)

184. UK Industrial Strategy White Paper
<https://www.gov.uk/government/publications/industrial-strategy-building-a-britain-fit-for-the-future> (Date of use: 6 June 2020)
185. UK Department for Business, Energy and Industrial Strategy “Better regulation framework guidance August 2018”
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/872342/better-regulation-guidance.pdf (Date of use: 14 July 2020)
186. UK Engineering Council “Guidance for institution codes of conduct”
<https://www.engc.org.uk/engcdocuments/internet/website/Guidance%20for%20Institution%20Codes%20of%20Conduct%202017.pdf> (Date of use: 7 August 2020)
187. UK “Artificial intelligence and public standards report”
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/868284/Web_Version_AI_and_Public_Standards.PDF (Date of use: 16 August 2020)
188. UK House of Lords Select Committee on Artificial Intelligence Report “AI in the UK: ready, willing and able?”
<https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf> (Date of use: 16 August 2020)
189. UK Independent Review “Artificial intelligence and public standards”
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/868284/Web_Version_AI_and_Public_Standards.PDF (Date of use: 20 June 2020)
190. UK “Code of conduct for data-driven health and care technology”
<https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology> (Date of use: 23 July 2020)
191. UKGC “UK Caldicott Guardian Council” <https://www.ukcgc.uk/> (Date of use: 25 August 2020).
192. UNESCO “Artificial intelligence: towards a humanistic approach”
<https://en.unesco.org/artificial-intelligence> (Date of use: 9 July 2019)
193. UNESCO “First draft of the recommendation on the ethics of artificial intelligence”
<https://en.unesco.org/artificial-intelligence/ethics> (Date of use: 28 December 2020)

194. UNESDOC Digital Library “Planning education in the AI era: Lead the leap”
<https://unesdoc.unesco.org/ark:/48223/pf0000368303> (Date of use: 9 July 2019).
195. United Nations “Sustainable development goals”
<https://www.un.org/sustainabledevelopment/sustainable-development-goals/> (Date of use: 16 August 2020)
196. USAID “Artificial intelligence in global health defining a collective path forward”
<https://www.usaid.gov/cii/ai-in-global-health> (Date of use: 22 May 2020)
197. US Food and Drug Administration “A history of medical device regulation & oversight in the United States” <https://www.fda.gov/medical-devices/overview-device-regulation/history-medical-device-regulation-oversight-united-states> (Date of use: 5 July 2020)
198. US Food and Drug Administration “Artificial intelligence and machine learning in software as a medical device” <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device#transforming> (Date of use: 11 August 2020)
199. US Food and Drug Administration “Software as a medical device”
<https://www.fda.gov/medical-devices/digital-health/software-medical-device-samd>
(Date of use: 12 August 2020)
200. US Food and Drug Administration “Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD)” <https://www.fda.gov/media/122535/download> (Date of use: 11 August 2020)
201. Valiani S “Public health care spending in South Africa and the impact on nurses: 25 years of democracy?” https://hasa.co.za/wp-content/uploads/2020/02/Public_Health_Care_Spending_in_South_Afr.pdf (Date of use: 18 May 2020)
202. Van Demark DC “The future of AI regulation: the government as regulator and research & development participant” <https://mcdermott-will-emery-2793.docs.contently.com/v/the-future-of-ai-regulation-the-government-as-regulator-and-research-development-participant> (Date of use: 29 June 2020)
203. Von der Leyen U “Political guidelines for the next European commission 2019-2024”
<https://www.europarl.europa.eu/resources/library/media/20190716RES57231/20190716RES57231.pdf> (Date of use: 16 August 2020)

204. Wilson HJ, Daugherty PR and Davenport C “The future of AI will be about less data, not more” <https://hbr.org/2019/01/the-future-of-ai-will-be-about-less-data-not-more> (Date of use: 20 May 2020)
205. World Development Report 2016 “Digital dividends” <http://documents.worldbank.org/curated/en/896971468194972881/pdf/102725-PUB-Replacement-PUBLIC.pdf> (Date of use: 11 May 2020)
206. World Economic Forum “Global AI in financial services survey 2020” http://www3.weforum.org/docs/WEF_AI_in_Financial_Services_Survey.pdf (Date of use: 14 May 2020)
207. World Economic Forum “Deep shift technology tipping points and societal impact” http://www3.weforum.org/docs/WEF_GAC15_Technological_Tipping_Points_report_2015.pdf (Date of use: 24 May 2020)
208. World Economic Forum “AI has started a financial revolution – here’s how” <https://www.weforum.org/agenda/2020/02/how-ai-is-shaping-financial-services/> (Date of use: 15 June 2020).
209. WHO “Blindness and vision impairment” <https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment> (Date of use: 16 May 2020).
210. WHO “Guidelines for drinking water quality” https://www.who.int/water_sanitation_health/dwq/chemicals/lead.pdf (Date of use: 16 May 2020)
211. WHO “Drinking water” <https://www.who.int/news-room/fact-sheets/detail/drinking-water> (Date of use: 16 May 2020)
212. WHO “mHealth: New horizons for health through mobile technologies” http://www.who.int/goe/publications/goe_mhealth_web.pdf (Date of use: 19 May 2020)
213. WHO “Patient safety” <https://www.who.int/news-room/facts-in-pictures/detail/patient-safety> (Date of use: 20 May 2020)
214. WHO “International community rallies to support open research and science to fight COVID-19” <https://www.who.int/news-room/detail/29-05-2020-international-community-rallies-to-support-open-research-and-science-to-fight-covid-19> (Date of use: 12 June 2020)
215. WHO “Data and innovation: draft global strategy on digital health” https://apps.who.int/gb/ebwha/pdf_files/EB146/B146_26-en.pdf (Date of use: 3 July 2020)

216. White House Report 2016 “Preparing for the future of intelligence”
https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf (Date of use: 11 May 2020)
217. White House “Guidance for regulation of artificial intelligence applications 2020”
<https://www.whitehouse.gov/wp-content/uploads/2020/01/Draft-OMB-Memo-on-Regulation-of-AI-1-7-19.pdf> (Date of use: 8 June 2020)
218. White Paper on Artificial Intelligence “A European approach to excellence and trust”
https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf (Date of use: 6 June 2020)
219. White Paper on Department for Business Energy and Industrial Strategy “Industrial strategy: building a Britain fit for the future”
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/664563/industrial-strategy-white-paper-web-ready-version.pdf (Date of use: 15 June 2020)
220. Wyber R *et al* “Big data in global health: improving health in low- and middle-income countries” <https://www.who.int/bulletin/volumes/93/3/14-139022/en/> (Date of use: 15 May 2020)
221. Yatich B “SophieBot AI wants to replace your doctor”
<https://www.afritechmedia.com/sophiebot-ai-wants-to-replace-your-doctor/> (Date of use: 24 June 2020)
222. Zhihao Z “AI development plan draws map for innovation”
<http://www.chinadaily.com.cn/a/201908/05/WS5d476b48a310cf3e35563d0d.html>
(Date of use: 16 August 2020)
223. Zhang B and Dafoe A “Artificial intelligence: American attitudes and trends (January 2019)” https://governanceai.github.io/US-Public-Opinion-Report-Jan-2019/us_public_opinion_report_jan_2019.pdf (Date of use: 6 June 2020)

Cases

South Africa

1. *Administrateur Natal v Trust Bank van Afrika Bpk* 1979 3 SA 824 (A)
2. *AN v MEC for Health, Eastern Cape* 2019 4 All SA 1 (SCA)
3. *Argus Printing and Publishing Co Ltd v Inkatha Freedom Party* 1992 3 SA 579 (A)
4. *Bayer South Africa (Pty) Ltd v Viljoen* 1990 2 SA 647 (A)
5. *Carmichele v Minister of Safety and Security* 2001 4 SA 938 (CC)
6. *Clarke v Hurst* 1992 4 SA 630 (D)
7. *Coronation Brick (Pty) Ltd v Strachau Construction Co (Pty) Ltd* 1982 4 SA 371 (D)
8. *Fair-Trade Independent Tobacco Association v President of the Republic of South Africa and Another* 2020 (6) SA 513 (GP)
9. *Fourway Haulage SA v SA National Roads Agency* 2009 (2) SA 150 (SCA)
10. *Goliath v The MEC for Health, Eastern Cape* 2015 (2) SA 97 (SCA)
11. *H v Fetal Assessment Centre* 2015 (2) SA 193 (CC)
12. *Herschel v Mrupe* 1954 3 SA 464 (A)
13. *International Shipping Co (Pty) Ltd v Bentley* 1990 1 SA 680 (A)
14. *Kroonstad Westelike Boere-Ko-operatiewe Vereniging Bpk v Botha and Another* 1964 SA 561 (A)
15. *Kruger v Coetzee* 1966 2 SA 428 (A)
16. *Khupa v SA Transport Services* 1990 2 SA 629 (W)
17. *Langeberg Voedsel Bpk v Sarculum Boerdery Bpk* 1996 2 SA 565 (A)
18. *Lomagundi Sheetmetal and Engineering v Bassonn* 1973 4 SA 523 (RA)
19. *Links v MEC Department of Health, Northern Province* 2016 (4) SA 414 (CC)
20. *Mankayi v Anglogold Ashanti Ltd* 2011 (3) SA 237 (CC)
21. *Minister of Police v Skosana* 1977 1 SA 31 (A)
22. *Ngubane v SA Transport Services* 1991 1 SA 756 (A)
23. *NM v Smith* 2007 5 SA 250 (CC)
24. *Oppelt v Department of Health, Western Cape* 2016 (1) SA 325
25. *Perlman v Zoutendyk* 934 CPD 151
26. *S v Makwanyane* 1995 (6) BCLR 665 (CC)
27. *S v Mokgethi* 1990 1 SA 32 (A)
28. *S v Ngubane* 1985 (2) All SA 340 (A)

29. *S v Humphreys* 2015 (1) SA 491 (SCA)
30. *Santam Versekeringsmaatskappy Bpk v Swart* 1987 2 All SA 443 (A)
31. *Savage and Lovemore Mining (Pty) Ltd v International Shipping Co (Pty) Ltd* 1987 2 SA 149 (W)
32. *Simon's Town Municipality v Dews* 1993 1 SA 191 (A)
33. *Smit v Abrahams* 1994 (4) SA 1 (A)
34. *Stoffberg v Elliot* 1923 CPD 148 148.
35. *Tuck v Commissioner for Inland Revenue* 1988 SA 819 (A)
36. *Universiteit Van Pretoria v Tommie Meyer Films (Edms) Bpk* 1977 4 SA 376 (T)
37. *Van Schalkwyk v S* 2016 (2) SACR 334 (SCA)
38. *Van Wyk v Lewis* 1924 AD 438
39. *Wagener v Pharmacare Ltd Cuttings v Pharmacare Ltd* 2003 4 SA 285 (SCA)

Germany

1. *Bundesgerichtshof [BGH] [Federal Court of Justice]* June 13, 2006, case no. VI ZR 323/04, NJW 2006, 2477 (Ger.)

New Zealand

1. *Bevan Investments v Blackhall and Struthers No 2* [1973] 2 NZLR 45

United States

1. *Borel v Fibreboard Paper Products Corporation* 493 F.2d 1076, 1088-89 (5th Cir.1973)
2. *Columbus McKinnon Corp. v China Semiconductor Co., Ltd.*, 867 F. Supp. 1173, 1182-83 (1994)
3. *Chatlos Systems, Inc v National Cash Register Corp* (1979) 479 F Supp 738 (DNJ) 740
4. *Data Processing Services Inc v LH Smith Oil Corporation.*, 492 N.E.2d 314 (Ind. Ct. App. 1986)
5. *Fancher v Southwest Missouri Truck Center Inc* (1981) Mo App 618 SW 2d 271
6. *Greenman v Yuba Power Products, Inc.* 377 P.2d 897 (Cal. 1962)
7. *Hospital Computer Systems, Inc. v Staten Hospital* 788 F Supp 1351 (DNJ 1992)

8. *Hyder v Weilbaecher* 1981 54 NC App 287 283 SE 2d 426
9. *Invacare Corp. v Sperry Corp.*, 612 F. Supp. 448 (N.D. Ohio 1984)
10. *In re Toyota Motor Corp. Unintended Acceleration Mktg, Sales Practices, & Prods. Liab Litig* 754 F. Supp. 2d 1145 (C.D. Cal. Oct. 7, 2013)
11. *KW v Armstrong*, 180 F. Supp. 3d 703 (D. Idaho 2016)
12. *Li v Yellow Cab Co* 13 Cal.3d 804, 532 P.2d 1226 (1975),
13. *La Vine v Clear Creek Skiing Corp* (1977) 557 F 2d 730 (10th Cir)
14. *MacPherson v Buick Motor* (1916) 217 NY 382, 111 NE 1050
15. *MAI Basic Four, Inc. v Generic Bus. Solutions, Inc.* CIV. A. No. 9908, 1990 WL 3665
16. *Milwaukee & St. Paul Ry. v Arms* 91 U.S. 489, 492–93 (1875)
17. *Prooth v Walsh* (1980) 105 Misc 2d 653; 432 NYS 2d 668 (Sup)
18. *Perrine v Pac. Gas & Elec. Co.* 186 C.A.2d 442, 449 (1960)
19. *Rudisaile v Hawk Aviation Inc.* 92 NM 575, 592 P.2d 175 (1979)
20. *Saloomey v Jeppesen & Co* 707 F.2d 671 (2d Cir. 1983)
21. *S Cent Bell Tel Co v Barthelemy* 643 So. 2d 1240, 1245
22. *Summers v Tice* 33 Cal.2d 80, 199 P.2d 1
23. *The TJ Hooper v N Barge Corp* 60 F.2d 737 (2d Cir. 1932)
24. *Triangle v Underwriters, Inc v Honeywell, Inc.* (1979) 604 F 2d 737 (2d Cir)
25. *Ultramares Corporation v Touche* 174 N.E. 441 (1932)
26. *United States v Athlone Indus., Inc.*, 746 F.2d 977, 979 (3d Cir. 1984)
27. *United States v Carroll Towing Co.* 159 F.2d 169 (2d. Cir. 1947)
28. *Vaughan v Menlove* (1837) 3 Bing NC 468; 132 Eng Rep 490
29. *WalMart Stores Inc v City of Mobile* 696 So. 2d 290, 291
30. *Ward v Hobart Manufacturing Co* (1971) 450 F 2d 1176 (5th Cir)
31. *Winter v GP Putnam's Sons* 938 F.2d 1033, 1035 (9th Cir. 1991)
32. *Ybarra v Spangard* 25 Cal.2d 486, 154 P.2d 687

United Kingdom

1. *Anns v Merton London Borough Council* (1978) AC 728
2. *Bolton v Stone* (1951) AC 850
3. *Barnett v Chelsea & Kensington Hospital Management Committee* (1968) 2 WLR 422

4. *Bolam v Friern Hospital Management Committee* (1957) 1 WLR 582
5. *Blyth v Company Proprietors of the Birmingham Water Works* (1856) 11 Ex Ch 781
6. *Byrne v Boadle* 2 H. & C. 722, 159 Eng. Rep. 299 (Exch. 1863)
7. *Caparo Industries Plc v Dickman* (1990) UKHL 2
8. *Carr v Inland Revenue Commissioners* (1944) 2 All ER 163
9. *Donoghue v Stevenson* (1932) UKHL 100
10. *Eckersley v Binnie and Partners* 1988 18 Con LR 1
11. *Hedley Byrne v Heller and Partners* (1964) AC 465
12. *Home Office v Dorset Yacht Co Ltd* (1970) AC 1004
13. *Hughes v Lord Advocate* (1963) UKHL 31
14. *Independent Broadcasting Authority v EMI Electronics* (1980) 14 Build LR 1
15. *Jayes v IMI Kynoch* (1985) ICR 155
16. *Lamb v Camden LBC* (1981) QB 625
17. *Latimer v AEC* (1952) 2 QB 701
18. *Montgomery v Lanarkshire Health Board* (2015) UKSC 11
19. *MW Mfrs Inc v Friedman Corp* No. 97-C-8319, 1998 WL 417501
20. *Nettleship v Weston* (1971) 2 QB 691
21. *Overseas Tankship (UK) Ltd v Morts Dock and Engineering Co Ltd or Wagon Mound (No. 1)* (1961) AC 388
22. *Overseas Tankship (UK) Ltd v The Miller Steamship Co or Wagon Mound (No. 2)* (1967) 1 AC 617
23. *Rondel v Worsley* (1969) 1 AC 191
24. *Smith v Leech Brain & Co* (1962) 2 QB 405
25. *The Lady Gwendoline* (1965) 3 WLR