# The Legal Impact of Artificial Intelligence on the New Zealand Health System

A thesis submitted in fulfilment of the requirements for the degree of

**Doctor of Philosophy**

in the

University of Canterbury School of Law

by

Christopher Ryan Boniface

University of Canterbury

2021

# Abstract

This thesis provides an analysis of the legal impact that artificial intelligence (AI) technologies may have on New Zealand's healthcare system. The focus is on whether the established rights and protections afforded to patient's can be readily applied or interpreted in situations involving an artificial intelligence system. Using the Code of Patient's Rights[1] as a framework for the rights afforded and the underlying ideals contained, this thesis engages with both the codified rights as well as their common law applications. Doing so provides a robust and comprehensive analysis of what features of an AI and which current formulations within the law are incompatible or at least require new interpretations to interact.

The thesis begins with an overview of the background to both the healthcare landscape and the technology of AI in general. And then follows this with an overview of the rights afforded to patient's within New Zealand, and the ideals which can be inferred from this which will be used as the foundation for the thesis' substantive analysis. The thesis then engages with four of the rights outlined – discrimination, privacy, informed consent, and negligence – chosen to provide the most widely applicable discussion, as well as those which are considered the most integral to a patient's care. Included in the substantive analysis of negligence is a discussion of the Accident Compensation Act 2001 and the role it may play in mitigating some of the commonly lamented issues with AI technology. The thesis then ends with a discussion of prospective reforms that may assist with the interaction with AI and any associated issues with them, and then a set of recommendations into how New Zealand should proceed in the coming years.

[1] Full name the Health and Disability Commissioner (Code of Health and Disability Services Consumers' Rights) Regulations 1996

# Acknowledgements

I would foremost like to thank my two supervisors, Associate Professor Debra Wilson & Professor Stephen Todd in the Law School for their support. Their feedback, advice, humour and support throughout these four years has been invaluable to ensuring my work was both completed and that my prospective academic career was entered with enthusiasm and clarity of purpose.

I would also like to thank the larger Law School for its support over the years in terms of research funding, administrative support, work experience, and career support. I have been provided with opportunities to grow as a teacher, tutor, researcher and member of the legal community thanks to a multitude of people. A special thanks must be given to Associate Professor Christian Riffel for providing the motivation to pursue postgraduate study during the penultimate year of my LLB. Without his enthusiastic request for me to pursue a Master's degree, I am unlikely to have taken the leap myself.

Special thanks must be given to those I met throughout the years at academic conferences who have become both friends and colleagues on my academic journey. Too many to name have been supportive and motivational voices which provided much needed boosts at different points throughout this thesis' development.

And finally, endless appreciation to my family and my amazing wife Vanessa, who have supported me throughout this journey in more ways than can be listed. Without the support network provided both at home and within my day-to-day life, this thesis would not have been possible and I will be forever grateful.

# Statement of Publications

During to the course of enrolment as a Doctoral candidate at the University of Canterbury, portions of this thesis have been published elsewhere as stated below.

Additionally, work has been published as articles separate to my thesis which have been utilized within this thesis.

Medical negligence in the age of artificial intelligence [2020] NZLJ 136

Except those sections which are listed above, the text of this thesis is original and has not been published elsewhere prior to submission and examination.

All work presented within this thesis is my own, except where materials are quoted, cited and attributed to other authors in footnotes and the bibliography.

# List of Acronyms

Throughout this thesis, many acronyms are used to refer to both concepts and legal instruments in an easier way. Below is a comprehensive list of the acronyms used within this thesis.

| | |
|---|---|
| "Scenarios" 1, 2 and 3 | SN1, 2 & 3 |
| Accident compensation | AC |
| Accident Compensation Act 2001 | ACA |
| Accident Compensation Commission | ACC |
| Ad Hoc Expert Group | AHEG |
| Artificial intelligence | AI |
| Black-Box Problem | BBP |
| Consumer Guarantees Act 1993 | CGA |
| Control of Design | COD |
| Crown Entities (NZ) | CE |
| Departments of State (NZ) | DOS |
| District Health Boards | DHBs |
| European Union | EU |
| Fair Trading Act 1986 | FTA |
| General Data Protection Regulation 2016 (EU) | GDPR |
| General Practice | GP |
| Health and Disability Commissioner | HDC |
| Health and Disability Commissioner Act 1994 | HDCA |
| Health and Disability Services Act 2001 | HDSA |
| Health Information Privacy Code 2020 | HIPC |
| Health Insurance Portability and Accountability Act of 1996 (USA) | HIPAA |
| Health New Zealand | HNZ |
| Health Practitioners Competence Assurance Act 2003 | HPCAA |
| International Covenant on Civil and Political Rights | ICCPR |
| Māori Health Authority | MHA |
| Ministry of Health (NZ) | MOH |
| Ministry of Justice (NZ) | MOJ |
| National Health Service | NHS |
| New Zealand | NZ |
| New Zealand Bill of Rights Act 1994 | NZBORA |
| New Zealand Public Health and Disability Act 2000 | NZPHDA |
| Primary Health Organisations | PHO |
| Primary Health Units | PHU |
| Personal Injury | PI |
| Personal injury by accident | PIBANA |
| Sale of Goods Act 1908 | SOGA |
| The Code of Patient's Rights | CODE |
| The Court of Appeal | COA |
| The High Court | HC |
| The Prime Minister's Business Advisory Council | BAC |
| The Privacy Act 2020 | PA |
| The Supreme Court | SC |
| The United Kingdom | UK |
| The United Kingdom House of Lord's | UKHoL |
| The United Nations | UN |
| The United States of America | USA |
| Universal Declaration of Human Rights | UDHR |

# Table of Contents

# Chapter 1: Introduction

---

"If a risk falls within the bounds of an existing regulatory regime…the policy discussion should start by considering whether the existing regulations already adequately address the risk, or whether they need to be adapted to the addition of AI"

The US National Science and Technology Council.[2]

"Although there might be no part of the regulatory array that is specifically dedicated to the emerging technology, and although there might be gaps in the array, it will rarely be true to say that an emerging technology finds itself in a regulatory void."

Roger Brownsword and Morag Goodwin.[3]

## 1.1 Introduction

Artificial Intelligence (AI) has started to generate a great deal of academic interest in the 21st Century, as it begins to move from the realm of science fiction into reality. There has been growing interest from a wide-range of disciplines, and therefore such the law has begun to grapple with the issues raised by this emergent technology in a variety of contexts. This thesis aims to contribute to this growing field of knowledge by analysing the potential impact that AI will have on the healthcare system of New Zealand (NZ) focusing on the rights afforded to patients, and their associated legal mechanisms for dispute resolution. NZ healthcare operates with a carefully crafted network of rights, duties and obligations for patients, medical practitioners, and associated organisational bodies which has developed for over a century. It is crucial to understand how the system is affected by AI technologies introduction for two reasons: (1) to identify the adaptability of the system, and how it may best respond, and (2) to identify the incompatibilities that cannot be managed are which warrant reform. Healthcare is a setting in which technological innovation has the potential to

---

[2] The US National Science and Technology Council, "Preparing for the future of artificial intelligence" (2016)" available at <https://publicintelligence.net/whitehouse-preparing-artificial-intelligence>

[3] Roger Brownsword and Morag Goodwin *Law and technologies of the twenty-first century* (Cambridge University Press, Cambridge, 2012) at 64

not only revolutionise but also endanger. The system must be prepared for potential paradigm shifts and unknowns as it affects those at their most vulnerable, and the consequences of ineffectual preparedness can be serious.

This thesis will engage with this interaction in a two-fold inquiry:

(1) how the legal rights and operations are affected in practice, and the potential issues associated with this (the 'doctrinal analysis'), and

(2) the conceptual ideals embedded within these rights and how they may be impaired by the involvement of AI systems, as well as the potential attempts to mitigate the impacts of (1) (the 'ideals analysis').

The doctrinal analysis consists of analysing of the way in which specific legal mechanisms operate, and whether in practice (i.e. within a judicial action) they would be applicable, or how they might be interpreted. The source of the rights to be discussed is the Health and Disability Commissioner (Code of Health and Disability Services Consumers' Rights) Regulations 1996, known commonly as the "Code of Patient's Rights" (The Code). The Code sets out 10 rights that a patient engaging with any health professional in NZ is afforded, and highlights the principles which underpin the countries healthcare system. The rights which will be focused on are: Non-discrimination, privacy, informed consent, and negligence (or care of a particular standard, as it is referred to in the Code). Throughout the thesis, these rights will serve as the framing discussion for each substantive chapter, however analysis will largely engage with the wider common law concepts of each right. For example, Chapter 8 will discuss common law negligence, as opposed to just the formulation of the right in the Code. In doing so, the analysis aims to have the widest possible application to healthcare, to illustrate issues or strengths of the system that exist across a variety of possible circumstances, actions and situations.

The ideals analysis relies on the idea that there are some ingrained thematic principles to healthcare and its associated protections, and discussing the impact of a new technology is not limited only to its practical efficacy, but also its impact on these principles. This thesis will highlight and champion two primary principles identified through historical research and analysis of regulatory intention, on which healthcare in NZ is built. These themes are respect and trust and serve as the basis on which prospective recommendations will be made. With the identification of issues of application or interpretation identified, these principles will be

used to determine to what degree, change may be necessary to facilitate these ideals, and what form this might take.

There are two foundational presumptions to this thesis which will inform its analysis. Firstly, that the distinct features of NZ's healthcare system, such as "accident compensation" (AC)[4] puts the country in a prime position for the transition to smart healthcare. Such features will serve to facilitate a smoother transition, and result in a legal framework which is both flexible and readily adaptable to new challenges. Secondly, that NZ's healthcare system has developed and evolved to be inclusive and flexible due to its biracial history and in response to historical occurrences, which make it equally adaptable to the coming challenge. This also means that the focus on the embedded ideals of the healthcare system is critical to ensure that the continued trajectory of healthcare from any prospective changes continues this historical trend of inclusiveness. This thesis seeks to then illustrate not only the potential failings of the NZ system, but also its strengths that may be worth emulation in other sectors (or jurisdictions) when considering similar issues.

The key conclusion of this thesis is that New Zealand's current legal framework surrounding healthcare is proficient in many ways to manage the introduction of AI systems. NZ's system is historically built with an adaptability that is shown in a number of systems, such as AC, and also in a broad-application approach, such as discussed in respect of discrimination in Chapter 5, that allows the law to deftly apply to the new situations. Outside of a few specific aspects, such as causation, which will experience notable contextual difficulties, the application of existing legal tests and formulations will continue with relative ease. However, as discussed repeatedly throughout this thesis, there is an arguable position that some of these applications may be unduly harsh or unfair on those the test would typically impose liability. The specific characteristics of different AI systems may result in outcomes that warrant further debate, even if the law could readily be applied.

The main barriers for application or clear interpretation that remain after this thesis are rarely the fault of the legal framework, and instead due to the technical realities or uncertainties that surround AI technologies. For example, for those AI systems which operate

[4] Commonly known as "ACC", but for this thesis will be referred to as simply "AC" to distinguish the compensation regime itself and the body responsible for its administration – the Accident Compensation Commission (ACC); see [3.2.3.1] and [8.2] for an explanation of the concept and its application.

as a “Black-Box” there will be substantial difficulty around application of causation tests, and matters of intent will become increasingly diffuse. Additionally, the manner in which AI operates at a scale to which the law has rarely been applied will cause significant practical issues with both bringing actions against tortfeasors, but also for applying equitable and manageable remedies. And finally, the possibility for AI’s to operate internationally, transnationally, and through a variety of different corporate entities will create significant practical problems in much the same ways.

In light of these conclusions drawn, the recommendations proposed are largely aimed at the immediate short-term, to further facilitate a smooth transition towards smart healthcare. And the future research proposed is intended to focus on these broader, innate practical problems which warrant more robust, and perhaps institutional level, solutions.

## 1.2 Relevance of research

AI has been a topical focus of academia and industry since the term was coined in 1956[5] and this interest has rapidly increased in recent years due to the boom of “Big Data” tech.[6] Much of this interest has been focused on Silicon Valley and the international tech conglomerates of the United States of America (USA) such as Google, Amazon, and Facebook. With AI having a seemingly limitless range of applications – even within the limited scope of healthcare – matched only by the immense variety of academic inquiries into different facets of AI, this section will outline where in this growing field this thesis is focused. In doing so, this section will establish the relevance, and intended benefit, of this research to NZ’s developing technological future.

Unless otherwise stated this thesis is based on available research and information as of the 1st of July 2021. Due to the rapidly changing technology and legal landscape around this area, precise analysis of information could not continue indefinitely.

[5] The term “artificial intelligence” was coined in a proposal by J. McCarthy and others for a summer research project as part of the 1956 Dartmouth Conference. The history of the concept extends far earlier than this, but this is the first academic reference to the concept as it is now known.
[6] Becky McCall provides an overview of some ways Silicon Valley is utilising Big Data for health in recent years, and the rapidly growing industry surrounding it. See ’15 Ways Silicon Valley is harnessing Big Data for health’ Nature Medicine, Vol.26 7-10 (2020).

### 1.2.1 Existing literature in this area

Within NZ there is no current literature on the impact that AI might have on the healthcare sector from a legal perspective. There have been several small analyses done on the potential economic and business impacts,[7] although these too are not legal analysis but instead commercial or public focussed. The health sector itself has begun to show considerable interest in AI and its prospects, with conferences and symposiums[8] and Orion, Health Informatics New Zealand, and The Ministry of Business, Innovation and Enterprise (MBIE) beginning to conduct large inquiries into its benefits.[9]

Academic interest in AI has thus far been largely limited to the public, criminal and employment sectors, with the University of Otago's *Artificial Intelligence and Law in New Zealand* project[10] being the principal driver of this. Funded by the New Zealand Law Foundation this project aims to investigate the legal, ethical and practical challenges facing NZ due to the emergence of new technologies such as "driverless cars, crime prediction software and 'AI lawyers'". Its first major publication *Government use of Artificial Intelligence in New Zealand*[11] provided a broad overview of conceptual issues and background to the public sector and its broad interactions with AI. While this research does not specifically mention health law it does provide useful background to some of the principles on which many of NZ's systems are based and some of the fundamental inconsistencies that arise when applying conventional legal systems to situations involving AI. Otago's research also helpfully provides an analysis of the problem of defining and identifying what "AI" technology is,[12] and this thesis will aim to remain consistent with that analysis.

---

[7] The Ministry of Business, Innovation and Enterprise (MBIE) partnered with AI Forum to produce *Artificial Intelligence: Shaping a Future New Zealand* (2018) which provides an overview of research and development being done in NZ's AI sector.

[8] The annual New Zealand Institute of Healthcare Engineering's conference theme in 2021 was *Smarter Futures* as one such example.

[9] MBIE maintains a comprehensive list of its Research into this area, collected as "Digital Economy Research" which is available here < https://www.mbie.govt.nz/business-and-employment/economic-development/digital-economy/digital-economy-research/>

[10] This project is a part of the Centre for Artificial Intelligence and Public Policy (CAIPP) and was funded by the New Zealand Law Foundation. Information on the project is available here <https://www.cs.otago.ac.nz/research/ai/AI-Law/outputs.html>

[11] Colin Gavaghan and others (2019)

[12] Discussed in more in depth in Chapter Two

AI Forum is a collaborative think-thank and community forum which formed to provide the means for different developers, government bodies, and public institutions to convene on the future of AI and to try to bring about its introduction more cohesively.[13] AI Forum hosts conferences and symposiums across NZ, although focused heavily around Auckland, to facilitate debate and knowledge-sharing between these different parties on a wide variety of topics. The 2018 report *Artificial Intelligence for Health in New Zealand*[14] made brief mention of some of the policy and legal issues that AI present, although this only provided an overview of the issues.[15] The issues identified here were: bias and errors, safety, explainability, and malpractice.[16] All of these parallel the focus areas of this thesis and appear in Part C. While heavily focused on business and broad public sector discussions, AI Forum's conferences and speakers have provided some ideas which will be relied upon within this thesis. Notably, Andrew Sporgle and others have highlighted some of the unique challenges and ideas that must be considered in NZ due to the nation's obligations to Māori and under Te Tiriti o Waitangi (The Treaty of Waitangi).[17]

Internationally there has been much more work in this area, possibly due to proximity to major corporate bodies that are driving the technological developments. Much of the work available has been concerned with the regulation and integration of AI technologies into different aspects of healthcare. Some examples include the work from Colleen Flood and others at the University of Ottawa into the impacts and potential roles of AI in the health sector of Canada and the legal challenges that may arise. Flood and Régis' analysis of the issues posed by AI in *AI and Health Law*[18] provides an overview of some of the areas of concern and legal interactions which are complicated by the involvement of AI. Particularly they highlight the issues that arise in tortious liability and privacy laws under the Canadian regime. Similarly, Piers Gooding at the University of Melbourne has published extensively on the integration between AI and mental health services within Australia and the ethical and

[13] Information on the organisation and some of its sponsors and partnerships can be found here < https://aiforum.org.nz/about/>
[14] Available here <https://aiforum.org.nz/wp-content/uploads/2019/10/AI-For-Health-in-New-Zealand.pdf>
[15] This report is supplemented by two other reports that make mention of the ethics and regulation of AI generally, but this report is specific to health.
[16] At 49-53
[17] Hack Aoteraroa, NZ's 1st Artificial Intelligence in Healthcare Conference (Auckland: January 2019)
[18] Colleen M. Flood and Catherine Régis, *AI and Health Law*, in Florian Martin-Bariteau & Teresa Scassa, eds., "Artificial Intelligence and the Law in Canada" (Toronto: LexisNexis Canada, 2021)

legal challenges raised, particularly in respect of privacy and discrimination, within this area.[19]

At an institutional level, the United Kingdom House of Lords (UKHoL) released its report *AI in the UK, Ready, Willing, and Able?*[20] which investigated whether UK society at large is prepared for the introduction of AI. This report was prompted by the Theresa May governments' call to arms on the development of the UK as a forerunner in AI technologies development and use.[21] This call also specifically highlighted the ways in which AI could benefit the National Health Service (NHS) but cautioned the need for thorough investigation into the ethical and legal challenges. Healthcare did (although still quite economically focused) feature prominently in the report with discussion of the ways in which the National Health Service (NHS) was already engaging with AI development and use and the concerns that arose from such developments. The focus within this discussion of healthcare was on patient privacy and consent for use of patient data, but largely provided an overview of the concerns which need to be addressed.[22] Much like the UK, Australia has been experiencing an accelerating challenge on how to approach and best manage the regulation of AI in healthcare due to its rapidly growing development sector.[23] However so far the country has not engaged with doctrinal analysis of the interaction between AI and patient rights, and instead focused on the regulation of its implementation into the systems at large. One notable analysis in respect of healthcare rights was carried out by Richards and Hutchison when analysing whether "innovative technologies"[24] warranted a new standard or test of informed consent for patients.[25]

---

[19] An overview of the ethical and legal concerns of AI in the area of mental health can be found at: Piers Gooding, Timothy Kariotis, "Ethics and Law in Research on Algorithmic and Data-driven Technology in Mental Health Care: Scoping Review" (2021) JMIR Mental Health

[20] House of Lords Select Committee on Artificial Intelligence, AI in the UK: ready, willing and able? Report of Session 2017-19 (Published 16 April 2018), available from https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf

[21] Theresa May's speech and relevant details are cited in *Ready, Willing and Able?* At [15]-[18]

[22] In both instances, again with an economic context (the selling of data for example) but the analysis of those two rights is still applicable to healthcare generally.

[23] In 2018, the Australian government announced that AI development would be at the centre of its Government Digital Economy Strategy moving forward. Information on this and the sectors involved can be found at < https://www.austrade.gov.au/news/insights/insight-artificial-intelligence-to-drive-australia-s-economic-growth>

[24] This was not specific to AI and was more generally about new technological developments, but can be applied aptly to the discussion of AI.

[25] Bernadette Richards and Katrina Hutchison, "Consent to Innovative Treatment: No Need for a New Legal Test" (2016) Journal of Law and Medicine, Vol.23:938

Perhaps the most prolific contributor to this field, whose focus is most comparable to that of this thesis, is the work of W. Nicholson Price II in the United States of America (USA). Price has published extensively on the different issues associated with AI in healthcare, focusing on a number of different rights and different areas of medical practice. He has also engaged with specific questions of interpretation, protection and application in respect of AI itself, privacy, discrimination, and consent.[26] Price's work will be mentioned throughout this thesis however due to the immense differences between the USA and NZ systems (both health and legal) specific analysis cannot be appropriately applied. Instead the ideas and principles discussed by Price serve to provide guidance on some of the areas of concern, and approaches to potential changes, which arise throughout this discussion.

At the international and inter-government level, there has so far only been principle based discussions. No international treaties or agencies exist governing the use, development, or ethics of AI. The United Nations (UN) has in recent years begun to discuss the challenges of AI and facilitate discussion on the values and principles that best be championed in the coming years. Much of the focus at this level has, understandably, not focused on healthcare[27] and instead of the militaristic and economic impacts of AI. The use of autonomous weapons or "killer robots" has been a topical point in recent years,[28] as well as the transnational investment, development, and "fair use" of AI in economic development. The ethical considerations of the UN are useful in the context of this thesis as they provide a roadmap of ideals to ensure are maintained throughout any prospective critiques and reforms. In keeping with NZ's role in the international community and obligations under the UN or any prospective arrangements, it is important to ensure that the direction of this thesis is in keeping with the ideals presented by the UN.

---

[26] Examples of some of Price's publications which relate to this thesis are: W. Nicholson Price II and I. Glenn Cohen, "Privacy in the Age of Medical Big Data" (2019) 25(1) Nat Med; W. Nicholson Price II "Describing Black Box Medicine" (2015) 21 Boston University Journal of Science and Technology Law 347; "Artificial Intelligence in HealthCare: Applications and Legal Implications" (2017) 14 The SciTech Lawyer

[27] It can be said that healthcare that is not concerned with emergency response or the uplifting of impoverished communities is a largely domestic pursuit, and thus the kinds of analysis this thesis engages with does not arise at this international level.

[28] Most recently NZ has called for a complete ban on autonomous weapons or "killer robots" and called for the development of an AI weapons treaty. Sam Sachdeva, "NZ to push for autonomous weapons treaty" (30 Nov 2021), Newsroom, available at <https://www.newsroom.co.nz/nz-to-push-for-autonomous-weapons-treaty>

The first major step in this regard came from the decision of UNESCO's[29] General Conference at its 40th session through Revolution 37.[30] In this the member states agreed that the Ad Hoc Expert Group (AHEG) be tasked with the preparation of a recommendation on the ethics of AI the following year. Published in 2020, the *Recommendation on the Ethics of Artificial Intelligence,*[31] signed by 193 member states recognises:

> the profound and dynamic impact of artificial intelligence (AI) on societies, ecosystems, and human lives, including the human mind, in part because of the new ways in which it influences human thinking, interaction and decision-making, […] AI technologies can be of great service to humanity but also raise fundamental ethical concerns, for instance regarding the biases they can embed and exacerbate, potentially resulting in inequality, exclusion and a threat to cultural, social and ecological diversity and social or economic divides […][32]

This publication sought to consolidate the aims, concerns and recognised boundaries of the AI technologies in question, and the associated definitional and ethical issues that arise from its development and implementation. The report outlines a number of values and principles[33] which should be at the forefront of any discussion of AI at an international level, many of which are relevant to and will applied throughout this thesis. Those are[34]:

1. Respect, protection and promotion of human dignity, human rights and fundamental freedoms
2. Ensuring diversity and inclusiveness
3. Proportionality and do no harm
4. Safety and security
5. Fairness and non-discrimination
6. Privacy
7. Responsibility and accountability
8. Awareness and literacy
9. Human oversight and determination; and
10. Transparency and explainability

---

[29] Full name The United Nations Educational, Scientific and Cultural Organization
[30] Records of the General Conference, 40th session, Paris, 12 November-27 November 2019, volume 1: Resolutions (40 C/RESOLUTIONS VOL.1 + CORR)
[31] Ad Hoc Expert Group (AHEG) for the Preparation of a Draft text of a Recommendation the Ethics of Artificial Intelligence "First draft of the Recommendation on the Ethics of Artificial Intelligence" (SHS/BIO/AHEG-AI/2020/4 REV.2)
[32] At 1-4
[33] At 6
[34] At 7-15

Other values and principles focused on within this report are matters of institutional concern such as environmental impact, international cooperation, and militarisation.

The listed principles could be applied to a discussion of healthcare generally, however the report did specify the importance of healthcare in *'Policy Area 10: Health and Social Well-being'.*[35] This focused strongly on the use of big data and its associated protections, and the ways in which AI and human beings interact within the healthcare environment. The conclusion of this section is that stakeholders and governments should develop guidelines and intermediary dispute resolution mechanisms for human-robot interactions, and the prospective circumstances in which these technologies may be used within healthcare and public health generally. Guidelines and intermediary developments of this kind should ensure to champion the listed values and principles, and do so in a way consistent with the overarching goals and purposes of the UN (such as the development goals).[36]

### 1.2.2 Current regulatory landscape of AI in NZ healthcare

NZ does not currently have any legislation dealing with the regulation, definition, or status of AI. The Government Digital Services (GDS) has published the *Towards a Digital Strategy for Aotearoa*[37] which outlines the countries commitment to safe and ethical technological developments, and mirrors the language used by Theresa May when calling for the country to become a forerunner of AI and smart systems. Within healthcare there is notably a number of guidelines and advice frameworks which have been released by the Ministry of Health (MoH) to guide health professionals in their use of AI systems. These guidelines have largely served to highlight the different areas of concern and where caution should be exercised by health professionals in areas such as patient privacy, consent and bias.[38]

As of writing, there has been no judicial analysis of the interaction between AI and conventional legal mechanisms in any context. This means there is no case law which affects the interpretation of AI itself, or the way in which the legal mechanisms discussed in this

---

[35] At 20
[36] At 23
[37] (2019) Available at < https://www.digital.govt.nz/assets/Digital-government/Strategy/Towards-a-Digital-Strategy-for-Aotearoa-discussion-document.pdf>
[38] The Ministry of Health "Emerging Health Technology Advice" series and "Introductory Guidance to Emerging Health Technology" series are both available at https://www.health.govt.nz/our-work/digital-health/vision-health-technology/emerging-health-technology-advice-and-guidance

thesis should be applied in relation to AI. With this in mind the thesis will proceed on the assumption that any hypothetical judicial proceedings discussed would occur on the basis of applying the legal tests in their current formulation first.

#### 1.2.3 Role of this thesis and conclusion

This thesis aims to provide a roadmap on which future legal analysis can be based, or legal actions can apply, to help facilitate a smoother transition towards NZ's future with these emergent technologies. To do this, the thesis will strive to provide four main contributions:

1. An overview of the technology in question and its unique features which cause concern for legal mechanisms;
2. An overview of the values and principles that should be sought to protect and empower when considering any changes in respect of (1);
3. An analysis of the different issues that arise when applying current legal frameworks and protections to situations involving AI; and
4. A set of recommendations beneficial for the immediate future, to assist with the early integration and adoption of AI technologies in healthcare.

## 1.3 Overview of thesis

This section provides an overview of the different components of this thesis and any limitations or qualifying factors involved. First will be an explanation of the scope of the thesis and an explanation of the limitations imposed on this research as a matter of scale or focus. Next will be an overview of the research questions which will inform the overall thesis and its analysis to achieve the stated goals in [1.2]. Following this is an overview of the methodology employed in this thesis and the different sources of information which are utilised throughout. And finally, there is an overview of the structure of the thesis.

#### 1.3.1 Scope of research

This thesis will consider the impact of AI on New Zealand's healthcare system only, and will only engage with areas of the law in so far as they are relevant to that purpose. Each

chapter's analysis will not attempt to provide an exhaustive list of issues that may arise, and will also not attempt to provide solutions to, or prospective reforms to, every issue identified. The intention within this research is to provide an overview of the issues in question and some of the core ideas which need to be considered, and to provide a set of recommendations which will at least mitigate the most prominent concerns.

#### 1.3.1.1 Limitations and omissions

The scope of this thesis will be limited in a number of ways. These limitations will be to tailor the discussion around the key question of adaptability within the system. The following limitations will be applied:

A) The focus will be on "narrow AI"

This thesis will not include a detailed discussion of general AI, except for the purposes of defining the technologies in question in Chapter 2. General AI requires a far broader, philosophical look at issues of personhood and ethics than narrow AI and including it would confuse the overall focus of this inquiry.

B) The rights discussed will be limited to four (as opposed to the codified ten) from the Code of Patient's Rights

Attempting to give worthwhile attention to all ten rights is impractical due to the requisite scale of analysis involved. The four rights chosen are those which engage with major areas of common law and statutory protections, and are widely applicable beyond the healthcare sector as well. In this way, the analysis conducted will be of use in other contexts, even though those contexts are not themselves analysed within this research.

The broader principles of the healthcare system will be discussed in Chapter 4 to underpin the four rights chosen, and serve as a foundation on which the ensuing discussions will be based. The rights chosen, and the underlying principles, are outlined at length in Chapter 4.

C) This thesis will specifically focus on the NZ healthcare system

This will not be a general discussion of the impact of AI on medicine. Instead, this thesis will focus on the features and idiosyncrasies of the NZ system that create notable differences, with special attention to the accident compensation scheme which interacts with the conventional law. It may, however, refer to and make suggestions based on analysis of

overseas healthcare systems where it is thought these will provide guidance for NZ. Additionally, resources that provide insight into some of the concerns of AI, as well as potential solutions, that are framed against overseas healthcare systems, will be discussed and adapted to the NZ system. If they are unable to be adapted appropriately, this can provide potential insight into reforms for the NZ system to mirror the overseas possibilities.

#### 1.3.1.2 Research Questions

To fulfil the purpose set out in [1.2.3] the research questions in this thesis are:

A) Whether the involvement of an AI system impacts on the current obligations and responsibilities owed to patients under NZ law; and
B) Whether a patient treated by an AI instead of, or in combination with, a human healthcare professional, is protected to the same extent as a patient treated solely by a human healthcare professional, and whether any additional protections are needed due to the involvement of the AI; and
C) Whether, in light of the findings in 1) and 2), the current NZ healthcare system requires reform to appropriately respond to the introduction of AI, and if so, what form this should take.

### 1.3.2 Methodology

#### 1.3.2.1 Doctrinal research

Part Two will provide background information on AI and the NZ Healthcare system which will frame the later chapters of discussion. The two main sections to the background will be a doctrinal discussion of what "AI" is and ways in which it is currently, or foreseeably, going to enter use within medicine (Chapter 2) and an overview of the history, structure and principles of the New Zealand healthcare system as it currently is, to establish the scope of AI's potential impact (Chapter 3). Both will involve doctrinal research to analyse and describe existing knowledge.

Part Three will discuss the key principles that unpin the NZ healthcare system contained within the Code. It will focus on the principles contained within the Code and associated legislation, and how these are identified and refined down to the key themes of the thesis.

Chapter 4 (Code) and the subsequent rights chapters (5-8) will use doctrinal research to analyse the Code, its themes, mechanisms and how they are applied within medicine. This doctrinal research will also be used to analyse the way in which accident compensation will apply to situations involving an AI, and the application of this scheme.

#### 1.3.2.2 Analysis of primary sources

The principal focus of this research will be an analysis of the existing frameworks in the NZ healthcare system. These include, but are not limited to, the Code of Patient's Rights, Health Competency Assurances Act 2003, the Health Information Privacy Code 1994, the Health Act 1956 and the Accident Compensation Act 2001. The intention of this research is to identify the shortcomings of these existing primary sources of law, and the common law that has been developed alongside them, in the wake of what is understood by, and has been seen in the current uses of AI.

In Part Two, Chapter 3 will analyse the legislation and Codes that establish the NZ healthcare system, and lay the foundation for all the mechanisms or rights within it. This will provide a robust roadmap of all the areas and potential issue areas that will arise throughout the rest of the thesis.

In Part Three, Chapter 4 will discuss the legislative authority and basis of the Code of Patient's Rights. The succeeding four chapters will analyse specific aspects of the Code by reference to its language and to disciplinary proceedings under the Health Practitioners Disciplinary Tribunal that enforce the Code.

#### 1.3.2.3 Comparative research

Intermittent throughout the thesis, but primarily focused in the recommendations and reforms section in Part 5, will be comparative research from a number of different international sources. These sources will only be compared where appropriately applicable, and will serve to provide insight into the possible trajectory of the NZ law and situation. For example, NZ tort case law often draws from the UK and therefore recent developments in that jurisdiction can provide insights into the sorts of questions or developments which NZ may wish to emulate (or avoid).

#### 1.3.2.4 Empirical research

Due in part to the lack of available literature, and the fact that healthcare is intimately related to the public and individuals, I conducted a survey to provide empirical data for analysis. This provides unique primary data for analysis on the opinions of those who will potentially interact with AI technologies and to gauge their understanding of, or concerns of, the technologies. An explanation of the importance of this in relation to the thematic principles of the thesis, namely trust, is provided in Chapter 4.

#### 1.3.2.5 Three Scenarios of Application

To investigate the potential impacts of AI in practice, Chapters Five to Eight will each engage with three scenarios of application after their doctrinal analysis. Each chapter will include a specified framing situation which is altered by the three scenarios, discussed below.

These scenarios are designed to cover a broad range of potential situations which can arise from the use of AI within medicine. All three are currently being tested or developed in different parts of the world, although their prospective timelines of integration vary greatly. Each involves a different level of "independence" for the AI to operate within, and the involvement of pre-existing health system components in diagnoses, or the administering of treatment become increasingly removed. The inquiry will focus on these scenarios in their most general, "common world" form, and there will be discussion and attention paid to specific details or features that might alter the outcomes involved. The three scenarios, and notable sub-categorizations, are:

1. An individual goes to a doctor, who utilises an AI tool in their treatment ("SN1")

In this scenario, the primary treatment is performed by a doctor, although the doctor is aided by the use of an AI tool. There are a number of potential sub-scenarios within this, with the main two being:

- SN1A: where the AI is the sole tool used; and
- SN1B: where the AI is one of many tools used, with the others being non-AI tools.

SN1 will largely focus on the process of diagnosis and treatment decision-making; it thus includes all associated components of these steps, such as the collection and storage of patient data used in the treatment.

2. An individual is treated only by a machine (“SN2”)

In this scenario, the treatment is carried out by a machine, without the involvement of a human in the diagnosis or treatment decision-making. The key difference with SN1 is that the AI is not acting in unison with a human agent but acting independently of them. This results in two major sub-scenarios to discuss:

- SN2A where a human agent is present but not participating (in an oversight capacity); and
- SN2B where a human agent is not present (the AI is acting entirely independently).

The former will raise issues of the meaning of ‘treatment’ and what it means to be ‘treated’. Specifically, it will consider all three stages of a medical inquiry: diagnoses, treatment prescription, and actual treatment (presuming some form of physical action is required on the part of the “treater”). The latter is not commonplace, although mechanisms to enable this are currently being tested in several jurisdictions and will be the principle focus of this scenario to distinguish it from SN1.

3. An individual uses an application or online AI tool as a substitute for visiting a doctor in person. (“SN3”)

In this scenario, the individual is diagnosed and treated remotely, through an online app. This final scenario is the most removed from conventional healthcare (although as discussed later is a common suggestion for the future of smart medicine). While both SN1 and 2 engage with patients in conventional spaces – clinics and hospitals – SN3 shifts healthcare into wider society where protections granted to someone receiving medical treatment apply less clearly. It should be noted that where the application was sourced will result in a difference to what legal principles apply: a government or institutionally provided application may result in conventional medical protections, a publicly sourced (or private) entity’s application will most likely not.

### 1.3.3 Structure

This thesis is divided into three unequal parts, divided by content and intention. The first part, Part A (Chapters One to Three) is focused on providing necessary background and contextual information for the later substantive inquiry. Chapter Two outlines the technology

at issue and aims principally to highlight the difficulties in both accurately defining, and clearly identifying the technology in question. Attention will be given to the reasons why a generally accepted definition of AI is both difficult and controversial within academia, and how this impacts inquiries such as this in practice. It will provide an in-depth discussion of two main categories of AI; narrow and general, and justify the focus in this research on narrow AI. This chapter will provide a robust outlining of exclusions and scope for the technology, so as to limit the wider thesis for simplicity. A comprehensive discussion of AI itself would require more time than this thesis provides, even when focusing on a singular area like medicine. Chapter Three will then outline the healthcare system in NZ as it operates currently. It will provide a contextual history to the system to establish some conceptual principles which underpin the system to influence the substantive discussion. This will be followed by an overview of the major regulatory and functional components of the system, including Accident Compensation (AC) and the Health and Disability Commissioner (Code of Health and Disability Services Consumers' Rights) Regulations 1996, which will be referred to as "the Code." It will briefly include a discussion of the regulation of those who operate as part of the healthcare system, known as "healthcare professionals" although for the purposes of this thesis, these will largely be referred to as "doctors" unless specified. Importantly, the NZ healthcare system is undergoing extensive reform in 2022 under the Labour government.[39] While these reforms will be mentioned where relevant, they are not themselves thoroughly analysed or applied. The focus of this chapter is largely on legal rights contained within the system, as opposed to the structure of the system itself, so this omission will be of little importance in practice.

Part B (Chapters Four to Eight) begins with a discussion of the principles of the health regime in NZ in Chapter Four, and provides the "Code" as the main representation of these principles. This chapter acts as a bridging chapter, providing both important context and framing to the thesis, and also engaging with the doctrinal concepts which underpin this part of the thesis. It will also provide an explanation as to the selection of topics for the following chapters, as not all components of the Code can be given adequate attention. These topics are selected by synthesizing the ten rights contained within the Code into central themes, or issues, which can be readily compared to broader legal concepts. This chapter will also

[39] Announced in April 2021, the official press-release is available here https://www.labour.org.nz/news-health-reforms-2021-system_that_works_for_everyone

introduce the persuasive theme of this thesis; the importance and role of public trust in both healthcare and its regulation. To do this, the results of the conducted survey will be analysed, in which participants identified their own concerns about the emergent technologies and their expectations.

The subsequent four chapters analyse substantive areas of the law, to identify in what ways their application is affected by the introduction of AI technologies. These four chapters are: non-discrimination and bias, privacy, consent, and negligence. Each of these chapters will include a brief explanation of these concepts' application and functions, and then discuss the issues that arise once an AI is introduced to conventional situations. Each of these chapters will also include an analysis of what legal outcomes are likely in the three scenarios discussed, to identify areas that require reform or perhaps are sufficient.

Chapter Five will discuss non-discrimination and bias and the issues inherent to AI's nature that make conventional formulations of these rules difficult to apply. Non-discrimination rules have developed in response to human prejudices, and the alleged capacity of people to manage or "put-aside" these precepts. The way in which AI operates challenges these options, and the tension between non-discrimination and AI results in a discussion of which is considered most important.

Chapter Six discusses the issues associated with privacy when including AI systems into healthcare, due to the way they aggregate information, often transnationally, and utilise this information at large. The issues identified here are not only problematic for the conventional application of the Code and the Health Information Privacy Code 1994 (HIPC), but also present practical difficulties due to jurisdictional and identification problems.

Chapter Seven focuses on the issues for informed consent as a result of the involvement of AI within treatment and diagnosis. Legal consent is largely centred on the idea of knowledge and conveyance of information, which this chapter will highlight is not only an impracticality but also often an impossibility. Informed consent has the potential to limit the effectiveness of these AI systems, by preventing their use and eroding public engagement with the healthcare system. This chapter will also briefly discuss the potential for a different standard of consent to be applied to experimental treatments (and by extension AI) that has been proposed in Australia.

Chapter Eight focuses on the area of negligence, by first providing a comprehensive discussion of the role of ACC within NZ. ACC has the potential to circumvent the commonly discussed issues of AI-doctor liability in healthcare by barring potential civil actions that may arise. After identifying how ACC works, and whether or not the three scenarios themselves would qualify for cover, the chapter will focus on the issues that may arise for applying common law negligence to these situations. The three requirements of negligence; a duty of care, breach of that duty, and causation for the breach, all encounter significant problems when introducing an AI.

Part C (Chapters Nine and Ten) contains two chapters: reform and recommendations, and the conclusion. The former will provide an overview of some potential ways in which the identified issues in part two could be managed, with reference to proposals that have arisen both domestically and abroad. These proposals will not be limited to medico-legal ideas, but will draw on the broader discussions of AI and the law. This chapter will also recognise the areas of the law which are adequate in their current conceptions, and may offer existing solutions without the need for comprehensive reform. The final chapter, Chapter Ten, will provide a conclusion to the overall inquiry and reiterate the areas of concern. It will also identify future areas of research that are necessary which could not be conducted within this thesis, or were previously identified as important but largely omitted (such as product liability).

Throughout these chapters, other issues associated with the use of AI in healthcare will become apparent. For example, issues of medical registration, licensing and disclaimers. These issues will not be discussed in-depth or engaged with and will only be referenced for completeness. Any omitted ideas such as these will be outlined in Chapter Ten when concluding on future research that is warranted.

# Part A

# Background

To analyse the impacts a new development may have on an established system, it is necessary to first establish the functionality and boundaries of the system involved and to understand the nature of the development. This part will provide the necessary background and definitional information on which the remainder of the thesis is built.

Chapter 2 outlines the technology in question, Artificial Intelligence, and discusses its functionality, its recognition within the law, and finally the dominant issues associated with it.

Chapter 3 provides an overview of the healthcare system in New Zealand, focusing in on the components which are most relevant to a discussion of patient rights. Alongside this structural overview, the chapter provides a brief history of the system and its genesis, to illustrate the core thematic principles of the system. In highlighting these principles, it aims to provide a foundation on which the subsequent discussion of impact will occur and a guide for potential goals of any proposed reforms.

# Chapter 2: Artificial Intelligence in Context: Definition, Issues and Considerations

---

> "Nobody phrases it this way, but I think that artificial intelligence is almost a humanities discipline. It's really an attempt to understand human intelligence and human cognition."
>
> Sebastian Thrun[40]

## 2.1 Introduction

As technologies advance and develop, the ethical, legal, social and cultural issues that arise become more identifiable and often more complex. Artificial Intelligence (AI) reaches into the realm of science fiction in a way that few technologies have done before, and functions in a way that remains obscure even to those involved in its development. To properly begin to understand the issues created by AI, it first must be established what AI actually is. This chapter will strive to generate a workable definition of AI that will be relied on in the remainder of this thesis. Secondary to this, the chapter will analyse how AI's "nature" causes both theoretical and pragmatic issues for conventional legal mechanisms and reasoning. While this secondary aim will be laid out and identified, an attempt to rectify or reconcile the issues at play will not be possible within the scope of this thesis. The principal issue of focus within this chapter will be with what is known as the "Black-Box Problem" (BBP); a conceptual problem with the ability to interpret information, or ascertain reasoning, behind AI decision-making. From the outset it is acknowledged that not all AI systems operate as a black-box, and in fact significant work is being done on "verifiable" AI systems. However, much of the discussion in respect of legal problems and applications will focus on situations

[40] German Computer Scientist and CEO of Kitty Hawk Corporation during the Google Ethics of AI Symposium (March 2017)

involving black-box systems, as they are the ones with the greatest potential for issues arising.

### 2.1.1 Difficulties in defining AI

AI is a term often used in science fiction, media, and academia with little clarity on the specifics or boundaries of the term. Studies have suggested that public understanding of AI is broad but "not deep" and that in general there is a lack of real understanding of the term.[41] This creates a problem for potential legal discussion of AI as an unclear definition creates difficulties for analysing the potential impacts of AI, in identifying shortcomings of the law and therefore drafting future regulation. In addition, discussing legal hypotheticals without clear boundaries or limitations means that any conclusions drawn are easily attacked or dismissed in practice, which limits or defeats their usefulness. There is debate around whether a clear, concrete definition of AI is a necessity for legal discussions, or legal regulation. It is likely that a sufficiently broad definition that can be applied to a variety of technologies and circumstances is more appropriate than a restrictive-but-definitive definition of AI systems. The immense variance of what qualifies as AI, discussed in [2.2], will make clear that attempting to finalise an appropriately specific definition is likely a fools errand. Instead, this thesis will opt for a flexible definition, albeit one with clear boundaries on how advanced the discussed technology is. It will also become clear that due to the conclusion that NZ's law is adequate to manage the introduction of AI already, there is less urgency on a clear and precise definition at this time.

"Artificial Intelligence" thus far has no statutory definition within New Zealand, indeed, in any other country.[42] A difficulty associated with defining AI is that oftentimes it relies on a comparison to human, or organic, intelligence. However, definitions of organic intelligence are similarly contested within the scientific community, making a comparative definition equally as unclear.

---

[41] Chris Holder, Vikram Khurana and Mark Watts, Artificial Intelligence: Public Perception, Attitude and Trust, (2019) Bristows, London, at 5

[42] A number of countries have begun to release reports on the potential impacts of artificial intelligence, as well as usage guidelines, often times including a section discussing a definition. However, so far none of these have resulted in statutory changes.

A problem that will become apparent throughout this chapter is that many definitions of AI presented in literature vary depending on their source.[43] Compounding this problem is the immense variety and depth within AI as a concept, which often is not appropriately addressed or signalled in literature. This can result in a definition of AI being given that is only intended to cover one specific use or configuration of the technology, but is presented as an overarching definition. Conversely, definitions can be too vague as to encompass technologies that present no real legal concern due to how "simple" some AI can actually be.

This chapter is not intended to generate an authoritative position on how AI should be defined, and instead only aims to illustrate the specific parameters in which this thesis will operate. A normative approach will be taken to synthesize a definition for use, after establishing the specific parameters which can be considered relevant. Since 2017, a number of state law commissions, judiciaries, and university research centres have engaged with the discussion of AI and provided their own definitions. These will be used to form a representation of the status quo in academic thinking, and therefore the norm. In doing so, the definition derived by this chapters' conclusion will be intended as a holistic synthesis of the norms present across legal academia.

## 2.2 What is Artificial Intelligence? A working definition

Defining AI is a task worthy of its own extensive thesis; sciences, life sciences and humanities have grappled with this issue since the term was first coined in 1955.[44] At its most generally accepted, AI is used to refer to computational programmes that simulate intelligent (generally considered 'human') behaviour. To phrase it another way, 'AI' is used to refer to

[43] For example, see Matthew Scherer "Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies" (2016) Harvard Journal of Law & Technology, Vol.29 (2); S. Jessica Allain, "From Jeopardy! To Jaundice: The Medical Liability Implications of Dr. Watson and Other Artificial Intelligence Systems" (2013) L.L Rev Vol.73(4) for some definition discussions by academics in this area

[44] John McCarthy and others, "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence" (31 August 1955), at 1: Accessed at <http://raysolomonoff.com/dartmouth/boxa/dart564props.pdf> on 5 February 2018

technologies that allow outputs that would generally require "intelligence" if they had been performed by a human.[45]

The United Kingdom House of Lord's (UKHoL) report *Ready, Willing and Able?*[46] is perhaps the first major executive analysis of this issue, and concluded:

> "There is no widely accepted definition of [AI]. Respondents and witnesses provided dozens of different definitions."[47]

However, despite this disparity, there is a loosely accepted array of factors or requirements associated with what can be called an "AI".[48] While this is not exhaustive, it does help to narrow the otherwise incomprehensible umbrella of the term. An AI must generally display the following forms of behaviour that are commonly associated with human intelligence:

- planning;
- learning;
- reasoning skills;
- problem solving;
- knowledge representation;
- perception of states;
- motion;
- manipulation.

In addition, although much less common – social intelligence and creativity are also mentioned.

AI itself is not a singular technology, or even a singular type of technology. The list of factors identified above indicate that AI is an umbrella term referring generally to types of technology that are vaguely associated through their outputs. In this way, it is helpful to think of the term "AI" analogously to the term "vehicle"; a boat and a car are not the same kind of transportation, but they both fall under the broader umbrella definition of a vehicle. In the same way, a famous chess playing AI system (DeepBlue) and a facial recognition AI utilised

[45] New Zealand AI Forum Artificial Intelligence: Shaping a Future New Zealand, (2018) at 14
[46] House of Lord's Select Committee on Artificial Intelligence, above n20
[47] At [9]
[48] Artificial Intelligence as a concept can be used as a term for the umbrella concept, or as a plural reference. "An AI" is generally the use when discussing a singular system or device.

by police both function in very different ways, with entirely disparate purposes, but both fall under the definition of AI.

In order to form a working definition for the purposes of this thesis, it is necessary to explore the scope of AI in more detail.

### 2.2.1 Categories of AI

#### a) General vs. Narrow AI

AI can first be divided into two main categories beneath the umbrella definition – general and narrow AI. The former is what is often represented in popular culture and science fiction media; a machine capable of thought and planning, indistinguishable from a person. General AI is still considered a pipe-dream,[49] and while many predictions exist for its occurrence within the next century[50]; processes and technologies have not sufficiently advanced to make these timeframes realistic and in fact some have gone so far as to suggest they will never develop.[51] General AI is therefore not within the scope of this thesis, although it might be referred to as a comparator. Narrow AI is that which currently dominates the technology market, and will likely dominate it for the foreseeable future, due to its widespread prevalence in both everyday technologies, and institutional uses. Even some of the most lauded AI, such as self-driving cars, are still in their commercial infancy and will be unlikely to dominate wider commercial markets for the immediate future. As a result, from herein, AI will be used to principally refer to Narrow AI. Narrow AI is an AI system that is focused on one specific task or functionality. These systems operate within specific pre-determined rulesets, utilising specific data-sets, to achieve one task, such as natural language processing in virtual assistants like Siri or Google Assistant. Unlike general AI, which is a machine that exhibits "human-like" intelligence and can perform *any* task, a narrow AI cannot function beyond its single purpose without direct alteration of its design.

#### b) Reactive machines, limited memory AI and theory of mind AI

[49] Margaret A. Boden "Perspective: AI – Utopia or Dystopia?" (World Economic Forum) 2019
[50] Vincent C. Miller and Nick Bostrom Fundamental Issue of Artificial Intelligence (Synthese Library, Berlin: Springer, 2014) at 11-13
[51] Ragnar Fjelland "Why general artificial intelligence will not be realized" 2020 Humanities & Social Sciences Communications at 2

Beneath this Narrow/General distinction, there is another loose separation of narrow AI's into what are called "reactive machines" or "limited memory AI" – the two most common in the current market.[52]

Reactive machines (RM) are designed for singular, specific tasks in which a large database can be drawn from that is prepared in advance. These RM do not create their own databases of experiences, and instead apply static knowledge and attempt to filter out sub-optimal outcomes to achieve their result. The aforementioned DeepBlue and similar game-playing AI AlphaGo are perhaps the most famous examples of this type of machine.

Limited Memory AI (LMA) are systems capable of relying on historical information to better prepare or extrapolate. This information however is not retained long term and not used to further develop, meaning its capacity to continue learning is very limited. The limited memory is used to interpret the environment or circumstances the system is exposed to, beyond the capabilities of a reactive machine, but this interpretation is then lost.[53] Many of the technologies used within medicine are currently some form of LMA, but this is not absolute. Within healthcare, an MRI that stores and identifies changes in a patients scans to analyse patterns for cancer diagnosis would be an LMA system.

A third form commonly discussed is called "theory of mind". This third form is discussed principally as a theoretical research area, and not a practical, available technology. As a result, theory of mind systems will not be discussed within this thesis.

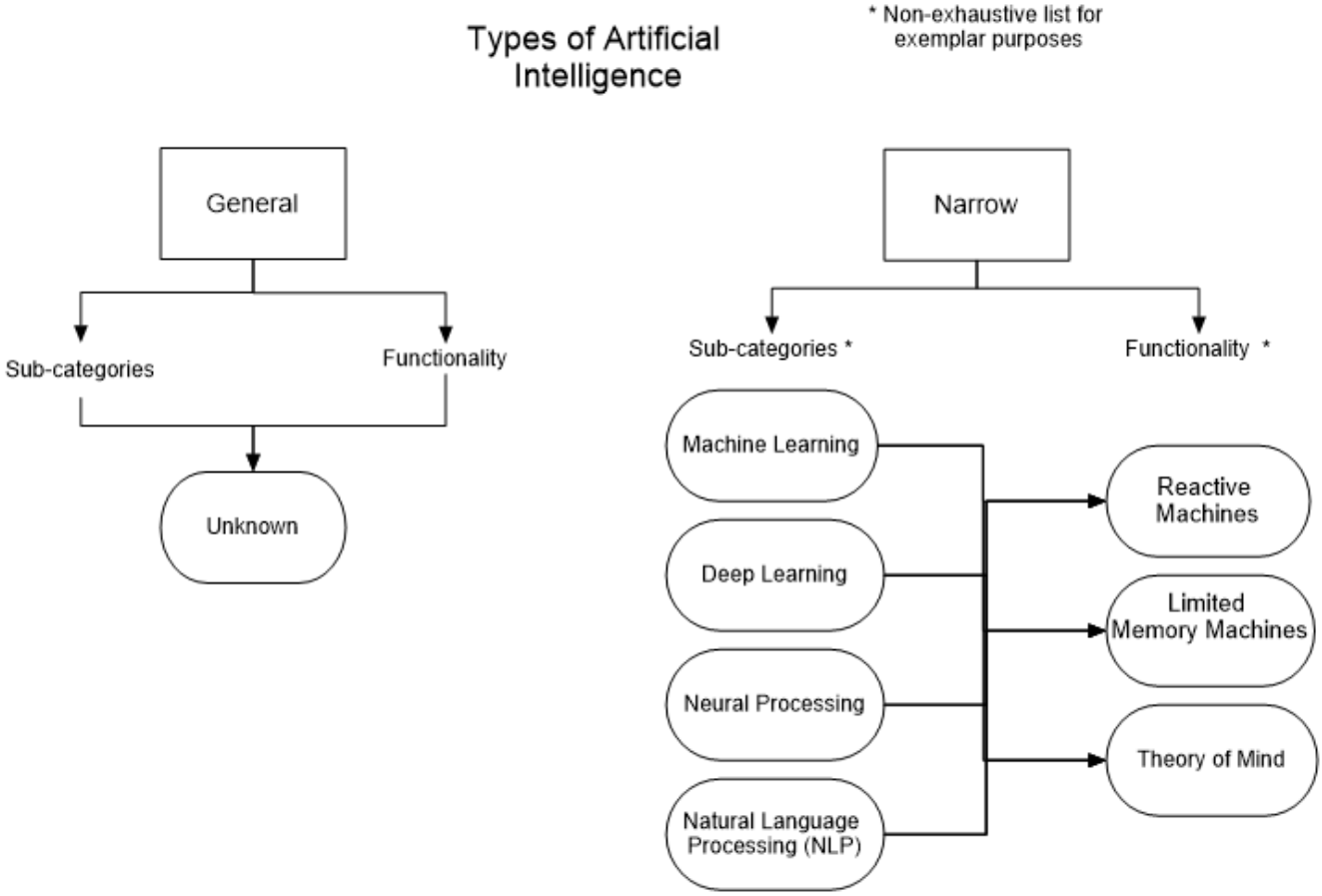
These three forms are not the only forms of AI that exist, but they are the ones individuals are likely to have interacted with before, and many of the innovations or speculative machines discussed later in [3] fall into these categories.

c) Conclusion

AI is therefore a broad umbrella term that encompasses a number of different technologies, and a broad range of disciplines or approaches to the concept of intelligence. A

[52] While these distinctions are not specific to Narrow AI, they are generally only discussed in the context of Narrow AI due to practicality.
[53] CMS Law AI in Life Sciences: Legal perspectives on the opportunities and challenges of AI for life sciences companies (2019) at 4-5

number of these sub-fields are often spoken of interchangeably with AI as a whole.[54] Common sub-categories of AI that are often represented as the whole include, but are not limited to: machine learning, deep learning, neural networks, natural language processing, and computer vision systems. Machine learning is perhaps the most rapidly developing area of AI, and the one in which confusion around terminology or distinction arises the most as well. These specific types will not be of much focus within this thesis; instead AI will be treated as a homogenous block – the exact process of their function being less important than their output or purpose. To summarise this brief overview, a diagram has been provided for a basic outline of AI.




[54] See Kristian Kersting, "Machine Learning and Artificial Intelligence: Two Fellow Travelers on the Quest for Intelligent Behaviour Machines" (2018) Frontiers in Big Data Vol.1(6) for a discussion of the distinctions, and brief discussion of conflation of the two in media.

### 2.2.2 Algorithms and AI

One, often confusing, disparity worth addressing is the relationship between the terms AI and "algorithms." It is becomingly increasingly common to see governments or institutional bodies release guidelines for the use of "algorithms" and the problems or cautions associated with their use. However, it is often unclear whether the terms AI and algorithms are being used interchangeably, or as distinct technologies.

Algorithms are best thought of as instructions; they are the pre-determined, rigid, recipe that a system follows or executes when triggered. This means there is a defined input, which leads to a defined (and expected) output. This is a specific programmed behaviour, or journey, from A to B that emulates formulaic decision-making. A very simple example of a non-AI algorithm are quick-sorting functions in software like Microsoft Excel, for example the function to alphabetize a list. This executes a pre-determined ruleset to determine the order of a specified dataset.

Algorithms can be categorized as "complex" when implementing a wider range of rules and calculations, and often involve numerous layers of input and output that interact. Whilst extremely reductive, AI instead can be considered as a group of algorithms that interact together, or in sequence, to arrive at their outcome. Within this category, AI can be either adaptive or locked; the former is capable of modifying their own algorithms, datasets or processes, whereas the latter are set to a restricted and highly specific method of functionality.

Adaptability and "changing" is not ubiquitous amongst AI and is, in reality, a capability of very few systems. This thesis will focus on such systems however, because these are both more likely to operate with a number of the issues discussed in the next section, and also more likely to function in ways that create evidentiary, causative and identifiable issues with established legal doctrine. It must be noted that any comments or conclusions throughout this thesis that relate to an AI's "nature" as something changing, opaque, or in some other way separate from the control of a human agent can likely only be applied to such adaptive AI systems. Locked AI systems are, at least in the kinds of scenarios discussed within this thesis, more likely to be akin to any other existent technology system.

This ability to change is its "intelligence." In this case, the outputs are not defined, but the intention of the output is designated through the complex curation and mapping of data and design. Each output then multiplies this process, resulting in further permutations of output. This more closely resembles the concept of human intelligence by AI interacting with an adjusting of its output based on response and environment. If an algorithm is the recipe, then an AI is the act of cooking; recipes are used and called upon, but changed, altered, abandoned or modified based on a wider variety of needs or circumstances.

Many of the guidelines released by the NZ government in respect of algorithms will be referred to throughout this thesis, as they are often the only reference in NZ to a technology that might look like AI. However, it is important to note that algorithms often fall short in terms of both complexity and capability of even a basic AI system, and therefore these guidelines will inherently also fall short in addressing issues that may arise.

### 2.2.3 Definition Relied Upon

The discussion above shows that defining AI and distinguishing it from processes like algorithms can be complex. However, for the purpose of enabling the discussion within this thesis, a clear definition is required. It should be emphasised that the definition used here is not intended to be exhaustive or authoritative, and is chosen in recognition of the complications of providing such an exhaustive or authoritative definition. However, as NZ currently lacks any major legislative or government thesis into AI and its impacts, a definition needs to be sought from elsewhere.[55] As a matter of convenience, this thesis will rely on a definition given by a high authority in an appropriately similar jurisdiction. The United Kingdom House of Lords (UKHOL) is the only appropriate organisation to have conducted a large-scale AI thesis to date.[56] Its definition is specifically for narrow AI, as it considered that general AI is too distant and complex a discussion to conduct simultaneously with narrow AI.[57] *Ready, Willing, Able?* relies on a definition previously established in the

[55] There have been a number of University inquiries into AI, which provide their own definitions, see Colin Gavaghan and others, "Government Use of Artificial Intelligence in New Zealand" (New Zealand Law Foundation, Otago University, 2019).
[56] House of Lords Select Committee on Artificial Intelligence, AI in the UK: Ready, Willing, Able?, Report of Session 2017-19 (16 April 2018)
[57] At [16]

UK Industrial Strategy White Paper,[58] and could be described as the most "average" of AI definitions.[59] Under this definition, AI refers to:

> Technologies with the ability to perform tasks that would otherwise require human intelligence, such as visual perception, speech recognition, and language translation.[60]

The additional requirement applied to this definition by the UKHOL is that "AI systems today usually have the capacity to learn or adapt to new experiences or stimuli".[61] This definition closely aligns with that given by the NZ AI Forum, a non-profit organisation dedicated to promoting the understanding, adoption, and integration of AI within NZ,[62] which defined AI as:

> "…advanced digital technologies that enable machines to reproduce or surpass abilities that would require intelligence if humans were to perform them."[63]

As remarked by the University of Otago, the AI Forum definition is potentially too loose to meet a regulatory target,[64] so the UKHOL definition, which could be argued to be more defined but along the same reasoning, will be adopted. The additional requirement imputed onto this specifically by the House of Lords is that the AI is capable of, in some way, learning or changing its skills based on experience.[65] This component will be a critical focus within this thesis, discussing the way in which an AI can act differently over time, and the unpredictable nature of this.

The definition of AI used for this thesis is therefore:

> **A technology with the ability to perform tasks that would be considered to require human intelligence, such as visual perception, speech recognition, and**

---

[58] Department for Business, Energy and Industrial Strategy, Industrial Strategy, Building a Britain fit for the future (November 2017)
[59] For example, see Matthew Scherer "Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies" (2016) Harvard Journal of Law & Technology, Vol.29 (2); S. Jessica Allain, "From Jeopardy! To Jaundice: The Medical Liability Implications of Dr. Watson and Other Artificial Intelligence Systems" (2013) L.L Rev Vol.73(4) for some definition discussions by academics in this area
[60] At 37
[61] House of Lords Select Committee on Artificial Intelligence, above n56, at [11]
[62] AI Forum could reasonably be considered the closest thing to a "government" definition so far, due to its close working relationship with the NZ government.
[63] AI Forum Artificial Intelligence: Shaping a New Future New Zealand, (2018) at 26
[64] Colin Gavaghan and others, above n55, at 5
[65] House of Lords Select Committee on Artificial Intelligence, above n56, at [11]

**language translation. These technologies have the potential to learn and adapt to new experiences and stimuli, which enables their performance to reproduce and perhaps surpass human intelligence and capabilities.**

## 2.3 The Issues associated with AI

AI has only recently transitioned from the realm of pure science fiction to reality, and with it many unforeseen and complex issues have arisen. The problems that stem from the use of AI largely fall into four categories: interpretation; control; design; and personhood. These will be discussed in turn.

### 2.3.1 The issue of interpretation: The Black-Box Problem

A problem within computer science is known as the "Black-Box Problem" (BBP). This refers to the fact that the processes of an AI seem like a "black box"; the reasoning or processes involved in decision-making cannot be seen or readily analysed. This creates a new layer of difficulty in discussing problems with regulation, liability, and even sometimes definition because it results in a decision-making system that is opaque to later analysis.

The black box is called a problem, not a feature, because the algorithms used in these AI are unavoidably opaque, as opposed to being opaque by a matter of design.[66] This is primarily seen as a by-product of deep learning in modern AI systems. Deep learning relies on a "family" or network of algorithms that facilitate deep layers of unsupervised learning to fulfil tasks and also adapt to new tasks and circumstances autonomously. An example of the use of deep learning is photo or video recognition; a combination of convolutional neural networks, and long short-term memory networks, can identify what is occurring in a photograph in a way representative of human visual recognition. Because of the way these deep learning algorithm families operate on a multi-layered approach, it is impossible (or at least

[66] Scherer, above n59, at 370-372 when discussing opacity; also see Joshua Kroll and others, Accountable Algorithms, (2015) 165:633 University of Pennsylvania Law Review, at 650-690 when discussing transparency and black box identification problems.

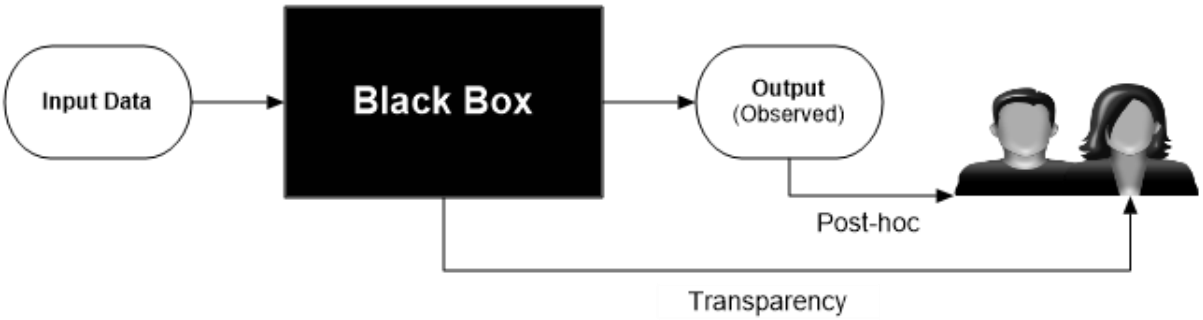
immensely difficult) to identify the factors which lead to the decisions, or the identification of the image.

Examples of the BBP within medicine are becoming increasingly common; while a doctor may be less accurate with their diagnoses than an AI, they would at least be able to explain the reasoning at each stage of their thesis as to why they came to that diagnosis. A deep-learning algorithm might have far better diagnosis rates, but there would be no way to determine why it reached its conclusion. As a result, learning from the decisions made from an AI is a moot point, and therefore issues of causation become clear. Medicine is a primarily communal knowledge system that relies on the sharing of diagnoses, situations, treatments and relies on an element of trust between doctor and patient. While machines might be able to achieve positive results, the general public might not trust a machine-exclusive system of treatment being employed. But a machine-doctor tandem system is not viable when the human component cannot learn from, or advance, with the machine's use. Without the ability to learn from the benefits provided by the AI, the technological capabilities will advance faster than their human counterparts, shifting AI medicine into a situation where the technology are more akin to prophets than engageable resources.


Work is being carried out by researchers to create what are dubbed "accountable algorithms" in an attempt to not only resolve the BBP, but also to create systems that do not result in the problem arising at all.[67] Kroll and others identify a few methods being developed, particularly the work of Zemel et al. to create "fair synthetic data" which assigns data into clusters which represent portions or examples of some feature of the total data set.[68] By doing this, it creates a series of "sign posts" for identification later, which can assist researchers in avoiding treating all systems as black boxes. This is considered a form of "transparency AI" model. An alternative method has been to develop AI systems that solely attempt to deCode the decisions of *other* AI systems, dubbed ad-hoc networks. Similarly, the technique called "regularization" has led to "many useful modifications to standard tools" by introducing a

[67] Kroll and others, at 634-640
[68] At 688

form of penalty linked with undesirable behaviour, or model attributes, in a system.[69] These explanations can occur at two different points in the process, depending on how the decision is made, and which form of control is implemented. A simplified explanation of two forms of accountability, showing the two differences in approach as to when explanation should occur, is displayed below:




This yields a "generalizable" range of models that can be treated and tested in similar ways and can create parity within alike systems.[70] There are a number of other methods discussed by Kroll which in many ways make headway into limiting the effects of problems like the black box, or at least allowing more consistently accountable algorithm systems which can be verified and "trusted."[71] However, at the current stage of development, the BBP remains a real and prevalent issue – and a serious one for the discussion of regulation and liability. As a result, throughout this analysis attention will be given to the problem of verification of information, or liability in lieu of an absence of verifiable decisions. Some of the techniques and approaches discussed by Kroll will be drawn on as potential solutions, however for the most part the analysis will focus on accepting the problems' existence and the potential to reconcile this within the law.

### 2.3.2 The issue of control

This issue is focused on the issues that arise once an AI is acting in some way – the control and understanding of an AI's outputs create another layer of issues for liability. Scherer separates these problems into three main categories: autonomy, foreseeability and

[69] At 688
[70] At 688-689
[71] Some of which will be drawn on at a later stage in this inquiry, however not discussed to the depth of Kroll and others' work.

causation, and then "control" overall.[72] AI are already showing the ability to develop investment portfolios, make medical diagnoses beyond those requested, and drive vehicles without supervision.[73] The increasing variety and independence of AI will cause unprecedented problems for the legal sector as issues of autonomy and liability come to a head in the future; how can an individual be blamed for the actions of an AI that acts autonomously in its decisions? But if you cannot blame the person, how can you then punish the AI? This issue then coincides with issues of foreseeability and causation; if you are not supervising the decision-making process of the AI, it can be almost impossible to predict how it is going to act (particularly in black box situations), and therefore be able to identify causative links between actions. This becomes especially apparent within NZ's tortious system, where the standard of causation is primarily known as the "but for" test. This asks the question whether the consequence of the action would have occurred but for the defendant's conduct.[74] The but for test also incorporates a number of elements of foreseeability, and in most cases (excluding strict liability) some element of awareness of the result.[75] There are other elements of causation within NZ law, like standards of proximate cause, independent sufficient causes, and concurrent causes[76] – however all of these in some way require an element of foreseeability, at least to remove the situation from a "true" accident.[77] Scherer discusses the concept of "satisficing" decisions being made by statistical AI – meaning decisions that are a satisfactory solution as opposed to an optimal one – often resulting in decisions that are unpredictable for a human actor, even the designer of the system.[78] This ability to come up with decisions that might not have been considered, or even identified as possible[79] creates huge problems for the principles of foreseeability and causation, and makes the liability issue a cluster of competing ideals and concerns. A human actor who employs the

---

[72] 365-370
[73] Scherer, above n59, at 364; Neil Johnson and others, "Abrupt Rise of New Machine Ecology Beyond Human Response Time," (2013) 3, Scientific Reports
[74] *Clements v Clements* [2012] SCC 32, per McLachlin CJ
[75] AP Simester and WJ Brookbanks, Principles of Criminal Law, 2007 at 55-56
[76] See cases like *Ogwo v Taylor* [1988] AC 431; *Rahman v Arearose Ltd* [2000] QB 351; and *Bonnington Castings Ltd v Wardlaw* [1956] AC 613
[77] *Lamb v London Borough of Camdem* [1981] QB 625 (Court of Appeal); *Cambridge Water v Eastern Counties* [1994] 1 All ER 53; An in-depth analysis of this concept can be found in [8.3.1] "The test of negligence"
[78] Scherer, above n59, at 364; citing Herbert Simon, "Rational Choice and the Structure of the Environment" (1956) 63(2) Psychological Review
[79] See the stroma diagnosis robot C-Path in Andrew H. Beck et al. "Systematic Analysis of Breast Cancer Morphology Uncovers Stromal Features Associated with Survival" (2011) Sci Transl. Med, Vol.3(108)

use of an AI would be very difficult to blame for an action that is entirely unforeseeable to them, especially as many AI's behaviours are dependent on the post-design experience and those who design, or intended actions of the AI are still often unable to predict what will result.[80] Scherer offers a brief solution that AI could be said to be a "superseding cause" – "an intervening force or act that is deemed sufficient to prevent liability for an actor whose tortious conduct was a factual cause of harm."[81] While this could be a satisfactory solution to the issue of causation, it does not then resolve who is liable and how liability could be enforced – particularly against a non-human actor.

It should be noted that the term "fairness" is used a number of times throughout this thesis, in several contexts: the fairness of an action taken by an AI, the fairness of decisions around data-curation, fairness of attributing liability, fairness of absolving one of responsibility and more. I acknowledge that the concept of fairness is both a contestable and abstract ideal on which to base any conclusions. Where fairness is used as a point of argumentation, it will be accompanied with reasoning and an attempt to draw on comparators or examples which effectively illustrate the intention behind the conclusion. However, it is understood that this still leaves room for ample debate on some of the conclusions drawn throughout this thesis.

A final issue mentioned by Kroll, and touched on briefly by Scherer and others, is the issue of "discrimination" within an AI system.[82] This is an issue that could be seen as an extension of "control of action", it also incorporates elements of understanding as to how a decision is made. Discrimination in a system can come in a number of ways, materialising in issues of race, gender, and even socio-economic issues. An example of this could be: an AI that diagnoses and recommends treatment plans on a large scale might recommend different plans based on a person's race, due to statistics about socio-economic capabilities of that racial group. This is an extreme example, but is an example of a way in which an AI can make judgements using factors that the public, or the user, might not intend or necessarily desire the AI to be making.[83] Kroll discusses how many designers are beginning to attempt to insert non-discrimination into AI development in their initial design and algorithms to ensure

[80] Scherer, above n59 at 365
[81] At 365 - 366
[82] Kroll and others, above n66, at 634, 642, 660
[83] A more comical, somewhat benign example, is the recent incidents of AI chat bots becoming racist, homophobic, or xenophobic in their speech due to outside "trolling" influences; see <https://www.forbes.com/sites/parmyolson/2018/02/26/artificial-intelligence-ai-bias-google/>

the "accountable AI" that society desires.[84] However, he goes on to identify how automated decisions can also complicate legal doctrines on disparate treatment and disparate impact, and actually create difficulties at a more fundamental level of our understanding of discrimination.[85] Healthcare is a system that is inherently linked with sensitive, personal issues for a person and involves a very intimate involvement in their life. Ensuring an AI acts in a way that does not jeopardize this relationship is vital for creating a system that can both be trusted and relied upon when the worst happens for an individual who seeks help. Discrimination and bias will be discussed more in depth as a component of respect, in Chapter Four.

### 2.3.3 The issue of design

This focuses on complications that occur during development, or at least prior to the AI's actual use. This is primarily focused on issues of identifying parties involved and their intentions in the design of a system; identifying those who could fall under the umbrella of liability or the fairness of expanding this umbrella. Control of design (COD) problems can be described in three terms: diffuseness, discreetness, and discreteness. Diffuseness is the issue that anyone and everyone is capable of developing AI, and these projects can involve actors in different locations, with different awareness of each other. Discreetness is the issue that the capability of developing AI is not limited to large-scale actors, and instead is possible to be developed by even a single person on their laptop. Finally, discreteness is the problem that these projects often utilise "discrete technologies", and the potential of these is not known until utilised in action alongside every other component.[86]

While generally issues that occur during development are a matter of product liability or similar regulation, AI is unique in that the issue is not necessarily that the machine is faulty, but that it changes to act in a new way due to unforeseen combinations. The first two issues are relevant for liability issues involving AI because it recognises that AI can be developed

[84] At 692-695
[85] At 693
[86] At 369

with a wide range of actors, some not even aware of who the others are, where they are, or why they are participating.

The issue of diffuseness is unusual in large-scale industrial AI, but is common for online collaborative works, or work where components of an AI are sourced from the wider community. Discreetness in respect of AI is in contrast to earlier technologies that posed a public risk, as Scherer identifies, because those were often only possible for large-scale national actors, therefore simplifying the regulatory process.[87] These create a problem for AI liability disputes because the people involved in the chain of production could be entirely anonymous, or in other jurisdictions. And their Code could be implemented in different stages of the process, by different parties, creating a web of issues for identifying the sources of faults. These problems could potentially be simplified through the principle of "assumption of risk", but this would rely on an extended discussion of what is considered "fair" generally. This principle will be discussed in depth in the context of negligence. Discreteness in AI can be contrasted to a more commonplace, and regulated technology of automobiles. Automobiles are often made using parts and techniques from multiple sources, all put together to form a single vehicle. The "sum" of these parts is not able to be properly identified until they are all put together, and therefore unexpected results or interactions are possible. For example, an AI system may pass testing and regulatory requirements within a controlled, laboratory environment. However, when integrated with a wider healthcare operating system, it may produce unexpected results and cause harm before these can even be identified. The ability to control these interactions is stymied the more complex the interactions become, and issues of fairness in holding parties liable become increasingly difficult. Kroll discusses the common argument that transparency, within a legal context, is often a "trump card" solution to these sorts of problems, however this still has both practical and theoretical limitations.[88] This approach will be discussed later, in chapter 10, after having highlighted some of the issues associated with liability in specific areas.

### 2.3.4 The issue of personhood

Although much less of an active concern currently, the issues of personhood or legal status need to be mentioned. As the technology becomes increasingly sophisticated, and capable of

[87] Scherer, above n59, at 369
[88] Kroll and others, above n66, at 633

replicating what is considered increasingly indistinguishable from human thought, the question of what status should be afforded to AI becomes more pressing. Opinions of how to classify AI vary; whether as software like an operating system, hardware like a computer, a simple algorithm or ruleset, analogous to a human being, or something new entirely. This issue of classification will not a focus of the following chapters in respect of NZ doctrine, however will be discussed in respect of potential reform. Chapter Nine will discuss this more in depth in light of the problems identified between, critically inquiring as to what status may be the most adept at resolving them as a part of wider reform.

## 2.4 Philosophical Considerations

While discussing the previous issues, much of the academic focus has been on trying to mitigate or remedy the conflicts or concerns identified. However, it is important to briefly mention the alternative; accepting these issues as necessary or worthwhile concessions. This is to consider and evaluate the value of potential solutions to these problems; to weigh the value of trying to overcome these solutions versus the benefit these technologies afford. For example, the BBP hinges on whether the value of understanding the process is seen as distinctly outweighing the value of an effective result. It is instead possible to decide that results are more valuable than the understanding, and therefore the BBP is simply a willing concession. This from herein will be referred to as "the value judgment"; judgments of this kind will be made throughout this thesis when referring to potential areas of law that are insufficient for the realities of AI. Choosing whether these inadequacies can be instead seen as appropriate concessions, as opposed to failures of the law, is a discussion that will need to occur when formulating ideas for reform. Hocquet penned an opinion piece called "Trust and Don't Verify" which discusses the BBP and how "trust" is an important metric of a systems viability in action; this metric however takes time to cultivate and develop.[89] He also remarks that "AI is based on statistical thinking – the art of uncertainty – and some tend to forget it." There is always an element of uncertainty to something built on a statistical model, but people develop a trust of it by understanding the science involved – similarly, a trust can be

[89] Philippe Hocquet, Trust and Don't Verify: The AI Black Box Problem, avail. on medium, accessed at <https://medium.com/@PhilippeHocquet/trust-and-dont-verify-the-ai-black-blox-problem-442c2b15e79e> on 15 August 2018

built in an AI even if it cannot fully be understood or guarantee a result. But if the concepts and mechanisms involved in reaching that result are loosely understood, this shortcoming might become null. Trust in a system can allow for circumvention of its shortcomings, however this will always depend on the aforementioned value judgment being made by those in the position to implement the technologies. While the law might often be seen as the enforcement of black-letter law, making judgments of these kinds allow for greatly flexibility and adaptation to new problems.

A value judgment of this kind could allow a government to determine that it is not concerned with where the results come from, or that there are certain limitations on rights occurring, so long as the results are verifiably correct.[90] For example, if a skin-cancer diagnosing deep learning algorithm is correct 98% of the time, it could be argued that how it made this diagnosis is of no concern due to how much better this result is to the human baseline.[91] This benefit is not exclusive to the BBP, and can also alleviate legal issues of both a procedural and pragmatic nature in other areas. The choice to arbitrarily assign blame or determine the standard on which to hold a medical professional liable could serve to simplify potential issues and create a more easily understood, agreed upon, and at least recognised standard. There are some potential issues with this system, such as the disproportionate assignment of fault, however clarity would allow whomever is in the "chain" of distribution that is targeted to prepare appropriately. This could lead hospitals to invest more into their own internal quality assurance, or industry standard verification processes. It could also mean a shift in how insurance is provided to healthcare providers or those seeking innovative treatments. The particulars of these problems and issues associated with them will be discussed in their appropriate chapters later in this thesis. Speculating on potential outcomes creates further uncertainty, but if the government were to elect to go this route it would create a more immediate, challengeable system for the industry to respond to. The outlined alternative would be many years of debate, patch-work fixes, and ever shifting standards that could lead to unfair, inconsistent or even dangerous outcomes in the interim.

Many recognise that objective knowledge is a practical impossibility, and even objective "science" can be seen with a degree of tentative caution. The idea of employing value

[90] Verification of this kind would result from consistent statistical analysis, as opposed to analysis of the process.
[91] Michael Phillips and others, "Detection of Malignant Melanoma Using Artificial Intelligence: An Observational Study of Diagnostic Accuracy" (2020) 10(1) Dermatol Pract Concept

judgments in the law is no more contentious, especially considering NZ's medical system is already built on an arguably equal judgment; ACC. ACC, as illustrated in the next chapter, is the codified recognition of the NZ health system that helping people following an accident is far more important than any punitive or negligence-based considerations in accidents. Whereas foreign systems, like that of the United States, have a reputation for being bogged down by often vitriolic negligence lawsuits for an endless variety of reasons, often comical or even nonsensical – NZ has made the deliberate decision to avoid these issues, barring exceptional circumstances.[92] This is a deliberate and conscious choice made that keeping its society healthy and functioning financially outweighs the benefits of established legal process. This has the added benefit of lightening the load on the NZ legal system and ensuring that medical compensation is swift and applied effectively. Like this, the decision could be made that the processes involved in AI are not of equal consideration as the results of such AI, and therefore only the output will be considered for regulation.

Depending on the result of this discussion, amendments could be made to make the issue of liability moot. This would be dependent on the determination that, as a society, NZ is not concerned with exact causation of decisions and their implementation, and that the results of these decisions are the law's core concern. This, as an issue, is far more suited for a philosophical or analytical analysis than a legal one. As a result, while this consideration will be mentioned intermittently throughout the ensuing chapters when the effects of this judgment would be relevant, this thesis will not attempt to resolve or give any greater attention than this to the idea. Instead, the discussion of reform will focus on attempting to find manageable solutions or potential regulatory changes that do not require this judgment to necessarily be made – at least not as firmly as suggested here.

## 2.5 Conclusion

In the nearly 70 years since the term has been coined, the definition of AI has continued to be debated, adjusted, and confused across disciplines and contexts of use. Establishing a

[92] Through the allowance of punitive damages, see section 317-319

working definition of AI is critical to discussions of its potential issues, as well as regulatory reform or management of these issues.

This chapter has discussed a potential definition of AI. It then considered specific technological issues raised by AI, and philosophical issues with its use.

The definition provided by this chapter is not suggested to be authoritative, or a conclusion to the search for a definition of AI, but it will provide a workable roadmap on which to discuss the remainder of this thesis. The issues identified relating to AI and its nature will be discussed as they arise in context, alongside different established concepts of law and rights within the NZ healthcare system. However, to do so the system in which these rights and rules operate must also be clearly defined, so the boundaries and limitations of the discussion and its application is clear. To reiterate, the definition of AI that will be relied upon hereafter is:

> A technology with the ability to perform tasks that would be considered requiring human intelligence, such as visual perception, speech recognition, and language translation. These technologies have the potential to learn and adapt to new experiences and stimuli, which enables their performance to reproduce and perhaps surpass human intelligence and capabilities.

The following chapter will provide a robust overview of the NZ healthcare system and its operation, so that the issues discussed in this chapter can be appropriately analysed in the following chapters. In doing so, these two chapters will form a foundation on which to present potential ideas of reform at the end of this thesis.

# Chapter 3: The Health System of New Zealand

## 3.1 Introduction:

To understand how the New Zealand (NZ) health system may be impacted by new developments in AI, it is important first to understand how it came to be in its current form, and how that form operates. This chapter aims to establish the principles and ideals on which the NZ healthcare system operates, and what mechanisms or processes would be engaged when adapting to new developments within the system. Doing so will demonstrate whether the health system is prepared for adaptation to new developments, and perhaps shed light on what, if any, changes should be introduced to improve its ability to adapt.

To do so, there will first be a brief overview of the history of the NZ health system, focusing primarily on its ideological underpinnings as opposed to structural developments. This will briefly discuss the two distinct systems, the European – predominantly British – settler and Māori indigenous systems of medicine and how they evolved in the early 19th century to form the foundation of what exists today. The settlement of NZ by British settlers in the early 19th century revealed the settler's toolbox medicine to be unable to cope with the new environment. Nor was the Māori traditional medicine equipped to deal with the arrival of European animal and person-carried diseases.[93] Whilst structurally similar to other Western healthcare systems, NZ's contains a number of distinct features and milestones which inform its current evolution. These features, and the systems evolution highlights the innate adaptability of the system operating today.

Following this there will be a discussion of the current structure and functionality of the NZ health system, focusing on the major bodies responsible for developments, oversight, and the administration of healthcare (in respect of both practical considerations and the fulfilment

[93] Sir Arthur Porritt BT, "History of Medicine in New Zealand" (Cambridge University Press, Cambridge, 1967) at 335

of obligations or duties). This structure will serve as a roadmap for the discussions in the ensuing chapters when discussing matters of liability and enforcement.

## 3.2 Development of the NZ system

### 3.2.1 Two approaches

Early Māori medicine had a strong focus on spirituality, like many aboriginal cultures, with *tohungas* – a form of witch doctor – working to remove *atua* – "spirits" from people who had fallen victim to them.[94] These spirits were believed to be a result of misbehaviour or violation of custom and not at this stage attached to bacteria or viruses however, Māori *tohungas* were already aware that certain foods, and the provision or treatment of foods could have a negative effect on people's health. They were also aware of beneficial uses of food for treatment. Treatment made use of the native bounty of flax leaves, kohu-kohu and matoutou leaves, and even the shell of pana being common remedies for fevers in tribal communities. Māori medicine, whilst seemingly very effective despite its comparatively primitive nature, relied on a sort of "trial and error" system. Solutions for problems had to be discovered over generations and then reapplied constantly to perfect them. This approach, while ultimately effective did not facilitate rapid responses to emergent problems, namely the outbreak of disease which greatly impacted the Māori population following the arrival of the European settlers.[95]

The European type of medicine that arrived with the early settlers[96] had benefitted from the Enlightenment in the previous two centuries and had a far more "modern" understanding of illness than that held by the local Māori population.[97] Only three decades prior to the

---

[94] Porritt, above n93, at 334

[95] This being furthered impacted by a lack of naturally established resistance to diseases like measles. See Ian Pool, 'Death rates and life expectancy - Effects of colonisation on Māori', Te Ara - the Encyclopedia of New Zealand, (http://www.TeAra.govt.nz/en/death-rates-and-life-expectancy/page-4) accessed 15 August 2018

[96] The first European to step foot on the new land was Dr Monkhouse, surgeon aboard the Endeavour in 1769, see Porritt, above n93, at 334

[97] Antoine van Leeunwenhoek's observation of microorganisms in 1676 could be said to be the first major step in the development of "modern" medicine. See Michael T. Madigan and others, Brock Biology of Microorganisms (12th ed) (Washington, Pearson Education Heg USA, 2018).

Treaty of Waitangi, an Italian scientist had proven that diseases were caused by microorganisms for the first time, validating theories that had slowly developed over the preceding three centuries.[98]

### 3.2.2 A union of approach

The arrival of whalers and settlers meant that new European diseases ran rampant amongst Māori communities.[99] Another common feature of British colonialism therefore reared its head; the inability of the local people to withstand the widespread death and illness brought by western arrivals. While the Treaty of Waitangi had been signed in 1840, tensions between local and emigrate communities continued to rise and boil into conflict as the increase in disease meant the new arrivals were seen as bringers of death. To maintain the peace and allow for even further integration of the settlers into New Zealand, a system was needed that incorporated Western medicine into the Māori communities and lands.

The more scientific approach brought by the settlers enabled the integrated medical systems of early NZ society to react more swiftly to new problems, like pandemic outbreaks of European illnesses among the Māori populace. This combined with the practical skill already in place amongst tribal groups,[100] their Mātauranga (or Māori knowledge) of the land around them, meant NZ had the groundwork laid for the ideals of the modern system. The intention was to create a healthcare regime that could effectively accommodate both people's without excluding (at least conceptually, historical discrimination against Māori was significant) anyone due to lack of effect redress, creating a "for-all" system.

The knowledge and skills available within the newly formed nation helped the developing NZ health system rapidly develop. The original Wakefield settlements made special provision for the arrival of doctors, with many choosing to set up local practices in whaling stations and settlements as early as 1838.[101] The first hospital was instituted in the new colony in 1846 for the 'sick and destitute Europeans and the free treatment of all Māori

---

[98] Michael T. Madigan, John Martinko, Brock Biology of Microorganisms (11th edition, Prentice Hall, London, 2006), at 405
[99] See Porritt, above n 93
[100] at 334
[101] Te Ara Encyclopedia, Dental Profession and Services, Te Ara Encyclopedia on New Zealand, edited by A.H. McLintock (http://www.TeAra.govt.nz/en/1966/medical-services/page-6%20on%2028) accessed on 03 March 2018

s.'[102] The divide between private and public healthcare was settled early in the country's history, with different hospitals receiving funding from: the fledgling central government, religious establishments or wealthy landowners in these formative years. The Minister of Health (MOH) became an established position in 1900, preceding England by nineteen years.[103] The development from this point on continued along a path of wide availability, accessibility and equal practice. Despite acting with pseudo-autonomy as a nation state in the early 20th century, NZ chose to primarily operate under the British healthcare rules and common law with only minor alterations. Changes were mostly focused on availability, particularly to do with cost, and the integration and availability for Māori people.

The country's relative isolation meant it was spared many of the widespread diseases of the 19th and 20th Century occurring throughout Europe, and as a result this meant it could focus on localised and systemic problems. Widespread public health initiatives did not become commonplace in Europe until the wake of World War I, following the 1918 flu epidemic,[104] but NZ had already laid a solid framework for these measures' decades in advance.[105]

The Health Act 1956, coming over a century after the signing of the Treaty of Waitangi, established the specific healthcare regime, structured from the ground up, that remains largely unchanged today. The ideals and principles underpinning this system are deeply rooted in universal availability and accessibility, continuing the tradition from which it originated in the 1840s.

### 3.2.3 Principles of the Unified System

European settlers brought with them an ethical framework within which medicine should operate. Māori medicine at the time was mostly built on a hierarchical system,[106] echoing the tribal framework it operated within. These principles stemmed from early Greek thought rooted in the Hippocratic Oath, which has been modified and adopted innumerable

---

[102] Porritt, above n93, at 336
[103] At 341
[104] John Barry, The Great Influenza: The Story of the Deadliest Pandemic in History, (Penguin Books, London, 2005) at 15-25
[105] See Porritt, as above n93, at 338-345 for a discussion of the development of certification standards, regulation of medicine, and accessibility to healthcare
[106] The medicine men of tribal areas were revered similar to shamans or witch doctors of other tribal cultures around the world. Whilst secondary to the Chief of the tribe in theory, they held great respect and status within the community. See Porritt, as above n93, at 320

times before reaching the modern age.[107] This influence can still be seen, with principles of non-maleficence and confidentiality persisting in almost identical formulations to earlier oaths.[108]

NZ's system developed with this ethos in mind, and it is still evident to the modern day in a multitude of health and medical standards, both formal and informal. What is commonly referred to as "The Code of Patient's Rights"[109] features many of the principles or underlying ideals of the Oath in a codified form in the modern health system, namely the aforementioned right of privacy[110] and non-maleficence.[111] Many of the accompanying legislation (discussed in [3.3]) also feature principles of the Oath and act as a complementary framework alongside the Code. It cannot be said that this development was swift or readily effective, and still to this day there are ethical issues, particularly around discrimination, present within NZ healthcare. However, recent work in regard to the Māori Health Authority and other initiatives (discussed in [3.3.2]) illustrate a continuing intention for progress. More work needs to be done to create a truly equitable and accessible healthcare system, but NZ's history of adaptability and inclusion lay an effective groundwork for such developments.

#### 3.2.3.1 Accident compensation

In the 20th century, reliance on traditional common law negligence principles began to highlight certain drawbacks. The uncertainty of result, cost of litigation, delays in the courts, and stresses involved in negligence actions meant those who had suffered harm often ended up far worse-off. 1956 saw the Workers' Compensation Act attempt to aid this in the area of work-place injuries by offering compensation for such injuries – this was ultimately criticized as not only insufficient, but also inadequate in its compensation.[112] The Social Security Act 1964 extended this to wider injuries and conditions, but only covered an individual for ongoing income they lost out on, as opposed to losses that had already occurred. The 1969

---

[107] Raphael Hulkower, "The History of the Hippocratic Oath: Outdated, Inauthentic, and Yet Still Relevant," (2010) Einstein Journal of Biology and Medicine 25, at 42-45
[108] G.K Daikos, "History of Medicine: Our Hippocratic Heritage," (2007) International Journal of Antimicrobial Agents 29, at 618
[109] Created as a schedule under the Health and Disability Commissioner (Code of Health and Disability Services Consumers' Rights) Regulations 1996
[110] Right 1
[111] Contained within various rights, such as Rights 1, 2, and 4
[112] Report of the Commission of Inquiry, Compensation for Personal Injury in New Zealand ("The Woodhouse Report") (Government Press, Wellington, New Zealand, 1967), at [225], [227], [228]

White Paper presented argumentation for and against this approach, and it appeared the Woodhouse Commission was in fact prepared to abandon wage-related benefits in 1967.[113]

The Woodhouse Report 1967 acted as the genesis for what is often colloquially known as "ACC": a universal compensation scheme for injuries. Five principles were established which would later become the foundation of the first Accident Compensation Act 1972. These were: (1) community responsibility, (2) comprehensive entitlements, (3) complete rehabilitation, (4) real compensation and (5) administrative efficiency.[114] This system would later evolve to the Accident Compensation Act 2001, known as "ACC" after the governing body that administers it.[115] The decades since Woodhouse have seen various politically motivated shifts in entitlements and coverage, but it has remained rooted in these principles. ACC's core commitment to a semi-universal healthcare protection for all truly articulates the NZ systems identity as developed through the early entanglement of European and Māori practices. Still unique globally, this system illustrates the strong health focus, and "for all" style of medicine that NZ operates.

#### 3.2.3.2 The Code of Patient's Rights

As the system continued to develop, many of the principles and ideals of the system began to be confused or overlooked in light of advancing technologies and new possibilities. The clearest example of this was seen in the 1987 Cervical Smear scandal in Christchurch. Professor Herbert Green had been conducting secretive research on women patients at the National Woman's Hospital. Green sought to prove that carcinoma-in-situ was not a pre-malignant disease, and to do so he followed women who had produced positive smear tests, but no colposcopic evidence of invasive cancer. These women were often discharged with misleading or deceptive advice on their condition, and this process ultimately led to the avoidable deaths of several women under his care.[116] His experimental treatment had lasted

---

[113] Richard Gaskins, Tort Reform in the Welfare State: The New Zealand Accident Compensation Commission, (1980) 18:2 Osgoode Hall Law Journal, at 251

[114] At 240-244; Gaskins also discusses other principals' core to this at 238-240; 244-248; the fifth principle remains a source of controversy for Accident Compensation

[115] For the purposes of this thesis, the organisational body will be referred to as "ACC", whereas accident compensation as a scheme/doctrine will be referred to as simply "AC."

[116] Charlotte Paul and Barbara Brookes, "The Rationalization of Unethical Research: Revisionist Accounts of the Tuskegee Syphilis Study and the New Zealand "Unfortunate Experiment", (2015) American Journal of Public Health Vol. 105(10) at e13

for nearly three decades until he was exposed by an article in *Metro.*[117] Metro's article led to the Cartwright Inquiry 1988[118], a commission setup to examine the claims made within the article. It found that there had been widespread and continuous failure to uphold ethical practices in respect of; communication, respect, information sharing, informed consent, surveillance procedures and new treatment.[119] Cartwright laid out a series of recommendations in light of these findings, which formed the foundation for the medical protections that exist today. The Inquiry ended with a series of recommendations and principled ideals that could be learned from the events. These sought to clarify, and embolden, some of the ideals of the system.[120] One of these recommendations was the establishment of what became the "Code of Patient's Rights."[121]

The Code sets out ten rights considered fundamental to the system, which act as a list of expectations and responsibilities on anyone working within the healthcare system in a way affecting patients.[122] These rights can be used as the basis for disciplinary proceedings by professional bodies, complaints under the Code itself, and malpractice lawsuits and carry the full force of law. The rights presented in the Code are the foundation on which treatment and care in the country is based. The specific rights contained, and how they will be approached within this thesis, is outlined in the following chapter.

### 3.2.4 Conclusion

The NZ health system has attempted to develop as a blend of the European and Māori systems that preceded it and was structured from the outset as a system striving for equity and accessibility. Whilst not always effective, with much work still to be done, the system has shown an adaptability to change and responsiveness to challenges that shows a path for future development. One such development is the introduction of ACC and the Code in the wake of systemic shortcomings identified in the twentieth century. This adaptable system, while

---

[117] Sandra Coney and Phillida Bunkle, An Unfortunate Experiment at National Women's, (1987) in Metro

[118] Silvia Rose Cartwright, The Report of the Cervical Cancer Inquiry: The Report of the Committee Inquiry into Allegations Concerning the Treatment of Cervical Cancer at National Women's Hospital into Other Related Matters, (Wellington, Government Printer, 1988)

[119] At 210

[120] Chapter 4 includes an in-depth discussion of the purpose and outcomes of the Cartwright Inquiry.

[121] Health and Disability Commissioner (Code of Health and Disability Services Consumers' Rights) Regulations 1996. Cited hereafter as "The Code of Patient's Rights"

[122] Section 2

flawed, does illustrate a potential to effectively adapt to the challenge of AI systems which will be discussed throughout this thesis.

## 3.3 The Current New Zealand System

### 3.3.1 Introduction

This section discusses the different components and operational functions of the current NZ health system. In 2020, the Labour Government announced a series of structural and institutional reforms spearheaded by Hon Andrew Little and Hon Peeni Henare. As this thesis was undertaken prior to, and during, the period of transition associated with these reforms, the following discussion presents the health system in the form it will take under the Labour reforms.[123] However, a brief explanation of the structures replaced will be provided to provide clarity on what this new system has changed.

Following this overview, there will be an explanation of the regulation of health practitioners within the outlined system, and how their role interacts with the systems legal protections.

### 3.3.2 Structure of the system

NZ's health system does not function as a singular administrative limb of government, but through the interconnection of several different types of entities. Each of these types have their own functions, intentions and governing legislation, but together form the wider national health sector.

The two main types of bodies to highlight are Departments of State (DOS) and Crown Entities (CE). These are not the only kinds involved, and are often reliant on various bodies established as extensions of themselves to effectively execute their directive and policy goals. However, they serve as the necessary level of abstraction for the purposes of this thesis.

[123] As this thesis is concerned with the application and interpretation of legal rights to patient situations, the structure of the system itself is only relevant in so far as to who is providing the care, and who may be responsible for it. In this regard, the structural changes are only of concern in respect of terminology.

Departments of State (DOS) are the various Ministries or Departments that govern particular areas of importance. In this thesis' case, the Ministry of Health heads the health system as the central authority on all things relating to public health within the country. As a DOS it acts as the overseer of all health and disability related regulation and issues, offering its voice as the government representative on any legislative matters.

Crown Entities are organisations established under the Crown Entities Act 2004 (CEA) which form NZ's state sector. These operate more akin to businesses or corporate bodies, where their management is separate from the organisational limbs associated with it. Where DOS' are responsible for overseeing a broad policy area, CE's are tasked with (generally) administering or achieving a specific purpose. Perhaps the most notable CE in respect of healthcare is the Accident Compensation Corporation (ACC) which is responsible for administering NZ's no-fault compensation scheme for accidental injuries.[124]

The functional bodies of the system are empowered, constrained or given purpose by their associated legislative or regulatory mechanisms. The Health Act 1956 (HA) is the principal legislation that governs all health-related matters in NZ, and forms the regulatory skeleton with which other pieces of legislation interact or amend.[125] One critical piece of legislation is the New Zealand Public Health and Disability Act 2000 (NZPHDA) which establishes the specific functional bodies for day-to-day management or healthcare. It is also the Act which contains the transitional and consequential provisions[126] which are being used to ensure the Labour Government reforms operate smoothly.

The next page provides two diagrams to illustrate the structural changes to the NZ health sector. The first diagram (A) is the system as it operated pre-reform, and the second (B) is the structure as it operates under the Labour reforms.[127] Both of these diagrams are taken from Labour's report *"Our health and disability system: Building a stronger health and disability*

[124] In this case, ACC operates akin to a public health insurance company as opposed to a government policy body.
[125] More purpose specific legislation will be discussed and outlined when relevant throughout the thesis, this explanation is only intended to highlight the overarching structure within which they operate.
[126] New Zealand Public Health and Disability Act 2000, Part 7, sections 93-114
[127] While the Labour Government did provide a roadmap of the reforms that would be undertaken, changes and variations throughout such a process are common. The version of the post-reform system that is provided is as known at time of writing: November 2021.

*system that delivers for all New Zealanders."*[128] Following these diagrams is a brief written explanation of the role of each body, and the key changes that occurred.

[128] (April 2021), available here https://dpmc.govt.nz/sites/default/files/2021-04/heallth-reform-white-paper-summary-apr21.pdf

**Diagram A** [129]

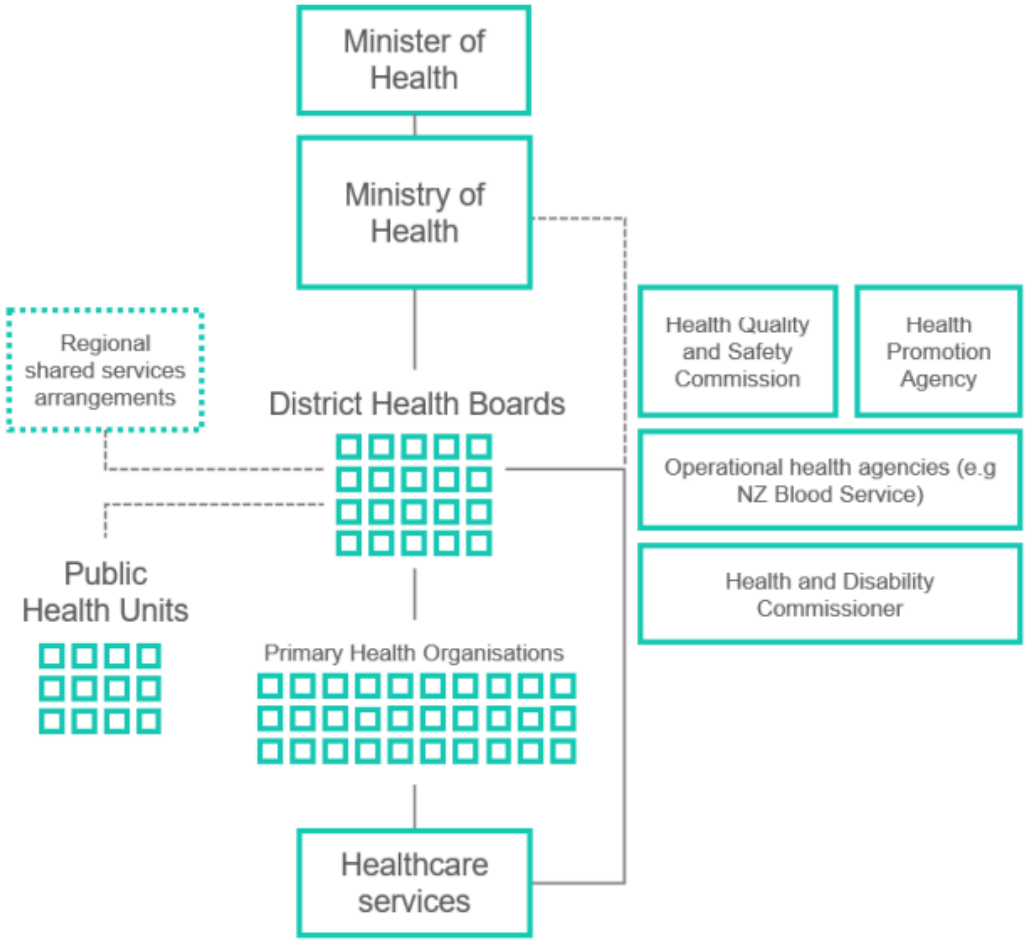
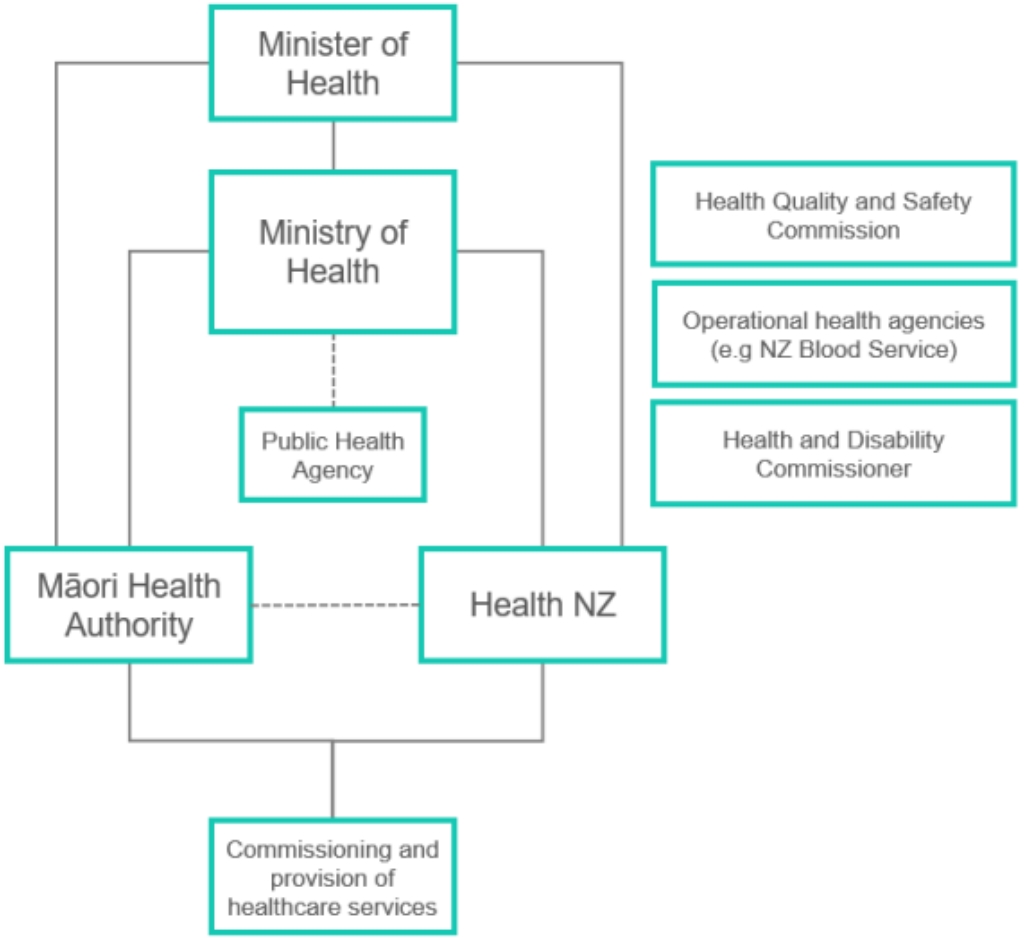



**Diagram B** [130]




[129] Page 5
[130] Page 6

#### 3.3.2.1 Diagram A: Prior to reform

The policy and strategic direction of the health system is determined by the Minister of Health, along with Cabinet and the government. The Minister is supported by the Ministry of Health (MOH) and a variety of business units and specialist advisory committees. Together this upper tier of the system serves to steward the health system as well as directly commissioning and manging some national services (such as maternity services).

The four agencies listed to the righthand side are responsible for more specialised or targeted areas of the system than general care and are responsible to the MOH. Not listed in the diagram, the Accident Compensation Commission (ACC) is a corporate business entity, accountable to the central government. While it plays an important role in the provision of healthcare in NZ, it is typically considered distinct from the primary tree of the system.

Beneath the Ministry, 20 District Health Boards (DHBs) existed to manage day-to-day public health and care. These DHBs were responsible for:

1. The direct commission of funding and administration to some care providers;
2. The management of 12 Public Health units (PHUs), responsible for things such as pandemic response;
3. Co-managing different regional shared services arrangements; and
4. The management of 30 Primary Health Organisations (PHOs) which served as the main funding and administration sources for most care providers.

The most important of these for patients are (1) and (4) which determine the ways in which they receive care, largely through their General Practices (GPs) and whom they interact with most. A PHOs can provide care either directly or through sub-contracted providers which they are responsible for.[131]

#### 3.3.2.2 Diagram B: Current system

The role of the top tier, the Minister and MOH, is largely unchanged. Still responsible for policy and strategic direction, the main difference is how they are associated with or

[131]Ministry of Health, About Primary Care Organisations, Online at <https://www.health.govt.nz/our-work/primary-health-care/about-primary-health-organisations> accessed on 26 March 2018

connected to the subservient components. Previously, in a strictly hierarchical system, the Minister was an abstract overseer to primary care separated by several layers of bureaucracy and management. In theory, the new system results in the Minister having a more direct relationship and engagement with the constituent parts of the system, thus creating a more cohesive system.[132] While the four side agencies are not critical to this thesis' purpose, of note the Health Promotion Agency was changed from one of these independent agencies to a business unit.

There is a new national body, the Public Health Agency, which is directly responsible to the MOH. This body is responsible for the management and protection of public health, most notably through pandemic preparedness and response (previously a component of the 12 PHUs). This body is now directly responsible to the Ministry as a national body, as opposed to a unit administered by a series of distinct local bodies.

The 20 DHBs and 12 PHUs have been consolidated into a single national agency, the Health NZ (HNZ). This change is intended to provide "true national planning" of health functions,[133] and will consolidate responsibility with a singular agency. HNZ will be administered through four regional divisions, and a range of district offices (called Population Health and Wellbeing Networks[134]).[135] To ensure that the system achieves equitable health outcomes for Māori, HNZ operates in tandem with the new Māori Health Authority (MHA). MHA is tasked with commissioning services in partnership with HNZ to ensure strong health outcomes and equitable care for the Māori population, due to previous concerns the complex system left them behind.[136]

Care continues to be provided by the same people and bodies as before; the same hospitals, GPs and health services with which patients interact are unchanged. The principal difference

---

[132] While the four side agencies are not critical to this thesis' purpose, of note the Health Promotion Agency was changed from one of these independent agencies to a business unit.
[133] Department of the Prime Minister and Cabinet, "The new health system" (2021), available at < https://dpmc.govt.nz/our-business-units/transition-unit/response-health-and-disability-system-review/information>
[134] Department of the Prime Minister and Cabinet, "Factsheet: Implementation Roadmap" (2021), available at <https://dpmc.govt.nz/our-business-units/transition-unit/response-health-and-disability-system-review/information>
[135] "Our health and disability system", above n128, at 6; the difference between these and DHBs is largely one of scale and independence. Instead of 20 distinct district bodies, there is a singular national service which administers a set policy.
[136] "Our health and disability system", above n128, at 5-6

is who these bodies are accountable to, and the administration bodies responsible for their funding and oversight have been simplified into national services.

### 3.3.3 Controls over Practitioners

The last major governmental components to the structure are a number of smaller regulatory bodies tasked with ethical or legal enforcement. These tend to be focused on very specific aspects of the health system and are usually limited in scope to issues of discipline or professional development within different areas. Notably, the Health Practitioners Disciplinary Tribunal (HPDT) is responsible for ensuring standards set for the profession are maintained in regard to privacy, consent, safety and dignity.[137] The HPDT is the enforcement arm of the rights and obligations contained within the health system. Bodies such as the HPDT prescribe qualifications and register practitioners and act as a mechanism to ensure standards are maintained and public confidence in the institutions ensues.[138] They exercise their powers with strong independent autonomy, however they are subject to audits and oversight by the Ministry to ensure that the exercise of their powers is within accordance with the overall policy objectives. These bodies are only of relevance in this thesis in regard to issues of punishment, discipline and malpractice.

An effective system requires there to be Codes of conduct and expectations in place to ensure those carrying out the policy vision do so appropriately. The Health and Disability Commissioner Act 1994 (HDCA) acts as the broad stroke outline of how medical professionals are to act, and within which boundaries. This Act, and the Code embedded within it, serve to outline the ideals and core principles that the health system rests upon. Made in response to unethical experimentation and an inwards disciplinary control of the medical institutions; the Act aimed to create greater external oversight and control over medicine, and the way it interacts with the populace. The purpose of the Act is stated as:

> To promote and protect the rights of health consumers and disability services consumers, and, to that end, to facilitate the fair, simple, speedy, and efficient resolution of complaints relating to infringements of those rights.[139]

---

[137] Health Practitioners Competence Assurance Act 2003, section 84
[138] Ministry of Health, Professional and Regulatory Bodies, Online at <https://www.health.govt.nz/new-zealand-health-system/key-health-sector-organisations-and-people/professional-and-regulatory-bodies> accessed on 15 April 2018
[139] The Health and Disability Commissioner Act 1994, Section 6

Prior to its enactment, disciplinary issues, complaints, and transgressions against patients were dealt with internally, by medical professionals acting as a disciplinary force towards other medical professionals. Without clear codification of rights and obligations, physicians and their peers tended to circle the wagons and insulate themselves from criticism or reproach. The Code acts as a Schedule within the wider Act, and spells out ten clear rights which all patients interacting with the medical system have, at all times. It also serves as components on which complaints or legal action can be hedged. The Code and its implications will be discussed in its own chapter[140] to provide a more robust overview of its purpose and role in the system.

Beneath the umbrella of the Code resides three main acts, some mentioned previously, that govern particular aspects of the profession more precisely. These serve as standards, regulations and guidelines which represent the vision and purpose of the HDCA and serve to create the framework necessary to fulfil it. These are: The Health Practitioners Competency Assurance Act 2003 (HPCAA), which sets the standard expected of those within the system[141]; the Health and Disability Services Act 2001 (HDSA), establishing the overall obligations within provided services[142]; and critically the Health Information Privacy Code 1994 (HIPC) which governs how information and data is to be utilised, stored, and treated within medicine.[143] The former two have clear problems with the advent of AI; new non-human parties involved in the diagnosis and treatment of patients leads to issues of circumventing, not being capable of, or not fulfilling standards set out. Importantly, the HPCAA only regulates the practice of practitioners who are in a position to cause "sufficient" harm to the public. Health professionals in low risk environments are not covered by this Act. Employer or self-regulating standards within a particular aspect of the health system is also common as a complementary, or alternative where the HPCAA does not apply. These standards tend to fall under employment or contractual law disputes, as opposed to the conventional negligence and malpractice issues that the HPCAA encompasses. The HDSA acts alongside this Act by providing the mechanism to certify different services and practitioners, combining a number of earlier pieces of legislation into a singular schedule.[144]

---

[140] See Chapter 4
[141] Health Practitioners Competency Assurance Act 2003 Section 3
[142] Health and Disability Services Act 2001 Section 3
[143] Health Information Privacy Code 1994 Section 4
[144] Health and Disability Services Act 2001, Schedules 3-4

Section 9 sets out the requirements to provide healthcare services, and acts as the entry-barrier that informs the competency requirements in the previous Act. These requirements, by their language, affect only human beings, or corporate bodies that are covered by the few institutional rules included.

Components of these Acts have interplay with a number of different rights within the Code and will each be discussed in more depth when they are applicable to specific situations discussed later. However, the latter, the Privacy Code, has potential to have the most devastating effects on the healthcare system come about. Big Data in medicine relies on the aggregation of immense amounts of data, from a huge range of patients and countries. While this information can result in seemingly miraculous diagnosis and treatments through AI, it also has to the potential to result in unprecedented privacy breaches, security intrusions, and will no doubt result in countless lawsuits as it is fine-tuned to the realities of medicine. Data hacks are one of the primary fears for legislators when pushing for new "smart cities" or similar developments, with incidents occurring in Singapore in 2017-18[145] and Hong Kong in 2016.[146] While not a focal point of this thesis, there will be numerous times when "public trust" is referred to as a core concept of the healthcare system. While this is not a codified or legally recognised principle, it will be discussed more as a conceptual ideal that the concept of the Code and its associated Acts depend upon. Privacy in particular will feature an in-depth discussion as to how public trust can be affected by both practical shortcomings, and perception of the technology itself. However, there will be mentions of this concept throughout the later discussions of the Code and specific rights.

## 3.4 Conclusion

This chapter outlined the history and development of the NZ health care system. The purpose of this was to illustrate what underlying principles or conceptual ideas can be reasoned to be embedded within the system. It is clear that the synthesis of European settler medicine and indigenous Māori practices helped create a system which is both adaptative to

[145] BBC, Singapore Personal Data Hack hits 1.5m Health Authorities Says, (20 July 2018) online at <https://www.bbc.com/news/world-asia-44900507> accessed on 10 September 2018
[146] Clifford Lo, After Singapore medical data hack, Hong Kong's Department of Health becomes latest cyberattack victim, (02 August 2018) South China Morning Post, online at <https://www.scmp.com/news/hong-kong/hong-kong-law-and-crime/article/2158023/after-singapore-medical-data-hack-hong-kongs> accessed on 18 October 2018

innovative and hardship, as well as inclusive of the countries peoples. This foundation will serve to inform the subsequent discussions of rights within the system, as well as the need for (and possibility of) prospective reforms to the system. The following chapter, Chapter Four, details the thematic ideals and rights within the system which will be discussed throughout the remainder of the thesis, and builds on the discussion in the first half of this chapter.

Following the historical outline, this chapter provided an overview of the key components of the health system relevant to this thesis. This illustrates which institutional bodies are to be kept in mind when discussing potential impacts of changes to the healthcare system, as well as who made be responsible for remedies, reforms and wrongdoings involved. The brief discussion in the Labour Government reforms in the health system also help to reinforce the inclusive and adaptative nature of the countries system, particularly with the reforms focus on simplifying and empowering care for Māori and at a local level.

This chapter concludes the background into the technologies at issue within this thesis, as well as the system which they may impact. The following part, Part B, engages with the substantive analysis into the four rights patients are afforded within NZ healthcare, and how those rights are impacted by the involvement of AI technologies.

Part B

# Legal issues and application

The previous part described and defined the technologies at issue and the health system in which they will interact. This part will now establish the areas of focus for inquiry. It will begin with an overview of the Code, then specific chapters will each deal with a specific right that patients in healthcare are afforded by the Code; non-discrimination and bias, privacy, consent, and care of an appropriate standard (or to not be treated negligently). In each chapter, there will be an overview of how that right functions within NZ's system, the potential issues in application that arise when utilising an AI, and a look at how these issues may manifest in legal situations in the future.

The purpose of this part is to examine how AI (as discussed in chapter 2) interacts with the health system (as discussed in chapter 3) and to highlight the areas in which the current formulation of the law cannot manage adequately. These conclusions will then inform the following discussion of reforms and recommendations in the next part.

# Chapter 4: The Code of Patient's Rights: Rights at issue and the Underlying Principles of Trust and Respect

## 4.1 Introduction

This chapter will discuss four key rights protected under the Code. It will explain why this thesis has selected these four rights to focus on, and will then establish the two underlying principles which will inform the discussions of these rights in the subsequent chapters. Before discussing these four rights, this chapter will consider two principles which underlie and inform the rights contained in the Code. The first is respect: explicitly codified as the first right of the Code, it can be argued as the prevailing theme throughout the ten rights (being mentioned numerous times throughout). The second is trust: not explicitly referred to within the Code, but arguably the lynchpin of healthcare as a system, and the requisite ideal for each of the recognised rights. Patients' trust in the system is necessary to ensure that patients engage with the system and, subsequently, feel they are respected within it.

The final section of this chapter will discuss the results of a survey in which participants outlined their personal preferences, expectations, and concerns about AI within healthcare. This will emphasise and demonstrate the importance of the selected rights, as well as identify the bases on what participants felt their trust most relied. These results provide important context for the ensuing discussions, in that they provide awareness of where willingness or resistance to change lies, which informs later discussions of potential reform. This chapter will not engage directly with the three scenarios, described in [1.3.2.5]; however a number of the survey questions relate to issues contained within one or more of these scenarios.

## 4.2 The Role of the Code

The Code of Patient's Rights ("the Code")[147] is the source of all medicine-specific rights for a patient who interacts with the NZ healthcare system. It also acts as the source of obligations on the part of medical professionals who interact with patients, by establishing

[147] Health and Disability Commissioner (Code of Health and Disability Services Consumers' Rights) Regulations 1996. Cited hereafter as "The Code of Patient's Rights"

clear requirements for them to fulfil. The Code attempts to consolidate established medical wisdom into a codified ruleset. It establishes ten rights afforded to patients within healthcare, each representing a commonly established human right or protection under law. Not all the rights within the Code will interact, or be affected by, AI and as a result do not require an in-depth discussion. However, many of the rights in the Code have parallels to other existing legal principles or obligations in common law and have potentially far-reaching consequences beyond just medicine. While the conceptions of the rights discussed will be as they appear within the Code itself, the rights which will be discussed beyond this chapter will draw upon their common law variations as well for the most holistic view of their principles and impacts.

#### 4.2.1 Background

In response to Herbert Green's cervical cancer experimentation, described in [3.2.3.2], a Royal Commission of Inquiry was convened to investigate what had occurred, and what reforms may be necessary to mitigate or prevent it in the future. This inquiry produced what is now known as "The Cartwright Inquiry".[148] Released in August 1988, Silvia Cartwrights investigation almost unanimously affirmed the claims made in Metro's expose of Herbert Green, detailing nearly three decades of experimentation. The report contains a series of recommendations for prospective medical law reform and procedural changes. Two critical recommendations amongst these were the development of a Code of Patient's Rights, and the establishment of an independent health overseer, which would come to be known as the Health and Disability Commissioner (HDC).[149] Some have argued that the real change wrought by the inquiry was its identification of, and challenge to, established power dynamics in medicine.[150] The Ministry of Women's Affairs, in their closing submission to the Inquiry stated:

[148] Full citation provided in fn 118. Cited hereafter as "Cartwright, The Report of the Cervical Cancer Inquiry, (1988)"
[149] Nie and Anderson, "Bioethics in New Zealand: A Historical and Sociological Review," Annals of Bioethics, ed. Peppin and Cherry (Lisse, Swetz and Zeitlinger, 2003) at 341-360
[150] Sandra Coney, "The Unfortunate Experiment: The Full Story Behind the Inquiry into Cervical Cancer Treatment" (Wellington: Penguin Books, 1988) at 17-18

> "Ultimately the issues are who controls medicine and how, about who benefits from it and who are its victims. Thus, as so many witnesses have clearly stated, the central issue, above all others, is power."[151]

The historical "doctor knows best" paradigm of healthcare had been broken by Green's experiment, and in its wake a new patient-first regime was erected in NZ.

The outcomes and purpose of the Inquiry have not been without criticism, however. At the time, some critics argued that the Inquiries' findings were disproportionately pro-women and were in some way influenced by a "feminist regime."[152] Linda Bryder argued against the need for the inquiry itself in retrospect, stating that Green had not conducted anything beyond "conservative treatment."[153] The inclusion of Bryder's argument, and other somewhat dissenting views of the inquiry, in the 2011 "*The Cartwright Papers: Essays on the Cervical Cancer Inquiry 1987-1988*" by Joanna Manning has raised some discussion amongst academics.[154] A number of essays in this collection point to negative effects that resulted on the medical profession as a result of the changes[155] are disproportionate to the harm they allege to be preventing, or unjust.[156] Some academics have discussed these later revisits to debates on unethical research as possible moralistic "rationalization" of the conduct.[157] While Bryder's position has been refuted by scientific investigation,[158] there is still a minority group that affirms Bryder's view that change was already underway in NZ and Cartwright did little to hasten it, dampening what they see as the trivial impact of the Cartwright Inquiry. Conversely, while the Inquiry can be seen to have brought about widespread positive change

---

[151] Charlotte Paul, Internal and External Morality in Medicine: Lessons from New Zealand, (2000) Biomedical Journal Vol.320(7233) at 500
[152] Anne Else, The "Unfortunate Experiment" and the Cartwright Inquiry, twenty years on: Why getting it right matters, (2010) Vol.24(2) Women's Studies Journal at 4
[153] At 3; quoting Linda Bryder, A History of "The Unfortunate Experiment" at National Women's Hospital, (Auckland, Auckland University Press, 2009)
[154] Else, as above n152, at 5
[155] David Skegg, "Foreword" in Joanna Manning ed, The Cartwright Papers: Essays on the Cervical Cancer Inquiry 1987-1988 (Wellington, Bridget Williams Books, 2009), at 93
[156] Charlotte Paul, "Medicine in Context", in Joanna Manning ed, The Cartwright Papers: Essays on the Cervical Cancer Inquiry 1987-1988 (Wellington, Bridget Williams Books, 2009), at 118
[157] See fn 150
[158] See McCredie and others, "Consequences in Women of Participating in a Study of the Natural History of Cervical Intraepithelial Neoplasia," (2010) 50(4) Aust NZ J Obstet Gynaecol, for the leading retrospective investigation of Green's research

towards the NZ healthcare system, some of its most vocal proponents feel change needs to continue.[159]

The Inquiry, in direct response to the specifics of Green's experiment, helped bring about one of the world's most successful cervical screening programmes and led to a dramatic reduction in deaths in the following two decades. The strong focus on patient-focused healthcare helped bring about the system that is in place today. This focus was directed by the newly appointed Health and Disability Commissioner (HDC), as well as the establishment of patient advocate positions, and nationwide ethics committees dedicated to the protection of research participants.[160] These ethics committees are of particular importance for the focus of this thesis, as many of the current uses of the most advanced AI in medicine are still being carried out as medical trials. The result of these recommendations is a higher patient participation in decision-making with consumer representation on decision-making boards being common practice.

As demonstrated in Chapter 2, the NZ system has shown itself to be adaptable, and capable of sudden change throughout its nearly 200-year history. It originated from two distinctly different systems, with different ideals, standards and understandings, and came together to form a unique and effective regime that was open to all. NZ's strong focus on public health from the very beginning meant that when science or conduct evolved in an unexpected, or undesirable, way there has been an immediate step towards a resolution.

The outcomes and developments of the Cartwright Inquiry can be seen as directly in response to Green's experimentation, and an example of NZ's adaptation to problems. AI represents a new major junction in scientific knowledge, and has the potential to be used malevolently, as well as in ways people simply do not understand. The current system will, in no doubt, be insufficient for all potential problems that may arise. However, the development of the Code does illustrate that the NZ health system is capable of adaptation and thorough responses to such problems. The remainder of this chapter will provide an overview of the ten patients' rights contained within the Code. Following that, the four main rights which will be

[159] An example being the Women's Health Action annual "Cartwright Anniversary" conference which focuses on social power structures in medicine, which they see the inquiry as having been in response to.
[160] Cartwright, as above n118, at 211-214

the focus of the rest of this thesis will be described, with an explanation as to why they have been selected for further consideration.

### 4.2.2 The Rights

The Code sets out ten rights considered fundamental to a well-functioning health system. These act as a strict list of expectations and responsibilities for anyone working within the healthcare system in a way affecting patients.[161] These rights can be used as the basis for disciplinary proceedings, complaints under the Code itself, and malpractice lawsuits. These rights are not simply recommendations or guidelines; they carry the full force of law and professional regulation. Complaints under the Code are investigated by the HDC who can deliver an opinion and make recommendations for remedial action. HDC decisions are publicly available as anonymised decisions.[162] Serious proceedings can result in professional disciplinary action through the Health Professional Disciplinary Tribunal or be referred to the Director of Proceedings who can institute disciplinary or civil proceedings in the Human Rights Review Tribunal.[163] These hearings can be the basis for fines, imprisonment, or removal of practicing certifications.[164]

It should also be recognised that the rights present in the Act are the foundation on which treatment and care in the country is based, and therefore many of these rights interact with other prevalent areas of the law. For example, the right to privacy is present in the Code and it also draws upon jurisprudence from beyond just medical law when addressing complaints. This means that even within a medical malpractice tribunal hearing, there will be reference to common law and statutory principles. The rights, listed in section 2, share some overlap in focus and specifics but the ten listed rights are[165]:

1. The right to be treated with respect;
2. The right to freedom from discrimination, coercion, harassment, and exploitation;
3. The right to dignity and independence;
4. The right to services of an appropriate standard;

---

[161] Section 2
[162] The latest decisions are posted here https://www.hdc.org.nz/decisions/latest-decisions/
[163] Health and Disability Commissioner Act 1994, section 45(f), 49(1)(a)
[164] Actioned under Part 4 of the Health and Disability Commissioner Act 1994, which contains the Code as its schedule.
[165] Code of Health and Disability Services Consumers' Rights 1994, Section 2, Right 1-10

5. The right to effective communication;
6. The right to be fully informed;
7. The right to make an informed choice and give informed consent;
8. The right to support;
9. Rights in respect of teaching or research; and
10. The right to complain.

Each of these rights comes with their own set of qualifiers, explanations and specific expectations and they aim to form a cohesive safety net for those undergoing treatment in NZ. These are given effect to by section 1, establishing that the Code applies to every "consumer"[166] and "provider".[167] These terms are defined under the Act.[168] A "provider" is defined as "a health care provider or disability services provider"[169] which may limit it explicitly to those certified and licensed as such.[170] A "consumer" is defined as a "health consumer or a disability services consumer"[171] These terms will not be used throughout this thesis, simply for ease of readability and understanding; a consumer will be known as a "patient", and a "provider" will be known as a "doctor" or "hospital" (when referencing the employer or controlling authority within which the doctor works). Any other parties involved in the following analysis will be specified where relevant (such as receptionists or administrative staff). Right 1 also contains the obligation to inform "consumers" of their rights under this Act, as well as ensure they properly understand them.[172]

Many of these rights, in varying ways, relate to the way in which a person is treated or given actionable choices in their treatment. As the Cartwright Inquiry concluded:

> "…had patients been . . . informed of the types of treatment available to them, informed of the risks of procedures which were not conventional, definitive treatment for carcinoma in situ, and given the opportunity freely to decide whether or not to be part of the trial, then the trial could not be so severely criticised."[173]

---

[166] Section 1(1)
[167] Section 1(2)
[168] Section 4
[169] Section 4, Definitions "provider"
[170] Developing jurisprudence perhaps suggests in situations where applicable, non-medical professionals may also be subject to these protections. A discussion of this occurs within Chapter Eight when discussing the duty of care, as one example.
[171] Section 4, Definitions "consumer"
[172] Section 1(3)
[173] Cartwright, as above n118, at 136

This highlights that the core issue the Code is concerned with is not success of treatment or harm caused by treatment (or failure to treat), but instead whether the patient was aware of, and understood, their options in advance; this is critical to the promotion of respect, and subsequently of trust in the care they are provided. Harm is often unavoidable within healthcare (particularly in the event of experimental or "cutting edge" treatment options), but patients being given the opportunity to appropriately understand and calculate that risk is vital to their understanding of that harm. Had Green explained what he was doing and why, some of the women may have approved his decision. However, the lack of the option to make a decision, based on proper information, was the paramount consideration in the Cartwright Inquiry.

Different rights will interact with AI in different ways, and each of these may warrant in-depth discussions of their own. Due to limitations on scale and scope for this thesis, all ten rights will not be able to be given equal, or in some cases any, specific consideration. Instead, four main rights, which correlate to common civil law doctrine, will be discussed. These four can be seen as a partial synthesis of the codified rights, as their different aspects often engage multiple rights. These four areas of discussion are:

1. Non-discrimination (right 2);
2. Privacy (right 1(2));
3. Consent (right 7); and
4. Negligence (standard of care) (right 4).

The specific features and requirements of each will be discussed separately in subsequent chapters. These four rights have been chosen due to their overlap with broader common law issues, which allows their discussion to have a wider impact than just within a medical context, but also provides a more robust breadth of resources which can be applied to their analysis. Additionally, all four of these (while not unique) have clear association with the ideals of both respect and trust, which underpin the NZ health system. Respect is explicitly referred to in the formulation of the first two, [174] whereas consent and negligence both relate to patients feeling they are being properly cared for, and included in the process. All four rights relate strongly to the concept of trust due to their focus on how their information and self is treated, utilised, and perhaps even recorded within healthcare.

[174] Right 3 contains the term "respect" as well, whereas right 2 deals with a number of moralistic elements of respecting an individual

It should be noted lastly that employers (hospitals or practices) are vicariously liable under section 72(2) of the Act, for ensuring their employees comply.[175] This means that failures to prevent an employee from violating the Code can result in action being taken against both the individual medical practitioners, as well as their wider employers. However, under section 72(5) an employer has an available defence if it can show it took "such steps as were reasonably practicable to prevent an employee from breaching."[176] This is often established through the provision of mandatory ethical training and reviews, and through the promotion of standards set by the Medical Council of New Zealand, such as its "Good Medical Practice – A guide for doctors."[177] How vicarious liability might operate in situations involving an AI will be discussed in Chapter Eight. However, it is noted now because it may be relevant when discussing the application of rights in earlier chapters, where it may be considered unfair to hold a doctor liable. This liability provides an avenue where perhaps where liability is necessary, it be enforced against the institutional body instead.

In practice, these rights are not applied by the HDC in decisions as strictly "legal" rights in the way a court would. Instead, these legal rights are applied in tandem with ethical conduct standards and obligations, such as those listed in the Medical Council of New Zealand's releases, the Medical Association Code of Ethics, and the Royal Australasian College of Surgeons policies. Additionally, these rights are often engaged with simultaneously as components of the same situation. A decision discussing privacy may also discuss the patients' right to be informed, as the reason privacy was violated was in an attempt to inform the patient of something.[178] Throughout this thesis the rights will be engaged with as purely legal concepts, and will not make attempts to analyse the way the ethical considerations will be impacted by the inclusion of AI. While the ethical and legal standards are often intertwined, they also sometimes lead to widely disparate standards for the same concepts, in different situations. And these rights will be discussed (where possible) as distinct, to highlight the shortcomings or benefits of each individually.

---

[175] Health and Disability Commissioner Act 1994, Section 72(2)
[176] Section 72(5)
[177] Medical Council of New Zealand, "Good Medical Practice" (2021) Available here < https://www.mcnz.org.nz/assets/standards/b3ad8bfba4/Good-Medical-Practice.pdf>
[178] For example, see Health and Disability Commissioner (NZ), Decision 00HDC03977 (07 June 2002) where this interaction was discussed <http://www.hdc.org.nz/ decisions--case-notes/commissioner%27s-decisions/2009/08hdc20258>

### 4.2.3 Experimental Treatment

A final noteworthy feature of the Code is its reference to "experimental" treatment within the rights.[179] Experimental is referred to in Right (7)(6)(b), requiring written consent when a "procedure is experimental". Similarly, the role of "research" in healthcare is commonly referred to in Rights 6, 7 and 9 of the Code. The inclusion of a reference to experimental treatment shows that the system at least recognizes an innate shortcoming of legislation and due process; the inability to adapt fast enough to external advances in technology and instead has included a mechanism for addressing innovation. The ability for the law to adapt rapidly to changing technologies has become a commonly accepted concern amongst legal theorists in the new millennium as it is becoming apparent that the existing law may not only be insufficient for possible occurrences, but is often also drafted in such a way that limits its ability to change or adapt to these. Godfrey recognises that this problem is muddied by a series of conflicting interests, particularly economics, legal, and public welfare interests.[180] He notes that "new scientific abilities often lead to problems that can be very difficult to solve"[181] and comments that if rights are not properly defined, they couple with poor assumptions or interpretations and create a law that from the outset is insufficient for the realities of the technology.[182] This is the crux of the issue at heart with AI, as a new field of technology continues to rapidly expand in the 21st Century, many people are either in the dark as to the reality – or simply lack the understanding to process the developments. As a result, the law needs to take proactive and effective steps to ensure that when these things reach patients and practitioners, that a framework exists to either nullify the risks or at least mitigate them enough to be manageable. This experimental treatment reference is the only real recognition of the imminent future in NZ law.

The inclusion of the reference to experimental treatment creates a flexible judicial safety net for issues relating to new technologies and techniques. It also provides a means through which the introduction of AI into an existing system can be resolved. Practitioners, as an example, may be cautious to use AI to delegate medical tasks that might open them up to wider malpractice lawsuits, but this provision creates a minor safety net for them to rely on so

---

[179] Section 2, Right 7(6)(b)
[180] Brett Godfrey, Law must keep pace with increasingly sophisticated technology to protect public health, Vol.1:2 (2014) WL7247054 at 3-4
[181] At 6
[182] At 6-7

long as the remainder of the Code is adhered to. This becomes less valid as AI becomes more common place and integrated, because regular practice could no longer be considered experimental. As Scherer emphasizes, beyond scarce references to driverless cars and specific uses of AI in tax and finance, most Western countries lack reference to AI's specific and ground-level issues.[183] While the potential need to reform and adapt the existing law, particularly accident compensation, to accommodate the drastic science-fiction technologies of tomorrow, the experimental treatment provision shows a system willing to evolve alongside these innovations. By testing the limits and shortcomings of the four Code principles, one can start to more appropriately tailor a response to the problem and hope to effectively wrangle the complications of AI before they become too drastic. While this works for current purposes, a more robust suggestion will occur in Chapter Nine.

## 4.3 Respect

Respect is the most commonly appearing concept within the Code's ten rights; having its own explicit right (Right 1) and explicitly or implicitly referred to in two others (Rights 2 and 3). It is important to establish what respect means, and how (and when) it is practically applied within the law. This will then serve as the foundational principle on which the four substantive chapters (Chapters Five to Eight), and the potential reform in Chapter Nine, will be measured. What this means is; solutions which may remedy issues identified in the following chapters may be rejected because they jeopardise the necessary relationship of respect between patient and the healthcare system.

### 4.3.1 Respect's place in law

At its most simple, the term "respect" typically means to show due regard to a person, group, or place one interacts with and their/its dignity. This can include respect for their feelings, beliefs, wishes, rights or other ideals central to their person. This section will outline firstly the principle of respect, both as a broad concept and then within the NZ law. It will then establish what position this concept will occupy throughout the remaining thesis.

[183] Matthew U. Scherer, Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies, Harvard Journal of Law & Technology, Vol.29 No.2 (2016), at 354-357

Respect can be illustrated by proactive behaviours (such as protecting those characteristics) or passively (by not interfering or contesting those characteristics). Importantly, respect does not require that one agrees with, or even necessarily understands, the needs or positions of what they are showing respect to.[184]

Respect appears throughout the law in a number of different disciplines and contexts. Within the criminal law, both victims of crimes, and perpetrators, are to be treated with regard and respect.[185] Within contract law, the intention to engage, and quality of information provided, is respected between participants.[186] Respect is often closely associated with the concept of "dignity": the state or quality of being worthy of honour or respect, which is considered innate within people.[187] This is expressly referred to within Right 3 of the Code, where every consumer has the "right to have services provided in a manner that respects the dignity and independence of the individual."[188] Showing respect is an inherently ethical component of the law, which protects the status and value of people and their position within society. Much like the broader Code and its origins, respect is about treating people equally.

### 4.3.2 Respect's place in the Code

Respect is present in the Code in several places; both within its own individual right and as a component of other rights as well. Appearing as the first Right within the Code, it appears to rightly occupy the space as core principle, leading into the following rights and their application. Respect is established in Right 1,[189] with three separate components:

> Right to be treated with respect
>
> (1) Every consumer has the right to be treated with respect.
>
> (2) Every consumer has the right to have his or her privacy respected.

---

[184] A common example of this, particularly within the health sector, would be respect of the rights and views of transgender people. Those who are not transgender, commonly known as "CIS gendered" people may not understand or appreciate the experiences and personal struggles of a group they do not belong with but can still respect that groups needs and expectations.

[185] For one such example, see Ministry of Justice "Restorative Justice: Best Practice in New Zealand" (2011) Wellington, MOJ at 19

[186] Charles Fried Contract as Promise: A Theory of Contractual Obligation (Harvard University Press, Cambridge, 1981) at 20

[187] NZ recognises this legally within the New Zealand Bill of Rights Act 1990, section 23(5) when referring to the "inherent dignity of the person."

[188] Right 3

[189] Code of Health and Disability Services Consumers' Rights 1994, Section 2, Right 1

> (3) Every consumer has the right to be provided with services that take into account the needs, values, and beliefs of different cultural, religious, social, and ethnic groups, including the needs, values, and beliefs of Māori .

The succeeding two rights within the Code also relate to, or refer to, respect. Right 2, the right to freedom from discrimination, coercion, harassment and exploitation, states:

> Every consumer has the right to be free from discrimination, coercion, harassment, and sexual, financial or other exploitation.[190]

This is the source of the right to non-discrimination, which will be discussed next chapter, but is another clear reference to the integrity and value of persons. Finally, Right 3, the right to dignity and independence says:

> Every consumer has the right to have services provided in a manner that respects the dignity and independence of the individual.[191]

As discussed, dignity and respect are intrinsically linked and this Right re-iterates this point with reference to a person's own value and their independence.

All three of these rights establish respect as the foundation on which the Code is set; people and their value is intimately linked to their health and provision of care. Subsequent rights all protect particular applications of the person; their knowledge, their body, their expression and so on. Respect will occupy this established paramount position throughout the remainder of this thesis. Where conflict arises between the emergent technology and a person established values or protections, deference will be made to the person automatically to comply with respect. This approach also ensures appropriate focus is given to the second vital component of healthcare; trust in its system.

## 4.4 Trust

Trust is the belief in the reliability, truth, or ability of a person, group, institution or thing. It serves as the calculated value of one's willingness to rely on something other than themselves, often subconsciously. Lewicki explains trust as "the trustor making themselves vulnerable based on the expectations about the trustee's likely actions, intentions or

[190] Right 2
[191] Right 3

capabilities."[192] Trust is an integral part of all law; the social contract between society and governance is reliant on trust of said governance in its applied mechanisms, and the society fulfilling its role beneath it. The importance of trust, as Karen S. Cook says, is "rarely acknowledged until it begins to break down, threatening the stability of social relationships once taken for granted."[193]

LaRosa and Banks have argued, during their inquiry on AI in healthcare, that patient-doctor trust is one of the most intimate and familiar forms of trust within a society.[194] Laura Crompton has argued that trust has become a "buzzword" surrounding discussions of AI and constitutes yet another "platonic point on the AI-design checklist".[195] In her analysis *Critical Analysis of the Trust Human Agents Have in Computational and Embodied AI* she concludes that trust in these AI is actually already afforded by humans at higher rates than media and publication would suggest to be the case.[196] However, she notes that this trust actually opens human agents to deception and manipulation through the "illusion" of AI authority, and that trust as a concept should not necessarily be as pedestaled as it often is.[197] While this is an important consideration to keep in mind, the nature of control over AI within healthcare and the intimate relationship formed between healthcare providers and their patients leads many of her concerns to perhaps be less-critical in this thesis. Although, as SN3 revolves around the use of an application distinct from the normal healthcare environment (and thus beyond the normally entrusted dynamic) it may be of greater importance. This thesis will not give complete deference to trust as the decider of issues, but will instead consider it alongside respect and functional application of the rights discussed in the succeeding chapters, to form a cohesive patient-first approach to AI in healthcare.

This section will briefly outline the role of, and importance of, trust within the healthcare system and this thesis. Alongside respect, trust will serve as the metric by which conflicts are

---

[192] R.J Lewicki, E.C Tomlinson, and N. Gillespie, Models of Interpersonal Trust Development: Theoretical Approaches, Empirical Evidence, and Future Directions. (2006) Vol.32 Journal of Management, at 993
[193] Karen S. Cook, "Trust in Society" Editor Karen S. Cook, (Russell Sage Foundation, London, 2001) at 5
[194] David Danks and Emily LaRosa Impacts on Trust of Healthcare AI, (Carnegie Mellon University, Psychology Press, 2015) at 2
[195] Laura Crompton ""A Critical Analysis of the Trust Human Agents Have in Computational and Embodied AI" in" in J. Seibt M. Nørskov, O.S Quick (ed) Culturally Sustainable Social Robotics (IOS Press, 2021) 623-631 at 623
[196] At 630
[197] At 626-627; at 630

analysed; where a possible solution to an issue arises, whether or not it would facilitate or hinder trust will be considered to determine its practicality. Being a highly personal inquiry, this cannot be determined for certain, but the survey results discussed below in [4.5] provide a loose framework through which to undertake this discussion. Importantly, this section will not discuss trust as an epistemological concept or its social characteristics in depth. Instead, the focus will purely be on a simple outline of the social role that trust plays, and the risks that AI pose to it as a practical concept.[198]

### 4.4.1 Trust in Healthcare

Within healthcare, trust plays a similarly vital role to respect, and is similarly also often taken for granted. Healthcare is often discussed as an innate component of one's life, and a service with which a person engages only when necessary. However, one of the problems that has become clearer in recent years is when people's trust in both the science and administration of healthcare is shaken, they choose instead not to engage with – or in some cases act contrary to – the systems assistance.[199] Discussions of AI and the future of healthcare often take for granted that people will be interacting with these systems; in other words, as medicine advances, the way in which people receive medicine will change. But it is not often discussed whether people will want to receive medical treatment from an AI at all (or at least, how they may choose to alter their interactions). Ensuring public trust in the changes within healthcare is vital to ensure the value of those changes is collected. As LaRosa and Banks have discussed, there are numerous studies analysing the methods and techniques to facilitate trust between doctor and patient[200] but how this will function in the advent of smart medicine is still unclear. The roles and capabilities of AI will lead to potential restructuring of conventional relationships, particularly in the realm of support and mental health where people are often most vulnerable.[201]

---

[198] For a brief analysis of the different natures of trust associated with healthcare and artificial intelligence in an ontological view, see Emily LaRosa and David Banks, Impacts on Trust of Healthcare AI, Carnegie Mellon University, 2018.
[199] For one such inquiry, see Jennifer M. Taber, Bryan Leyva, and Alexander Peroskie, "Why Do People Avoid Medical Care? A Qualitative Study Using National Data," (2005) Vol30:3 J Gen Intern Medical,
[200] E.N.H Montague and co., "Trust in Medical Technology by Patients and Health Care Providers in Obstetric Work Systems," (2011) Vol.29:5 Behaviour & Information Technology, at 541-554
[201] Danks and LaRosa, above n194, at 1

### 4.4.2 Trust in the Code

Within the Code, trust is not defined or included in a specific right at any point. The only reference to "trust" within the Code is under the definition of exploitation:

> Exploitation includes any abuse of a position of trust, breach of fiduciary duty, or exercise of undue influence.[202]

The reference here to "position of trust" is important because it signals the role that healthcare takes within the patient-doctor relationship. The patient has sought help, and they trust it to be provided in a manner according to their rights, by the system they have engaged. Maintaining this trust, and executing on this expectation, is the role of those who interact and engage with patients directly. How trust is facilitated or engaged is a highly personal inquiry; individuals rely on different expectations or considerations when analysing their own interactions with the system. However, what can be said is that people tend toward trusting things they understand, or at least have been informed of, which establishes transparency as a crucial component.

### 4.4.3 Role of Transparency

Often subsumed within discussions of trust is the idea of transparency. For an individual to trust a system, they often rely on either understanding it, or at least being informed and made aware of its different components, interests and applications. While this is of course an integral component of ideas like informed consent, there is also a passive method of facilitating this trust; through being transparent without prompting. Open transparency, or disclosure, enamours a society to a system, body or group by allowing them to be aware – at their own discretion – of what they consider important.

The goal of NZ's legal and governance system is one of transparency. Transparency in areas like governmental disclosures, court procedures and, information access has become a growing area of discussion and concern as more data is collected and utilised by the public sphere. As healthcare also develops into a smart-discipline, and begins to integrate new technologies and interconnectedness, it is important to maintain the established trust.

---

[202] Code of Health and Disability Services Consumers' Rights 1994, Definitions

NZ has shown a strong interest in ensuring transparency within healthcare and in recent years a number of initiatives have been undertaken. In 2016, a complaint by Ombudsman Professor Ron Paterson led the introduction of an annual Ministry of Health report on "increasing transparency in New Zealand health care."[203] The purpose of this report was to show the health sector's progress towards a set of specific goals. These goals, measured through statistical outcomes, were:

1) Meaningful to healthcare consumers (patients);
2) Meaningful to clinicians who provide their care;
3) Meaningfully attributable to the clinicians or services providing that care; and
4) To increase the availability of information to the people of New Zealand.[204]

This illustrates an effort by the Ministry to produce this information, and to facilitate trust between not only patients and the healthcare system, but also between the interconnected parts of the system itself. Cynically, reports of this nature could be argued to benefit only institutional or professional bodies. The general public is unlikely to engage with released reports of this nature, or even necessarily comprehend the data in its disclosed form. This relates closely to the issues of interpretability and explainability, which were discussed in Chapter Two in relation to one of AI's core issues: the black-box problem.[205] The practicalities of transparency and its impact on trust is both difficult to measure and difficult to remediate. This thesis will not attempt to present or grandstand solutions in respect of trust, but instead simply try to emphasise its importance and role within the system throughout.

---

[203] Health Quality & Safety Commission New Zealand, "First annual update on increasing transparency in New Zealand healthcare" by Health Quality Intelligence (2017), available at < https://www.hqsc.govt.nz/our-programmes/health-quality-evaluation/publications-and-resources/publication/2962/>

[204] Office of the Ombudsman. "Request for Complications data by named cardiothoracic surgeon and neurosurgeon. Case numbers 402136/402138/402140/402142/402144." (2016) Available at <http://www.ombudsman.parliament.nz/system/paperclip/document_files/document_files/1635/original/402136_etc_-_request_for_surgical_complications_data.pdf?1467187036> accessed on 05 June, 2020

[205] Discussed in Chapter Two [2.3]

## 4.5 Public Concerns in AI: Empirical Research

In much of the literature in respect of emerging AI assumptions are often made as to what the public, or consumers of a system, consider important for its effective use. As a simple example, it is often expected that patients want to understand how an AI system is utilising their data, before sharing with a Big Data system to trust its uses. This next section will provide a brief investigation into this and similar assumptions, to further emphasise the concept of value judgments that have been talked about thus far.

Rather than rely on assumptions as to public concerns around the use of AI in healthcare, this section will discuss two pieces of empirical work carried out to ascertain this. The first was carried out as part of this thesis' research, and will be discussed in detail. The second, carried out by Bristows,[206] a London based law firm, will then be discussed to show concerns identified by a section of the UK public. Under each discussion heading, the results of the NZ survey will be discussed first, and then compared to those within the Bristows survey. As the Bristows survey was about AI generally, it is of little use when discussing the healthcare-specific components of the NZ survey, and thus comparison will only occur for the preliminary questions of the survey.

### 4.5.1 Explanation of the two surveys

#### 4.5.1.1 New Zealand Survey (conducted by thesis author)

In 2018-2019, I conducted a survey into the impressions, understandings, and expectations of AI in healthcare. The purpose of this survey was to gain an insight into four main topics:

i. What the common degree of understanding was to what "AI" is;
ii. What concerns people had for the integration of AI into public life;
iii. Fears or concerns people had about AI in healthcare; and
iv. Their expectations for the proper and fair implementation of AI into healthcare.

The survey and associated ethics approval are included as Appendices' A and B. Participants of the survey were aged 18-30 at the time of surveying and were all people either currently enrolled in undergraduate or postgraduate study within NZ. This participant group was

[206] Chris Holder, Vikram Khurana and Mark Watts, above n41

chosen for a combination of ease of access and communication, as well as the consideration that they a demographic highly likely to encounter these technologies in healthcare as a reality. Older patients have shown a preference for traditional methods of healthcare[207] or at least a preference for doctor-involved healthcare, even when presented with alternative, new methods.[208] A younger generation is more likely to be open to interacting with the early-adopted AI technologies, and will interact with the smart health system of the future. As a result, not only are they possibly more receptive to technology, they are in theory more vulnerable to any detriment as well. Participants were contacted through a variety of small, community-based groups – targeting student groups and their communities allowed both for easier targeting of the age-range, and also simple access to them. The survey received Ethics Approval from the Human Ethics Committee at the University of Canterbury[209]. The survey was administered online, using survey tool Qualtrics. The total number of participants who completed the survey was 242.[210]

The survey first asked participants to self-report on their perceived understanding of "AI" as a concept. Following this, participants were asked to answer a series of questions about areas in which they felt AI would be of benefit, and areas in which AI would be a detriment, to society as a whole, in order to situate healthcare as an area of people's awareness or concern. Next the survey shifted its focus to AI within healthcare. The focus of these questions were on what the participant considers important in the future of healthcare, including; results, understanding and outcomes. The survey concluded with a series of questions relating to more specific areas of concern with AI, which will be discussed throughout the body of this thesis: consent, use of their data after the fact, and disclosure.

The intention of this survey was to ascertain if established concepts of law were aligned with what a statistically significant quantity of people considered "important". This can then be used as a basis to discuss what concepts are important to correct issues identified later, or what areas may be open to willing concessions in exchange for progressive medicine.

---

[207] Danica Rotar-Pavlic, Igor Svab, and Raymond Wetzels, How do older patients and their GPs evaluate shared decision-making in healthcare? (2008) Vol.8 BMC Geriatrics

[208] Anne Lise Holm, Astrid Karin Berland and Elisabeth Severinsson, Older Patients' Involvement in Shared Decision-Making—A Systematic Review, (2016) 6:3 Open Journal of Nursing

[209] Approval is included as Appendix B

[210] Total participants was 516 however the remainder did not complete more than two thirds of the questions involved, making conclusions drawn from their answers difficult. As a result, only those who answered the substantive portion of the survey have been included.

#### 4.5.1.2 United Kingdom Survey (conducted by Bristows)

Bristows believes that the emergence of AI technologies requires a vocal and prominent debate, which engages with regulators, businesses and the public sphere. They feel that central to this debate is the idea of "public trust", because widespread adoption of new technologies is reliant on people being accepting of its integration and use.[211] Their survey aimed to identify the way the UK public perceived the technologies; the functionality and capabilities, its role and usefulness in their daily lives, and the ways they expected regulators to be involved.[212] This survey built on the submission made by Bristows on the House of Lord's Select Committee on AI, in *Ready, Willing, and Able?*[213] in which they highlighted the importance of public trust in the successful propagation and regulation of AI technologies.

Their survey, considerably larger, had 2103 participants across a one-week period in July 2018. It is important to note that the exact intention and design of the two surveys do not align, and as a result the questions are not always directly comparable. Comparisons drawn will only be useful as indicators of trends or broad opinions.

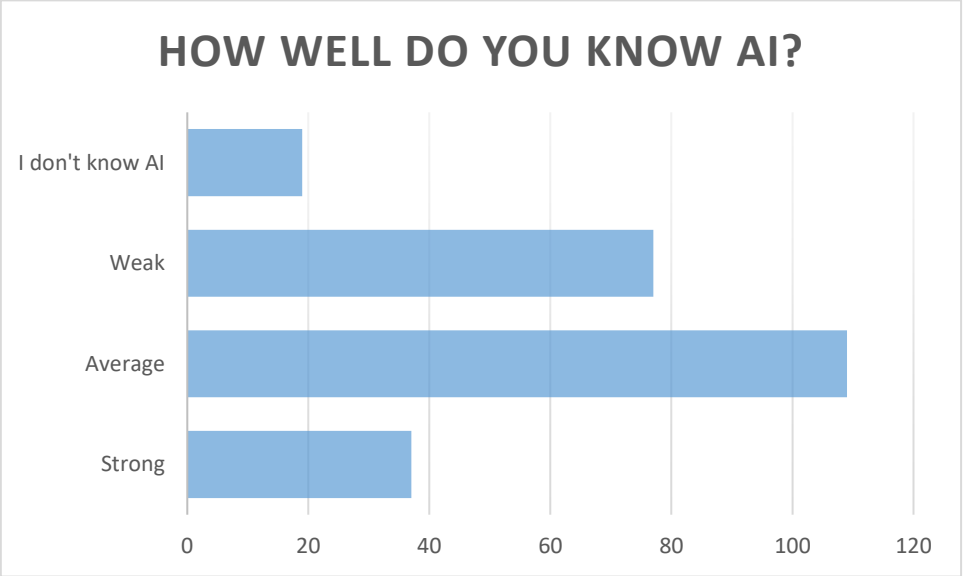
### 4.5.2 Participant Understanding (NZ)

The understanding of participants was self-reported and in no way examined or compared. This reflects the reality that trust is not built on verified rationality or accuracy, but is somewhat reliant on "instinctual" responses and feelings. For example, a patient could trust the efficacy of a blood test they are receiving, despite believing the blood test works via method X when in fact the blood is tested via method Y. Trust of a system is not inherently linked to understanding of it; or at least not verifiable understanding – in the case of complex systems.

---

[211] At 2
[212] At 2
[213] House of Lords Select Committee on Artificial Intelligence, AI in the UK: Ready, Willing, Able?, Report of Session 2017-19 (16 April 2018), above n 56

In Q1, when asked about their understanding of AI, the most common response from participants was that they had an "average" understanding of AI and AI concepts. 50% of the participants said they had no personal interest or association with AI through study or employment that shaped their views (Q2).




In terms of where the understanding of the other 50% originated from (Q2B), science-fiction and general entertainment (particularly novels and film) appeared the dominant reason for one's interest and association with AI as a concept. This was followed by social media, the news, or public forums (such as Reddit). Elon Musk[214] appeared commonly amongst responses, suggesting he plays a strong role in shaping popular understanding of AI as a public figure who often speaks about its dangers.[215] Four participants referred to their understanding of AI being built on conspiracy theories or public fears, or commented that they themselves were afraid of AI and thus tried to avoid contemplating what they "thought" it was. Those who referenced learning about AI in their studies came from four dicsiplines: data science, engineering, philosophy, and computer science. However, the largest section of participants stated they had no background or personal interest in AI. As discussed in chapter

[214] An entrepreneur and business magnate, Elon Musk is one of AI's most vocal and popular commentators. His companies (Tesla Motors, SpaceX, Neuralink and OpenAI) are all involved in cutting-edge, or publicly prominent, AI research and development. His business interests in AI range from driverless cars, to human neurological augmentation, to natural language processing and problem solving.

[215] Elon Musk is known to be a regular doomsayer in relation to AI, which may result in these participants have a disproportionately negative perception. For an example, see "Elon Musk warns A.I. could create an 'immortal dictator from which we can never escape'" (2018) by Ryan Browne, available at <https://www.cnbc.com/2018/04/06/elon-musk-warns-ai-could-create-immortal-dictator-in-documentary.html>

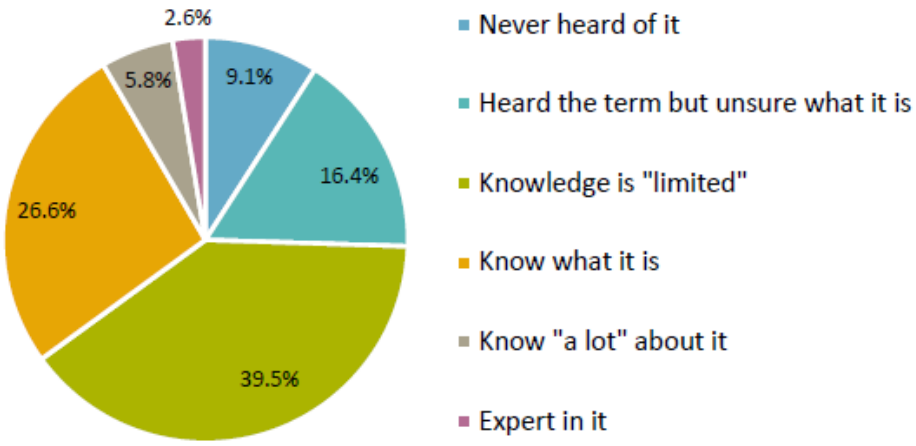
2, this likely leads to a skewed impression of both what qualifies as AI, and also the capabilities and realities of AI.

A recognised component of patient understanding is their own experiences with such technologies. Overwhelming participants responded that they believed they have interacted with AI, with over 50% answering definitively, and 38% suggesting they "probably" had (Q5). When asked for further information, the experiences they reported were ones linked to either entertainment or personal communications and organisational systems. Commonly given examples were Apple's Siri and Microsoft's Cortana. Similarly, algorithms for services such as Netflix, which provide recommendations for subscribers based on consumer viewing patterns were other common responses in this area. Interestingly, perhaps highlighting the skewed understanding, many responses in this area felt it necessary to qualify their answers, referring to their own uncertainty as to whether a technology counted as an example, despite already answering definitively of having encountered AI. And additionally, several participants chose to explain that while they felt a particular technology, such as their phone, was an AI, it was not "self-aware". This suggests some degree of self-correcting; participants might answer one way based on how a particular question is phrased, but might alter or shift their opinion when prompted with further, differently worded, questions.

Below is a word-cloud representation of the forms of AI that participants believed they had encountered in their own experience. The largest words represent the forms most often mentioned.


4.5.2.1 Bristows Comparison (UK)

The average evaluation of one's knowledge of AI by NZ participants was quite high, on average higher than that indicated in the Bristow's survey where the majority of participants professed "limited" knowledge.[216] Although the scale in Bristow's survey was more sub-divided, it would be reasonable to suggest that their "limited" knowledge is the equivalent to "weak."




[216] Holder, Vikram and Watts, above n 41 at 16

This may suggest that participants, perhaps due to age or sample size, over-estimate their own understanding or knowledge of a subject.[217] Without more information provided about participant age, demographic or background, the disparity cannot be analysed further.

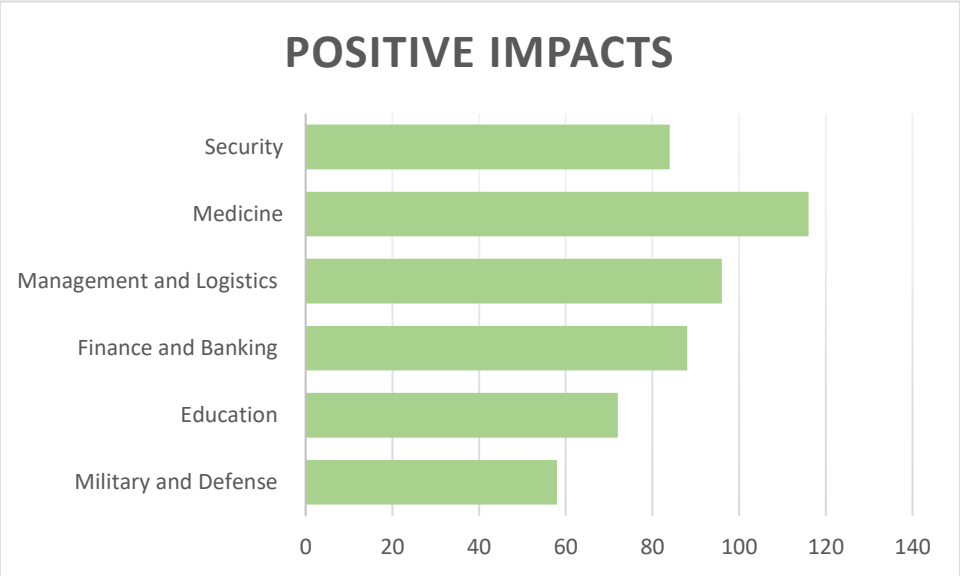
### 4.5.3 Sectors Impacted (NZ)

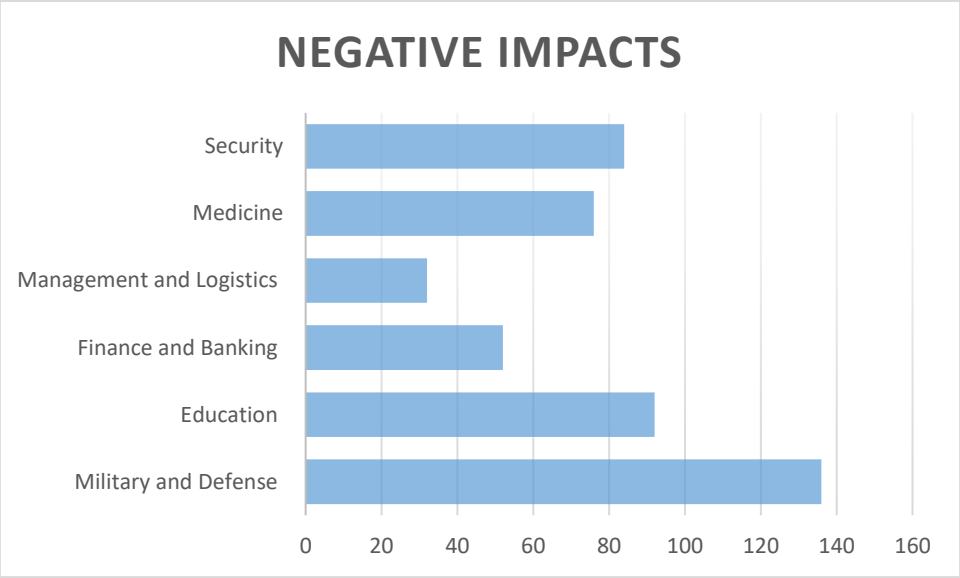
In Q3 participants were asked "Which area(s) of daily life do you think AI could provide the most positive impact?" And in Q4 they were asked the alternative question: "Which area(s) of daily life do you think AI could provide the most negative impact?" In both questions they were asked to select up to 3 areas out of a provided list. The list of sectors were the same between both questions: military and defence, education, finance and banking, management and logistics, medicine, and security. This list of potential sectors was chosen as a synthesis of those discussed specifically within *Ready, Willing, and Able?*[218] as well as being the areas in which the European Commission has noted concerns over.[219] Largely this list provides a broad cross-section of the areas in which individual's lives either operate using, or least are influenced by, AI. The outcomes in both questions seem to suggest a strong emotional response to this question, with the answers both ways favouring what could be inferred as the areas in which physical benefit or harm is possible.

In regard to positive impacts, medicine was the noted leader. Conversely, the area in which negative impacts were considered more likely, was within the military. However, medicine also ranked quite highly in this area, suggesting this is an area both with the potential for great benefits and widespread harms or concern.

---

[217] This is a well-documented phenomenon known as the Dunning-Kruger effect.
[218] House of Lords - Select Committee on Artificial Intelligence AI in the UK, above n 56
[219] European Commission Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions: Artificial Intelligence for Europe (2018)




As evidenced within these two charts, security scored highly in both questions. Almost equal responses in both recognised security as an area likely to be impacted. This is also of importance to the healthcare sector, as one of the principal concerns in respect of AI systems is privacy and data security for what is highly sensitive data.

Participants were asked to provide reasons to both questions (in Q3B and Q4B) as to why they ranked the sectors in their chosen orders. In respect of benefits, the answers given were highly varied and oftentimes personal. Those with experience in a particular sector appeared to favour that sector, suggesting their own experience was the deciding factor. However,

those sectors considered more personal required human involvement still, and thus would not benefit as well resulting in a somewhat paradoxical conclusion.

For the areas in which harm is considered likely, the reasoning was largely the same. Areas in which people could be physically harmed were of concern, and areas in which large amounts of sensitive data could be leaked or manipulated were also the central concerns of participants. Management and logistics appear to be the one sector in which participant opinion was nearly exclusively positive, perhaps due to its perceived nature as an already largely automated or computational field.

#### 4.5.3.1 Bristows Comparison (UK)

The Bristows survey did not engage with a specific separation of sectors, and instead asked their participants how quickly they believed the impacts of AI would be measurable, and whether these effects would be positive or negative broadly. The writers summarised their responses as:

> On average, respondents thought Al would start having a positive effect on human kind in four years (4.32) and a negative effect in five years (4.74), with one in four (28.6%) saying positive effects would appear within the next five or between five and ten years and one in five (20%) saying negative effects would appear within the same time period. This indicates that people see AI as a "game-changing" technology and therefore expect to see results relatively soon; that this feeling is more marked in those who are worried about possible negative impacts suggests this effect is caused by adverse media reporting of the potential harm that can be done by AI.[220]

The NZ survey did not contain a similar question about timing, instead focusing on areas of impact. Due to the growing perception of the technology and its integration shown in [4.5.2], it is reasonable to assume that the timeline may be accelerated in the NZ participant's minds. Bristow's also asked about which fields would be most impacted, however they provided a much broader spectrum of specific work-fields. Their questions, while similar, are difficult to compare due to the focus on employment impact and usability specifically.

[220] Holder, Vikram and Watts, above n 41, at 10

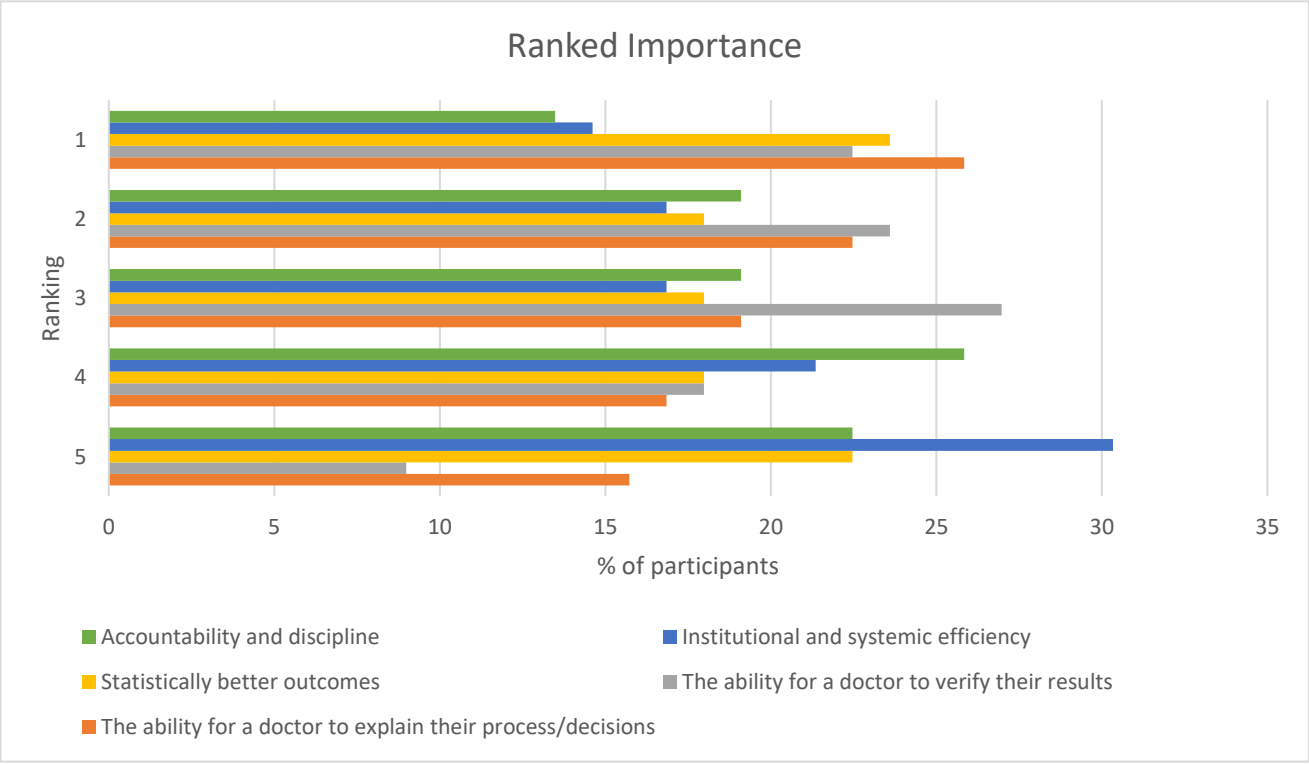
### 4.5.4 Healthcare Expectations (NZ)

Once participant's personal understanding and areas of concern were identified, the survey shifted focus to healthcare. This portion of the survey dealt with what participants expected to experience or be offered during their healthcare interactions, and what components of care they most valued. The intention here was to identify what is prioritized by participants, to help illuminate what things might be more necessary to protect, and what can conversely be conceded in the changing landscape.

In Q6, participants were first asked to rank five concepts in order of importance for themselves; in other words, what things do they prioritise in their healthcare interactions? The five concepts given were:

i. The ability for a doctor to explain their process/decisions;
ii. The ability for a doctor to verify their results;
iii. Statistically better outcomes;
iv. Institutional and systemic efficiency; and
v. Accountability and discipline.

Due to the ability for participants to rank these five in any order there was a high variance of rankings and answers. Below is a graph displaying the results of these rankings, by percentage of participants to choose each option:



Interestingly, these results suggest that the main concern is whether the doctor is able to perform their task, and continue to communicate the process to patients. It is ranked first in the "most important" category (or "1"), and then second highest in places "2" and "3", where the highest ranked concern was the "ability for a doctor to verify their results." Both answers suggest that participants consider the role of the doctor, and their expertise, to be the dominant factor still. The actual structural changes that AI might benefit were considered of greatly lesser importance overall.

This aligns with Q7 where participants were asked who they prefer to provide their treatment: a human doctor, an "AI doctor"; it is situationally dependent; or prefer not to answer. Overwhelming participants said they were unsure/it was situationally dependent, or they preferred a human doctor. Less than 4% of participants preferred their treatment to be given by an AI doctor. The most common reasons given for this were two-fold: an inherent or established trust in human doctors that they felt could not be easily replaced, and the lack of more "human attributes" like intuition and experience in an AI. Those who answered in relation to situation dependent choices seemed to make this distinction on the basis of complexity; common colds or basic triage could be done by an AI, but more intimate or extensive care defaulted to human preference.

As a final question in this regard, participants were asked to describe what they preferred for an AI systems appearance or physical aspects. The intention of this question was to ascertain what was considered important in how patients, and their doctors, interact with an AI system in their healthcare. Participants were given four images as reference and then the freedom to describe their preference however they pleased. Additionally, they were given the prompt to include references to pop-culture if that helped with their description. This answer had some wide variance as well, with some participants declaring that a humanoid AI is too "creepy" or "uncomfortable", whereas others insisting that humanoid features would increase their comfort with the involvement. A common preference answered was some sort of machine or computer-based body only and for communication to occur through text as opposed to a synthetic voice. Overwhelmingly participants preferred image D, shown on the next page,[221] indicating a strong preference for machines with no human characteristics.

[221] Also found in Appendix 2

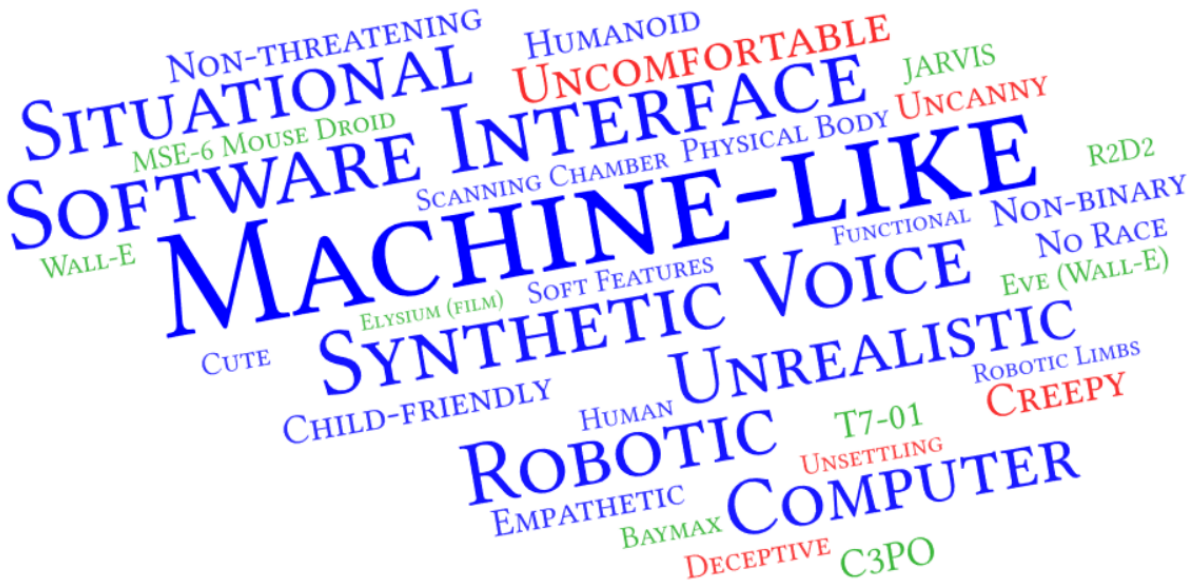
A: Robotic

B: Humanoid

C: Realistic

D: Machine

A word cloud is provided below to show the most commonly appeared words in these descriptions.



Interestingly these results highlight an important consideration; these preferences may be linked to broader cultural or social conventions, and warrant independent investigation in each jurisdiction. As an example, Japanese healthcare has already begun to trial different forms of robotic assistants for both in hospital and at home care (particularly for the elderly). Almost universally these companions are robotic, but humanoid; reminiscent of the science fiction robots of 1970's and 1980's cartoons or films.[222] This is more akin to Image B, as opposed to Image D which was overwhelming the preference by the participants in this survey. Japan's preference may be linked to deeper cultural associations of robotics, particularly through their entertainment industry with the prevalence of anime and manga depictions of science fiction.

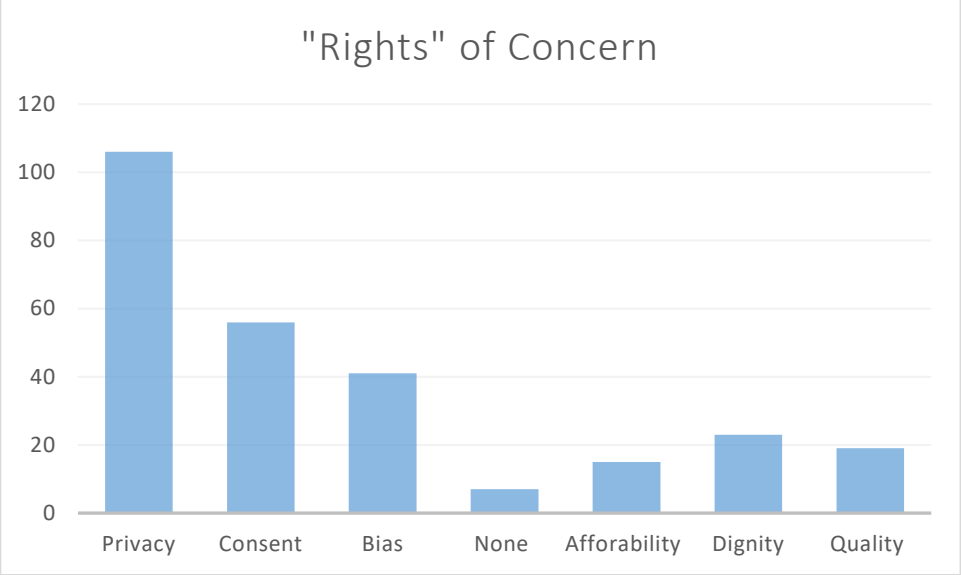
### 4.5.5 Rights and Concerns (NZ)

The final section of the survey asked participants about legal concepts and concerns, particularly focused on their rights and the controls in place for AI systems. These questions were not prefaced with explanations or guidelines on legal terminology or concepts, instead again relying on participants self-understanding of the ideas being asked. The reason for this is that while people often speak of their "rights" or protections, they often lack a technical understanding of its specifics. However, many rights still defer to the person's "reasonable" expectations, understandings, or obligations and as a result personal understanding of these concepts is an important measure of their import. As an example, privacy is often framed as the "reasonable expectation in the circumstance" – if a large proportion of participants indicate they think a particular situation violates their privacy, it provides insight to what the common reasonable expectation is.

The first question, provided as a form of framing question, asked participants what right(s) they believe they are entitled to within medicine might be affected by the involvement of an AI (Q17). Some key words were provided, these were: discrimination, consent, privacy, bias (racial, cultural, gender) and "informed." The question was not framed to suggest negative or positive implications, however notably the answers focused exclusively on negative impacts – infringements or failures being highlighted often. Although the participants wrote their own

[222] A discussion of this development is given by Luke Mahoney in "The Rise of Companions Robots in Japan," (2019) available at <https://japantoday.com/category/tech/the-rise-of-companion-robots-in-japan>

answers, so as to not skew their considerations by providing them limited options, overwhelming their answers referred to specific concepts (often with brief explanations as to why attached). Below is a graph representation of the answers given:

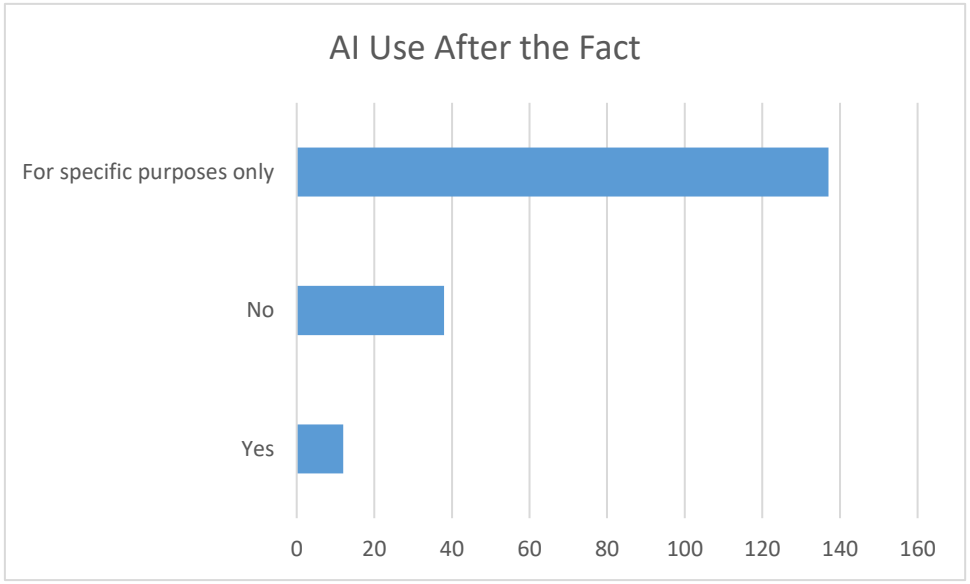


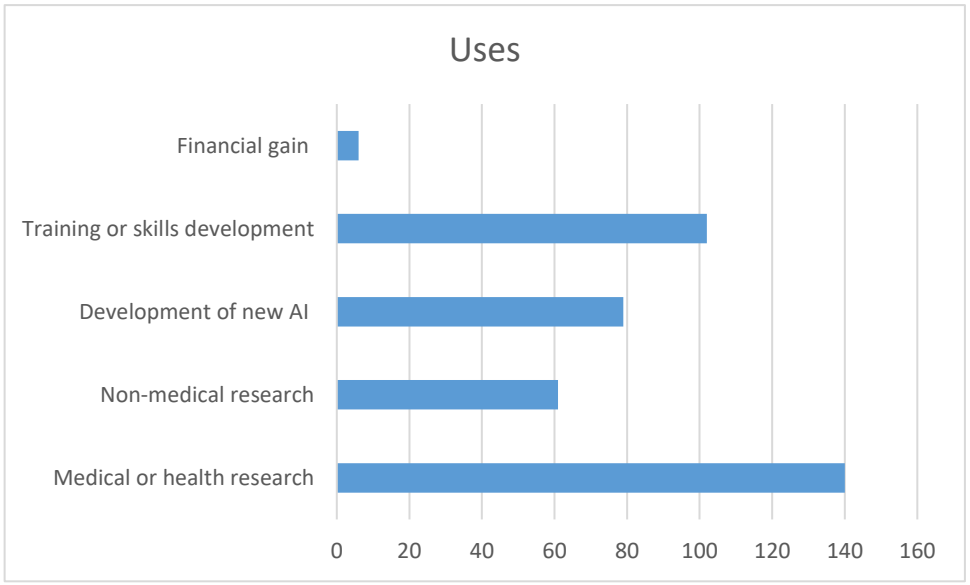

Note, both affordability of healthcare and its quality were mentioned by participants. This perhaps illustrates a lack of understanding of what "rights" means in this context, but does serve to show the considerations of people. Privacy is clearly the overwhelming concern for participants, with many making reference to how the information is both stored and processed when in use. This aligns with the subsequent question which asked what the participant considers the most important aspect of healthcare, with privacy and "superior medical outcomes" rankings first and second by a large margin.

Following this question was a more targeted question, asking which of the rights or concepts they felt would be impacted are incompatible with AI (Q17B). They were asked also to provide (if possible) an explanation as to whether this incompatibility was remediable, or even worthwhile. For this, the overwhelming response can be summarised as "I don't know." Most participants here recognised their own lack of understanding of both AI generally and their medical rights, and considered themselves ill-prepared to answer this question.

Participants were also asked about use of their data or information collected by an AI being made available to other parties (Q14), importantly *why* this information is shared was not yet mentioned. They were asked to choose between "yes", "no", and "for specific purposes only." In Q14B, participants were who answered the latter were asked to provide

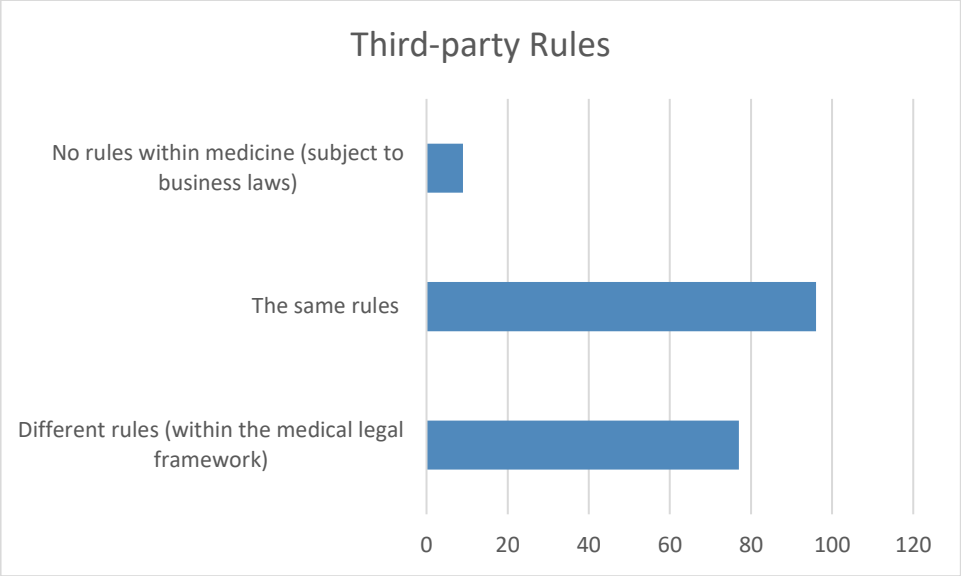
either choose a broad category provided, or a brief written reason, as to when it would be appropriate. Charts for both Q14 and Q14B are shown below:




An in-depth discussion of use after-the-fact will occur within Chapter 7 when discussing consent. A final closing question for this discussion of use was whether participants expected third-party organisations (the example of Google was given) outside healthcare, to be subject to the same rules as healthcare providers or different, unique rules (Q15). Interestingly, 42% of participants said they felt third-party corporations should be subject to different rules, within the healthcare framework. This appears to be recognition that structural differences

exist between sectors, and that many of the rules or obligations in place within healthcare are highly contextual and would be inapplicable to other sectors.




These final questions, while relevant to the discussion of trust and respect, are more relevant to discussions of privacy and consent. Within these chapters there will be a further discussion of these concerns, and how these results align with common academic concerns expressed thus far. To avoid redundancy, discussion of these questions will occur in these chapters.

## 4.6 Conclusion

The Code of Patient's Rights provides a robust and clear roadmap of the ideals of the current NZ healthcare system. These ideals are ultimately what both individuals interacting with the system, and those within it, expect from the system that provides such an integral service. Using the principles of the Code, surmised into four main pillars which represent the Codes main interactions with the wider law is the most effective means to identify the impacts of the AI revolution on healthcare services.

The survey conducted for this thesis shows that public perception of AI is still in its infancy, and that general understanding of the technology and its potential is lacking. Coupled with this lack of understanding, participants appear to strongly prefer the "human" element of healthcare, over any prospective results or improvements to efficiency. As a

result, the potential impacts of AI are even more crucial to ensure that this already tenuous acceptance of AI in healthcare is not jeopardised further by unclear rights or ineffective enforcement of protections in place.

The following four chapters will detail each of the selected themes one by one, discussing their function, issues when interacting with AI, and application to the three scenarios. Each of these will serve as separate guidelines on where the current system is either able to adapt to potential problems or is currently insufficient to respond to technologies that will soon reach the NZ shores. Where appropriate within each chapter, those rights that are not the focal point but are still relevant will be discussed to at least highlight their importance. In-depth analysis of the remaining rights, specifically those with little overlap to their companions, will be reliant on future research.

The proceeding two chapters will focus specifically on rights within the Code in which respect is the most clearly ingrained. Non-discrimination and privacy are both rights in which specific reference to respect, and oftentimes trust, is used when discussing their boundaries and appropriate application. These two chapters will act as an extension of this chapters discussion and provide a strong bridge to the later chapters of consent and negligence, which while still related to trust, are focused more onto actual physical harms.

# Chapter 5: Non-discrimination

> The era of big data is […] full of risk. The algorithmic systems that turn data into information are not infallible—they rely on the imperfect inputs, logic, probability, and people who design them. Predictors of success can become barriers to entry; careful marketing can be rooted in stereotype. Without deliberate care, these innovations can easily hardwire discrimination, reinforce bias, and mask opportunity.
>
> Megan Smith, The White House Archives of Barack Obama, 2015

## 5.1 Introduction

As one of healthcare's central tenets, a patients' need to feel respected and to have their personal characteristics treated with respect is paramount to effective healthcare. One of the forms that respect takes within the Code is contained in Right 2: the right to be free from "discrimination, coercion, harassment, and sexual, financial, or other exploitation".[223] This chapter will discuss the first ground within this right, discrimination. It will consider how discrimination may manifest due to the involvement of an AI, and how the application of the law around non-discrimination might be impacted by this involvement.

As machines are developed to supplant or augment professions that are so customer-focused as healthcare, those responsible will face difficult decisions about the variables and information utilised for the machine's functionality. Datasets utilised for AI systems are more than sets of arbitrary values and information; they contain stories and relationships between the data points and can carry the opinions and views of those responsible for their curation. Improperly curated datasets can contain embedded biases, and an AI system can then act on these biases, potentially against a large variety of people in a short time. As a White House report cautioned, "powerful algorithms can unlock value in the vast troves of information

[223] The Code of Patient's Rights, Right 2

available… but also raised the potential of encoding discrimination in automated decisions."[224]

In recent years, a number of examples have been published of AI exhibiting discriminatory biases against different peoples, within a variety of different contexts. A notable healthcare example was published in 2019, where researchers found that a system used to allocate health care to patients based on need was systemically discriminating against African-American patients.[225] It found that African-American patients were significantly less likely to be considered at "serious risk" than Caucasian patients, despite displaying otherwise equivalent symptoms. Similarly, within a criminal justice context, COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) is a system used by US courts to predict the likelihood of recidivism. It was revealed in ProPublica that COMPAS predicted twice as many false positives for African-American offenders (45%) than Caucasian offenders (23%),[226] resulting in a considerably higher recidivism "risk rating" for African-Americans. These are two examples of a multitude available, but the salient point is that evidence shows that the use of an AI can lead to a result that appears discriminatory.. Therefore, this chapter will not concern itself with how or why the use of an AI may result in a discriminatory outcome,[227] but instead focus on whether the legal definition and formulation of discrimination protections can be applied to such situations.

This chapter will begin with an explanation and discussion of the key terminology and factors involved in determining discrimination and liability in NZ. This will begin first with a discussion of the difference between "discriminatory" actions and "fair" actions. Discriminatory actions are when a difference in treatment occurs because of a protected class or characteristic (such as race), and this results in the claimant being harmed or put into a negative position. Fair actions are when this same information is used (such as race) in a way

---

[224] Executive Office of the Presidency of Barack Obama, "Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights" (May 2016), available here < https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf>

[225] Zaid Obermeyer and others, Dissecting racial bias in an algorithm used to manage the health of populations, (2019) 366(6464) Science, p447-453, at 451

[226] Jeff Larson and others, "How we analysed the COMPAS recidivism algorithm" (2016) Propublica. Available here https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

[227] The methods through which AI-driven decision-making can lead to discrimination is discussed in depth in the seminal paper of Barocas and Selbst, "Big Data's Disparate Impact" (2014) 571:104 California Law Review, at 5

that leads to a difference in treatment, but this difference is considered warranted, justified, or excusable. As a common medical example, giving a Caucasian patient and an African American patient different medicine (or sometimes even less medicine) for heart defects is commonplace.[228] This is a difference in treatment based on race (a protected class in most jurisdictions, including NZ) but in some medical situations it may be justified and thus fair. In most jurisdictions, the issue of fairness is dealt with when determining whether a defendant (or respondent) to a discrimination claim has a defence or exemption. Following the discussion of these two terms, an explanation of how these concepts are represented within the NZ law will be given.

Following the overview of the terminology involved and its formulation within NZ law, this chapter will give an analysis of the issues that arise when applying the law to situations involving AI systems. This analysis will broadly involve two main questions:

1) Can the test for discrimination, as it is formulated under NZ law, be applied to situations involving AI systems? And, if so;
2) Does the involvement of an AI system potentially create situations in which this application is undesirable?

This second question is not to suggest that patients should not be able to claim they have been discriminated against. Instead, it is asking whether the nature of AI leads to situations in which the current formulation of the risk against discrimination results in an unduly harsh application. The chapter will conclude with an application of the three scenarios, discussed in [1.3.2.5] to illustrate the conclusions drawn from these two questions.

[228] It should be noted that the impact of racial heritage on drug disposition is a hotly debated area of medical science. In A Ramamoorthy and others "Racial/ethnic differences in drug disposition and response: a review of recently approved drugs" (2015) 97(3) Clinical Pharmacological Therapy, it was found that 20% of newly approved drugs had known, measurable, racial/ethnic differences in disposition. However, Mario Cazzola and others in "How does race/ethnicity influence pharmacological response to asthma therapies?" (2017) 14 Expert Opinion on Drug Metabolism & Toxicology, found that racial differences were of a lesser effect than socioeconomic and environmental factors. However, this debate coupled with the black-box nature of AI only complicates further by obscuring prospective discussions of fairness and discrimination.

## 5.2 Terminology and the Law:

This section provides an overview of the different terminology involved when discussing discrimination, and an explanation as to how these concepts are represented within the law. First, the chapter will outline discrimination, and then the opposing concept of fairness.

### 5.2.1 What is discrimination?

A key connection to first identify is the relationship between "bias" and "discrimination." AI literature often refers to "AI bias" as an umbrella term for when an AI system acts in a way that is perceived as unfair to a party. However, the Code does not feature the term "bias" but instead refers to "freedom from discrimination, coercion, harassment and exploitation."[229] The terms 'bias' and 'discrimination' can be considered an extension of one another that forms a continuum from "idea" to "action." Bias is the "predisposition towards or against a particular thing, person or group".[230] Bias need not be negative or closed-minded, although the connotation of the term is often so. A simple form of bias is racial stereotyping. When this bias is unreasonable, against rationale, or in another sense destructive in nature, it becomes "prejudice."[231] This then distinguishes positive or neutral racial stereotypes from negative ones.[232] Finally, when this prejudice is actioned against a target party in a way to disenfranchise or cause harm, it becomes discrimination.[233] A visual representation of this progression is displayed below.


[229] Health and Disability Commissioner (Code of Health and Disability Services Consumer's Rights) Regulations 1995, Right 2
[230] Colin Gavaghan and others, Government Use of Artificial Intelligence in New Zealand (New Zealand Law Foundation, Otago University, 2019)
[231] Gavaghan and others, above n 230, at 43
[232] Of course, whether a stereotype is "positive" is an exclusively personal determination, but there can be distinctions made based on ostensibly negative stereotypes like "African American people are more violent" and ones that are (at least appearing) to be positive like "Asian people are great at mathematics!" A neutral stereotype might be "the Dutch are a tall people" as it is based on a commonly held belief or understanding of the Dutch, but has neither positive or negative connotations associated.
[233] Gavaghan and others, above n 230, at 43

### 5.2.2 Discrimination in NZ law

Protection from discrimination appears within the NZ law in a variety of different contexts. This section will detail first how the right to be free from discrimination appears in general human rights law under the New Zealand Bill of Rights Act 1990 (NZBORA) and Human Rights Act 1993 (HRA), and then specifically within the health system under the Code of Patient's Rights. Following this, an overview of how discrimination protections are applied within case law will be provided to illustrate the test involved in determining a breach of this right has occurred.

Within the NZBORA, the right to be free from discrimination is found in section 19, which states:

> Freedom from discrimination
>
> (1)Everyone has the right to freedom from discrimination on the grounds of discrimination in the Human Rights Act 1993.[234]

This provision acts in tandem with Schedule 2 of the HRA 1993, which provides a non-exhaustive list of protected characteristics that cannot be used to discriminate against an individual.[235] The list of protected characteristics is non-exhaustive, and generally relates to things of either personal belief such as religious or political opinion, or things of personal or physical character, such as race, gender, sex, sexual orientation, and age. In theory, any of these characteristics can be utilised to discriminate against a patient within healthcare.[236] These different characteristics are, at least in theory, considered equal; discriminating against a person for one characteristic is not better or worse than discriminating for another.[237]

---

[234] New Zealand Bill of Rights Act 1990, section 19

[235] Human Rights Act 1993, Schedule 2

[236] While physical characteristics like race and gender are the most obvious grounds on which to discriminate within healthcare, religion and political belief are also common. For an example of religious belief, see *B v DGSW* [1995] 2 NZLR 134 which was a case dealing with parental refusal of treatment for their son, due to their beliefs as Jehovah's Witnesses that a blood transfusion would deny their son access to heaven.

[237] In some jurisdictions, such as the United States of America, there is a hierarchy given to some of these protected characteristics altering their status by scrutiny or application (most notably race). See, for example, *San Antonio School District v Rodriguez* 411 US 1 (1973) at 29. New Zealand has yet to follow this route, at least not officially.

There has been some debate as to whether s19 NZBORA applies to healthcare scenarios, due to s3 restricting application of the Act to acts done by government bodies, or those with public functions, powers, or duties.[238] In *Ransfield v Radio Network Limited*[239] Miller J held that this provision excludes healthcare actions, especially those within the private medical sector. The reasoning given by Miller J was that the doctor-patient relationship is a private one, and not for the state to interfere in. However, this judgment has received criticism for presenting an impractical and illogical view of the role of healthcare,[240] and because of this will be treated with caution. Healthcare is a fundamental role of the state; caring for the public's wellbeing and overriding health is a responsibility of the state. The interpretation of Miller J appears to unorthodoxly restrict the application of NZBORA to state-ordered compulsory medical treatment or scientific actions. While this would prevent the historical atrocities that the NZBORA aims to prevent (such as the actions committed during World War II)[241] it is undoubtedly too restrictive of an interpretation for reasonable day-to-day application.

Within the Code, non-discrimination does not appear as its own distinct right, but instead within a series of concepts that a patient has the right to be free from. Right 2 states:

> Every [patient] has the right to be free from discrimination, coercion, harassment, and sexual, financial or other exploitation.[242]

Like the NZBORA, the Code refers to the HRA to determine what is considered discrimination, and the specific characteristics protected. The Code states:

> Discrimination means discrimination that is unlawful by virtue of Part II of the Human Rights Act 1993.[243]

While referred to explicitly in Right 2, freedom from discrimination could also be broadly construed within Rights 1 and 3. Right 1 requires the consumer to be treated with respect, and Right 3 requires respect for a persons' "dignity" and independence.[244] This results in an

---

[238] New Zealand Bill of Rights Act 1990, section 3
[239] *Ransfield v Radio Network Limited* [2005] 1 NZLR 233, (2004) 8 HRNZ 185 (HC)
[240] Andrew Butler and Petra Butler, New Zealand Bill of Rights Act: A Commentary, 2nd ed (Wellington: Thomson Reuters Press, 2015) at 11.5.
[241] At 11.5.1
[242] Right 2
[243] Section 4 "Definitions"
[244] Health and Disability Commissioner (Code of Health and Disability Services Consumer's Rights) Regulations 1995, rights 1 and 3.

overall protection that is both robust, and capable of being very broadly construed throughout a variety of circumstances.

As both the Code and the NZBORA rely on the HRA for its definition of discrimination, the application of these rights is largely the same in practice. As a result, the next section will largely discuss the test for discrimination in the context of the NZBORA due to the greater volume of jurisprudence available.

#### 5.2.2.1 Application

This section will provide an overview of the requirements of what it means to be "discriminated" against within a legal sense; how does one determine discrimination has occurred? Within healthcare, it is rare that overt, explicit discrimination has occurred; doctors are unlikely to shout racial slurs at a patient and deny them treatment for this reason.[245] Instead, a short inquiry must occur to determine whether the situation amounts to discrimination.

In *Quilter v Attorney-General*[246] it was held that "to differentiate is not necessarily to discriminate."[247] The case referred to what it called "permissible differentiation" and "impermissible differentiation"[248], a clear reference to the idea of fairness which will be discussed next. In *Ministry of Health v Atkinson*[249] this idea was modified and expanded into a simple two-step test. To determine that conduct is discriminatory, it must be shown that:

1) The defendant created a distinction based on a prohibited ground; and
2) The distinction caused a disadvantage.

For step (1), the person alleging to have been discriminated against (the "claimant") must be compared to another person in a similar situation to determine if a different experience has occurred. Then it must be determined if the prohibited ground is the cause of that different experience. It was cautioned in this case that when identifying the comparator, it was "critical to identify the correct and appropriate comparative group […] comparing apples to

[245] And if so, it would be of little value to this thesis' discussion due to there unlikely to be any issues of application in such a case.
[246] [1998] 1 NZLR 523
[247] At [54]
[248] At [35]-[39]
[249] [2012] NZCA 184

apples."[250] Because conduct is highly contextual, especially when factoring in different kinds of prohibited grounds between situations, it will often result in the comparison being made between two very specific persons.[251] The courts have not yet developed a specific test or framework to formulate a comparator, and it is largely a case-by-case hypothetical determination. The specificity of the characteristics of the comparator may be a determinative factor for judgments,[252] which may warrant special attention when factoring in the role an AI has played in determining differences. There has been concern expressed that this test is imprecise, particularly in respect of identifying a comparator. Emanuel has argued that the distinction between prima facie discrimination and justifications (or fair) decisions is unclear, and that this lack of clarity can result in an ineffectual protection against discrimination.[253]

For step (2), the disadvantage caused must result in "actual" or more than trivial harm.[254] Within healthcare this will often be the case. If treatment is denied for prejudicial reasons, both physical and mental harms are likely to result. But harm can also be financial (such as deliberately overtreating someone to incur expenses or loss of time) and reputational.

Previously, it was sufficient to show that the "prohibited ground is a substantial operate reason for the different treatment."[255] In *McAlistser v Air New Zealand Ltd,* it was held that this was no longer necessary. "By reason of" appears to suggest that the prohibited ground need simply be a "material ingredient" to the outcome.[256] In this regard, discrimination is about output, not process. A seemingly harmless, incidental action can result in discrimination occurring (often indirectly) due to improper consideration of different people. An example of indirect discrimination would be where a job listing includes requirements for height or hair length. These would be potentially discriminatory against a number of groups, such as women (both hair and height) or people of Asian descent (height), due to disproportionately affecting those groups. NZ does not require "stereotyping" or "prejudice" within its interpretation, unlike other jurisdictions.[257]

---

[250] At [55]
[251] See, for example *Purvis v New South Wales* (2003) 217 CLR 92 (HCA)
[252] *McAlister v Air New Zealand* [2009] NZSC 78, [2010] 1 NZLR 153, (2009) 8 HRNZ 801 at [34]
[253] A Emanuel "To whom will ye liken me, and make me equal? Reformulating the Role of the Comparator in the Identification of Discrimination" (2014) 45 VUWLR 1
[254] *Child Poverty Action Group Inc v AG* (2011) NZCA 397
[255] *Human Rights Commission v Eric Sides Motor Co* [1981] 2 NZAR 447
[256] *McAlister v Air New Zealand Ltd* [2009] NZSC 78
[257] Canada for example requires both of these requirements, see *Withler v Canada* 2011 SCC 12 [2011] 1 SCR 395 at [35]-[37]

Under section 19(2) of the NZBORA there is an exception of sorts to the right in s19(1): actions taken in good faith, for the purpose of helping a group of persons do not constitute discrimination.[258] Such an action is commonly called "affirmative action" and may be taken to try to remedy historical injustices or disadvantages. This is an area in which intention is important; not intended to be used as a defence, this is more a tool to identify what constitutes a breach. As *Coburn v Human Rights Commission*[259] said, an action which fulfils s19(2) must be "designed" to help the group at large, it cannot simply be a by-product or coincidental outcome. There are also reasonable limitations to the requirement of non-discrimination. Sometimes distinctions have to be made or are made for reasons that the courts have considered acceptable in context. For example, age-discrimination has been permitted on the grounds of administrative convenience in the UK when dealing with social security matters.[260] Similarly, resource allocation or limitations may be considered an acceptable reason to provide disparate treatment. As an example, some racial groups are less susceptible to some illnesses than others. The decision to do less testing on a person of Caucasian descent than on a person of African-American descent who is determined to be at higher risk, may be justified in an instance where there are not sufficient resources to perform both sets of tests on both patients, and resources are therefore better utilised in testing the higher risk group.

The degree of accommodation to be made for someone who requires distinct treatment is "reasonable"; if person A cannot treat person B the same as person C, due to B's personal characteristics, then A must make "reasonable efforts" to accommodate B.[261] This consideration is largely irrelevant to discussions of healthcare, as the situations it relates to are mostly subsumed within the reasonable limitations discussed.

Both of these previous points can be considered to be matters of "fairness." When is it fair to treat an individual differently than another, and when is it fair for deliberately differential treatment to be engaged in?[262] What is considered "fair" is highly contextual and the involvement of innovative technologies only further complicates the attempt to clearly define the concept. As an example, in [5.2.1] an example was given where COMPAS was alleged to

[258] New Zealand Bill of Rights Act 1990, section 19(2)
[259] *Coburn v Human Rights Commission* [1994] 3 NZLR 323
[260] *R v Secretary of State for Work and Pensions, ex parte Reynolds* [2005] UKHL
[261] *Smith, Valerie v Air New Zealand Ltd* [2011] 2 NZLR 171 at [15]
[262] Sandra Wachter, Chris Russell and Brent Mittelstadt, Review, Why fairness cannot be automated: Bridging the gap between EU Non-discrimination law and AI, 41 (2021) Computer Law & Security at 3-5

be racially discriminating against African-American parolees. In 2015, Sam Corbett-Davies and others commented in the *Washington Post* that the situation was not as simple as ProPublica had presented it, and that the decisions made by COMPAS were not necessarily discriminatory.[263] It is important however to recognise that a decision which is based on a prohibited ground, even a decision which causes significant harm, is not automatically unlawful discrimination.

As evidenced in recent HDC decisions, what constitutes "harmful discriminatory" behaviour is quite broad. In one instance, a doctor's insistence on "double-gloving" when providing a colonoscopy to a patient who was known to be HIV-positive was considered "confrontational, intimidating and predetermined."[264] The ensuing questions and confrontation between doctor and patient, where the doctor asked a number of questions with a "hostile tone" was indicated to amount to discriminatory behaviour, below the standards expected of a medical professional, and thus violating Right 2.[265] The decision indicates that the doctor was not believed to have adversely affected the patient's health (or contributed to his resultant death), but the communication was "inappropriate and insensitive" which amounted to unnecessary distress, humiliation, and agitation.[266] In this instance, the matter was referred to the Director of Proceedings who chose not to issue follow-up proceedings, due to the doctors already-actioned personal practice changes.

With the test and application of discrimination in mind, the next section will outline a number of issues that arise when attempting to apply the right to non-discrimination to situations involving an AI system.

---

[263] Sam Corbett-Davies and others, "A computer program used for bail and sentencing decisions was labelled biased against blacks. It's actually not that clear." (2015) Washington Post, (https://www.washingtonpost.com/news/monkey-cage/wp/2015/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/).
[264] Health and Disability Commissioner, Decision 03HDC13605 (18 June, 2004) at [15]
[265] At [19]
[266] At [25]

## 5.3 Issues in Application: Discrimination and Fairness in situations involving AI

In this section, the issues that arise when considering medical AI and discrimination will be discussed. The bulk of this discussion will seek to answer two questions:

1) Are there any issues with identifying whether discrimination has occurred when an AI is involved? And
2) Who is liable for the discrimination that occurs because of an AI's involvement?

This second question will include a discussion of whether holding the identified party liable is perhaps unduly harsh. It could be argued that a decision made by an AI was outside the identified party's foresight or control and holding them liable for it is imposing too harsh a burden on them.

Following these questions, two practical issues of AI and discrimination will be discussed:

1) the scale involved in claims as a result of the use of AI; and
2) the issue of time and changing biases in AI and datasets used.

This section will not concern itself with the technical aspects of AI that lead to discrimination, or that may exacerbate discrimination, but will instead focus on the features of AI that complicate applying the discussed components of discrimination.[267] In light of these issues, Chapter Nine will engage with the issue of prospective reform and its viability for discrimination.

### 5.3.1 Identifying if discrimination has occurred

As discussed in [5.2.2.1] the test for discrimination is concerned with the outcome of a situation. This section will highlight some of the issues associated with the steps of the test from *Atkinson* and *McAlister*, and some of the practical or conceptual hurdles that AI present. As mentioned briefly, there have been numerous examples in recent years of allegedly discriminatory AI. Two further examples include: Amazon's hiring algorithm, which was

[267] For a comprehensive discussion of those pitfalls, and their severity, see Sharona Hoffman and Andy Podgurski "Artificial Intelligence and Discrimination in Health Care" 2020 Vol 19(3) Yale Journal of Health Policy, Law, and Ethics at 12-22

shown to select male candidates at nearly three times the rate of female candidates during its vetting process,[268] and 'PredPol', which led to police unfairly targeting minority neighbourhoods because the system predicted that crimes were significantly more likely to occur in that neighbourhood due to racial crime statistics.[269] The question is, do these amount to legal discrimination under the NZ test? To re-iterate, the test contains two steps, that the action or decision in dispute:

1. Created a distinction based on a prohibited ground; and
2. The distinction caused harm or a disadvantage to the complainant.

For the purpose of this discussion, (2) will be omitted as this is a circumstantial concern and not one that can be discussed generally. Instead, this analysis will focus on (1) and dividing it into two sub-questions: did someone intend the discrimination (if yes, there is no issue), and (2) if lacking clear intent, can the distinction being based on the ground be proved?

#### 5.3.1.1 Intent

The first point to discuss is the role that intent plays in these inquiries. As mentioned in [5.2.2.1] intent is not determinative of whether discrimination has occurred; indirect discrimination, and entirely inadvertent discrimination, can result in a breach of non-discrimination rights within NZ. But it can provide useful guidance. Simply, if a person or party intended to make a distinction because of a prohibited ground, and the harm has occurred, then this is direct discrimination.[270] This section will briefly outline how intent may be relevant to situations involving AI. For completeness, before discussing the intent of other parties, a brief comment will be given on the idea of "intent" and an AI.[271]

---

[268] Jeffrey Dastin, "Amazon scraps secret AI recruiting tool that showed bias against women" (Reuters, 11 Oct 2018), available here < https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>. The reason for this bias appears to be that this system drew the conclusion that because men are disproportionately represented in the current staff population, they must therefore be disproportionately superior in some way.

[269] This is an example as well of circular biases; it is likely the case that racial crime statistics are themselves biased, and these further contributed to additional biased outcomes, and further fuelled said statistics. Kristian Lum and William Isaac, "To predict and to serve?" (2016) 13(5) Royal Statistical Society: Significance

[270] While this may appear to imply some requirement of malice, this is also not necessary. A doctor could choose to treat a patient of a particular race different from another, and cause harm because of poor treatment, due to a misguided belief but not because they view one race poorly.

[271] The intent of an AI would only be relevant insofar as the liability for this discrimination was imposed on the party responsible for the decision to utilise the AI (i.e. the hospital).

A) The intent of an AI

It has been argued that an AI itself cannot not be said to act with intent.[272] Intent is an inherently cognitive process, and far beyond the capabilities of any modern AI. AI act on the exact specifications and requirements outlined by their design; they are purely logical and mathematical minds, as opposed to the emotive, reactionary, and organic minds of people. As some form of action is required to deliberately apply the prejudices against the party, it could be argued that an AI is incapable of this as it only applies the data it has been provided as instructed. The actions of an AI could be likened to those of an automaton, and no actual action from which fault could derive has occurred. If intention exists in a situation involving an AI, it originates from somewhere else (for example, during the design of the system). Intention presupposes some form of awareness of consequences. While an AI can be said to "learn", what it is learning is how to perform its prescribed functionality more efficiently. This is simply the streamlining of an equation, as opposed to the development of an understanding as to *why* it is performing an action.

If an AI is incapable of acting with intent, any intent relevant to a discrimination claim must originate from another party involved. This will be discussed next.

B) The intent of other parties

This section will discuss the role that intent may play, when it originates from a person involved in the use, or development, of an AI system. This will be separated into two parts: the doctor (who uses an AI), and the developer (who creates or maintains the AI).

1) The doctor who uses an AI

If a doctor relies on an AI in treatment, and the AI has acted in a manner that amounts to discrimination, it might be questioned whether the doctor has also discriminated. If intention is required, can the doctor be said to intend to discriminate? The intent of doctors is unlikely to create an issue of application. A doctor follows the advice of an AI, which may be discriminatory, but they do so because they trust the science. Here, there is no intent, and the

[272] See Yavar Bathaee "The Artificial Intelligence Black Box and the Failure of Intent and Causation" (2018) 32:2 Harvard Journal of Law & Technology, or Thomas C. King and others "Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions" (2019) 25 Science and Engineering Ethicsfor a discussion of this in different legal contexts.

test is applied conventionally. Alternatively, a doctor chooses to interpret the AI's decision in a particular way, or only provide the AI with specific information, so as to act on their own explicit biases.[273] In this case, the doctor has intended to discriminate, and there no issue with applying the test either. Neither of these situations result in anything novel.

2) The developer of an AI

How the data that led to the discriminatory outcome was implemented, curated, or utilised however is more complex. Opponents of machine-learning applications argue that such systems should be subject to higher scrutiny through the invocation of circumstantial evidence with the purpose of convincing courts to infer discriminatory intent.[274] This is based on the assertion that the mere inclusion of some data-points, such as race, gender or sexuality, in an AI's analysis demonstrates an a priori intent to give disparate treatment.[275] If the designer of the system included these points, they must then intend for them to be used to create distinctions. However, as mentioned already, medicine is somewhat unique in that these characteristics are explicitly needed in many circumstances. A critique of their inclusion or exclusion must be considerably more nuanced, and conclusions about intention equally so, than for, as an example, a banking AI that approves loans.[276] However, some have then argued that the more important determination for protected characteristics is the age of the data. Attitudes towards different characteristics shift with time, particularly towards women and different racial groups. The inclusion of data in an AI system from an era that should reasonably be known to be skewed could illustrate that the creators should reasonably foresee disparate treatment occurring,[277] although this is not to suggest it is evidence of an intention for it to occur.[278] Similarly, the decision to utilise an AI system (i.e. by a hospital procuring it) which is known to possess racial biases could be construed as evidence of an

[273] An example might be where a doctor only gives the AI system certain information which they know will result in subpar care, or a biased determination.
[274] Wachter and others, above n 262, at 15; Talia Gillis "False Dreams of Algorithmic Fairness: The Case for Credit Pricing" (2019) at 47-49
[275] David Coglianese and Lehr Cary "Regulating by Robot: Administrative Decision Making in the Machine-Learning Era" (2017) 1147(105) Geo. Law Journal at 23-24
[276] The inclusion of racial data into a system that approves bank loans would almost certainly be evidence of the designers intention to discriminate. But that same data being included into a system that identifies genetic illnesses is far less determinative.
[277] This point is discussed in more detail in [5.3.4].
[278] An inference of intention is discussed by Coglianese and Cary at 24. However, this is a distinct mental state and requires a more explicit action than to have "reasonably foreseen."

intention to racially discriminate. In that case, identifying comparators and the distinctions made would be easier with this inference of intention.

The next section discusses some of the issues that arise when considering evidence of discrimination from an AI and the problem that determining step (1) of the *Atkinson* test may have.

#### 5.3.1.2 Can an AI's decision be discriminatory? The issues of evidence and causation

If there no clear intention to treat two patients differently due to a prohibited ground, it has to be determined if a distinction between them has been made on such a ground by appropriate comparison. This leads to a clear problem for AI; evidence of how decisions are made, and the processes involved, are often opaque due to the Black-Box Problem (BBP).[279]

The BBP raises an evidentiary barrier to applying the test for discrimination to these situations. Why a particular decision was made, or how a particular piece of data was utilised (or if it even was) in reaching that decision, is an obscured and (often) unknown detail. Take the example shown by Northwestern University in 2019. Millions of African-American patients were found to be referred less often than Caucasian patients who were equally sick to programmes for improved care.[280] In this instance, the fact that race was the determinative factor was identifiable thanks for the sample size of millions of patients. Could this trend of racial disparity be identified in a sample of just tens, or even hundreds of cases? The information that highlights this disparity (in this case, race) may not be identifiable without a significant portion of cases to analyse. And secondary to this, the basis on which this decision is being made is not necessarily still accurately verified. The AI could be making the decision based on race, but it could also be making use of another data point, besides the patient's race, to create this disparity[281] or making this distinction for a (unknown) justified reason. In the latter case, the AI is therefore not being discriminatory, but actually fair and would

[279] Discussed in Chapter 2 at [2.3.1]
[280] Ziad Obermeyer and others, "Dissecting racial bias in an algorithm used to manage the health of populations" (2019) 25(366) Science
[281] While this is extremely unlikely in the given example, if the assertion accepted in these discussions is that AI are capable of drawing connections or identifying new patterns, vastly beyond the capabilities of human beings, then the possibility that a different factor is responsible cannot be ignored.

engage with the "[acting in good faith] to benefit or help" requirement[282] or simply be a fair use of the information involved. It has already been presented that some evidence exists for racial dynamics influencing the provision of healthcare. It has also been regularly proven that an AI can identify new patterns, processes and causes to illnesses, that human doctors could not. Because of these two details, unless an AI is sufficiently self-reporting it could be argued that one cannot conclude an AI to be acting discriminatory with absolute certainty.

The previous standard of the prohibited ground being "a substantial operating reason for the different treatment"[283] would undoubtedly be complicated by the BBP. it was necessary to show that the prohibited ground is a "substantial operating reason for the different treatment" which the BBP would undoubtedly complicate. One of the complications in the COMPAS example is that the software code utilised is a proprietary software, and the protections afforded to it make the software code difficult to analyse for verification. When combined with the BBP, this early requirement is almost certainly impossible. This requirement was somewhat lessened in the case of *McAlister*[284]where now the prohibited ground need only be a "material ingredient" of the outcome. What this means when considering a technology capable of identifying new patterns, or synthesizing decisions through a variety of proxies, is unclear. If an AI does not possess information on race, but it does possess information on a variety of other points which when combined appear to disproportionately represent a race, is race a "material ingredient" of this decision? Or is it simply the by-product of other material ingredients? This is what is known as proxy datapoints; a collection of datapoints that individually do not relate to a particular thing, but when combined do relate to the thing. How proxies relate to the ground itself is unclear and warrants further judicial guidance.

The use of physical and immutable characteristics is not an automatic indicator of discrimination within healthcare, due to the practical realities of bodies and their differences. And how an AI uses or determines to use these characteristics can be difficult to determine. As the decision making involved becomes further removed from explicit data points, and

[282] This leads to a broader issue of prescribing "good intentions" to a system that cannot act with intention at all. In theory an AI that is making a justified decision, even if it cannot be understood, is "good." But this is a question better served to a philosophical inquiry, not a legal one.
[283] *Human Rights Commission v Eric Sides Motor Co* [1981] 2 NZAR 447
[284] *v Air New Zealand Ltd* [2009] NZSC 78

begins to make use of more broadly associated connections, the issue of how discrimination can be accurately identified becomes increasingly more difficult to determine.

### 5.3.2 Liability for the occurring discrimination

This section will now discuss how it is determined who is responsible for discrimination that occurs. The presumption will be that it can be shown that discrimination did occur, however discussion will occur for the differences in situations with clear intent versus those without. The principal consideration is: who is liable for the discrimination that occurs?

This issue of liability will be discussed with two broad considerations in mind: firstly, how liability for discrimination is determined and any issues associated with this when an AI is involved, and secondly, whether this methodology is unduly harsh when an AI is involved. This second point will be considered as a matter of policy; while the purpose of non-discrimination protections is to protect patients, it may be the case that the application of this right becomes too onerous or harsh and has negative effects on practitioners that outweigh the protective benefit enjoyed. The outcome of this determination could be one of two things: either the party considered to be liable for the discrimination is changed to an individual or entity more able to bear this burden (i.e. shifting from the individual doctor to the hospital at large), or that situations involving AI make a more precise distinction between culpable and non-culpable discrimination.

Liability for discrimination conventionally rests with the individual or party responsible for its occurrence. In instances where intent can be shown, it is the person who intended to discriminate who is liable. In cases of indirect discrimination, liability rests with those responsible for the factor which led to the discriminatory outcome.[285] But AI complicates this otherwise simple application; where in the "process" did the thing which led to the outcome occur? This issue relates to causation and will be a recurring point of discussion throughout the subsequent four chapters. It is one of the fundamental barriers for regulating and managing the problems caused by any AI system; their complexity, and the sheer volume of

[285] For example, in the example given in [5.2.2.1] of hair-length requirements in hiring processes being indirectly discriminatory against women, the business which included that requirement would be responsible.

different parties and components involved in their actualisation. There is debate as to which link in the chain of development is responsible for bias. It has been commonly claimed that biases originate within data; the collection or representation of data is where the biases in methodology are most apparent. Often it has been found that "bias and discrimination may emerge only when an algorithm processes particular data"[286] and that its specific use will steer an AI towards these outcomes. The Centre for the Future of Intelligence stated that:

> "identifying and correcting such biases poses significant technical challenges that involve not only the data itself, but also what the algorithms are doing with it (for example, they might exacerbate certain biases, or hide them, or even create [new ones].)"[287]

However, this view has begun to give way to the new view that bias can be embedded into an AI system at any stage of its development. A system arrives at its final functionality through a synthesis of its development process, and any one of these steps could lead to unintended (or intended) consequences elsewhere. A recent example of this problems complexity is Google's visual identification algorithms which failed to distinguish between gorillas and black people. After three years later, this problem has not been fixed, but searches for "gorillas" in specific Google products have been disabled.[288]

Barocas and Selbst[289] identify five areas where bias can become embedded within an AI system. These are:

1) **How the "target variable" and "class labels" used by an AI are defined.**

   This means how the different datapoints utilised by an AI are labelled, distinguished and combined into the curated dataset used. Examples of how to avoid bias here would be to limit the ways in which particular datapoints, based on prohibited grounds, interact with the wider dataset. An AI that determines eligibility for surgery should be restricted from applying racial data.

2) **Labelling and training data.**

[286] As the UKHoL has stated, "several witnesses pointed out that the issue could not be easily confined to the datasets themselves, and in some cases "bias and discrimination may emerge only when an algorithm processes particular data" at [54]. This trend in understanding has seen changes over the recent years, particularly as the understanding of the curation of data has shifted.
[287] Written evidence from Leverhulme Centre for the Future of Intelligence (AIC0182) House of Lords, above n56
[288] House of Lords Select Committee on Artificial Intelligence, above n56, at 142
[289] Barocas and Selbst, above n227, at 5

This is the same area of (1) but expanded to the data on which an AI is trained or "practices" it's skills. This data is often a considerably larger set than that which it operates with on a day-to-day basis, and in some cases is even publicly sourced through services like Facebook.

**3) Collecting the training data of a system.**

Data that is being collected for (2) may itself be biased due to its methods of collection. Those responsible for determining the methods, or enacting them, may impart implicit biases into this process which lead to disproportionate representation, or undue stereotypes being embedded.

**4) Feature selection; and**

Related to (1) and (2), this is the ways in which the functionality of an AI is determined and codified. The base process by which the AI operates may be biased due to its preference for particular methods, or its choice of actions.

**5) Proxies**

Perhaps the most complex because it is rarely an intended component of the AI's development. AI have been shown to be capable of utilising different data points in conjunction to act as "proxies" for other data. A simple example would be the use of socioeconomic and regional data as stand-ins for race.[290]

As evidenced by this list, within just the development stage of an AI there are numerous points in which numerous different parties can influence the outcome of an AI. Because of an AI's capacity to learn, the AI that is first provided to a hospital, and that which makes a discriminatory decision, may not be the same system in practice.[291] If a doctor made the decision to use an AI, it would be unreasonable to suggest that doctor is responsible for the discriminatory determination made. This is especially true when it is considered that a doctor is unlikely to have the appropriate technical understanding of the system itself,[292] or be aware of the earlier involvement and impact of different parties on the AI's decision-making. Similarly, if a doctor is provided with an AI diagnostic system and told of its seemingly superior capabilities, their deference to its discriminatory decision is also not unreasonable. Or if the AI makes a wholly unpredictable decision (which in turn is discriminatory and causes harm) a doctor being held liable for discrimination makes little sense.[293] In both

[290] An example of this would be the PredPol system discussed above in [5.3].
[291] This touches on issues of "wholly same product" protections in consumer law and the liability involved in these disputes. There will be a brief mention of consumer protections in the area of negligence in Chapter Eight, but otherwise this is research beyond the scope of this thesis.
[292] This is discussed in Chapter Seven as a matter of "understanding" in informed consent. See [7.3.2]
[293] This would likely qualify as an issue of informed consent; did the doctor appropriately communicate this risk (at least the risk of unpredictable decisions, as opposed to discriminatory ones specifically). See [7.3.1.2]

instances the doctor would likely qualify as having made an attempt, acting in good faith, to help the patient, and not to discriminate against them.

It is a patient's right under the Code to bring a complaint of discrimination against the practitioner they feel responsible for it. But caution must be exercised by the HDC when investigating such complaints. Identifying who is responsible for the decision occurring, and also identifying who most reasonably should be held responsible for it, are issues that will require significant judicial debate in the future. The individual subject to a complaint, and the individual most appropriately liable, could often be different parties. It may be the case that some form of vicarious liability is relied upon, and which the hospital should take responsibility for such occurrences as it was the one responsible for the inclusion of the system and is more readily capable of mitigating the impacts of such actions against itself.

A recurring point raised throughout this thesis is that caution should be exercised when applying conventional legal mechanisms to situations involving AI. The application of rules which are too harsh may result in disproportionate burdens being placed on medical professionals, who are ill-equipped to manage them. And in doing so, may have a chilling effect on the uptake of, and effective use of, AI systems.

The next two sections deal with practical issues which arise from the use of AI, as opposed to doctrinal issues of application. In both instances, it is not suggested that these issues make the application of discrimination protections untenable, but instead that the system responsible for doing so is ill-equipped.

### 5.3.3 A practical issue: A new scale of claims

A practical issue facing the management of discrimination is the ability to enforce or remedy problems that may arise within a court, tribunal or other mechanism. Unlike issues of consent or negligence, there is rarely a singular actionable moment in which a patient identifies that they have been discriminated against unless that discrimination is explicit. Whereas a human doctor is capable of explicit discrimination (by words or conduct),[294] an AI

[294] The use of slurs, mocking comments, or even denial of treatment for prejudicial reasons being common examples. See Thomas Rohner, "Nunavut woman claims mistreatment in Ottawa hospital" (2020) CBC Canada, available at (https://www.cbc.ca/news/canada/north/nunavut-woman-ottawa-hospital-mistreatment-1.5801844) for one such example, where an indigenous woman was denied water in her hospital people and heard her nurses talking about her negatively.

system is not likely to even interact with patients it makes decisions about.[295] For human doctors, complaints of discrimination are lodged for specific actions, of a decidedly more explicit nature, against a specific individual. In situations involving an AI, it is far more likely that a trend of discrimination is identified after the fact through analysis of the outputs or progress of the system involved. By the time any bias is identified, the system may have informed or independently decided on hundreds, if not millions, of patient's health data. This means the right to avoid discrimination requires enforceability on a far larger scale than the responsible bodies are equipped for. Wachter, when discussing this in the context of the General Data Protection Regulation (GDPR) within the European Union (EU), said:

> Cases have historically been brought against actions and policies that are potentially discriminatory in an intuitive or obvious sense. Compared to human decision-making, algorithms are not similarly intuitive; they operate at speeds, scale and levels of complexity that defy human understanding, historically protected groups, group and act upon classes of people that need not resemble and do so without potential victims ever being aware of the scope and effects of automated decision-making. As a result, individuals may never be aware they have been disadvantaged and thus lack a starting point to raise a claim under non-discrimination law.[296]

On day-to-day application this means, if discrimination can be identified, patients themselves could be righted through a medical tribunal. However practical change or prevention for the future would be harder to then implement.

To mitigate this issue, some scholars in the USA have debated how legal frameworks of "disparate treatment" and "disparate impact" used to decide discriminatory decisions could be reorientated for algorithmic systems.[297] However their discussion is focused principally on employment law and on a considerably smaller scale of effect. Within the EU, a growing support for "statistical fairness" has been developing, with a call for more "intuitive" approaches to anti-discrimination within human rights legislation.[298]

---

[295] In current applications, diagnostic AI's are often used as support tools to enhance or augment the decisions of a doctor, that is still interacting with patients in the room themselves.
[296] Wachter and others, above n 262, at 5
[297] Andrew D. Selbst and others, "Fairness and Abstraction in Sociotechnical Systems" (2019) ACM Conference on Fairness, Accountability and Transparency, at 5
[298] Wachter and others, above n 262, at 25-27

Whatever solution may be favoured, the combination of smart healthcare and discrimination protections is likely to create an almost unprecedented volume of claimants. The pressure that this could create on legal mechanisms; both within the judiciary, and within conventional medical tribunal or complaint forums, is difficult to foresee. It may be the case that the concept of discrimination in healthcare needs to move away from the protection of an individual's dignity to a more collective obligation on providers to treat all of its patients equally. In this way a patient would not be responsible for complaining that they have been discriminated against, but audits would identify where a provider, such as a hospital, has given disparate treatment unjustifiably. This would then be resolved in the public sphere, as opposed to treated as a personal dispute that current Code violations are concerned.

### 5.3.4 Biases as a Fluid Concept

Another practical problem for discussions of discrimination protections and AI is the passage of time. AI systems make decisions based on data, and the reality is that oftentimes this data, to provide any statistically significant benefit, has been collected over several decades. The timing of the AI's decision, its development, the data curation, and the data's collection all create an immense timeline of potential ideological shifts that can impact the data. So, while discussion in [5.3.2] was about who imparted the bias, here the issue is of *when* that bias originated. The problem behind this is simple: bias is not a static concept, and identifying when biases arose, or when they may result in disadvantageous outcomes, is difficult to determine.

As an example, an AI is designed to identify heart problems within a patient. Its "brain" is made up of several intertwined algorithms, all informed by millions of points of data. For these data points to be reliable they must first be curated. This process involves identifying the potential biases within the data (its' collection, inferences, management, and potential uses) then applying these data points to a wider dataset, to ensure that further biases are not empowered or given rise to throughout this process. Understandably, this is a difficult process where issues can be overlooked or inadequately managed at any number of stages. But additionally, the time involved in both this process, and the original timescale needed to gather the data, can also create imperfect biases. For example, historical medical data disproportionately represents, and therefore benefits, male patients. As Janine Austin

Clayton, the associate director for women's health research at the USA National Institutes of Health (NIH) said, "We literally know less about every aspect of female biology compared to male biology".[299] This has led to decades of systemic problems for female patients, resulting in misdiagnosis or ineffective care.[300] If an AI system makes use of datasets that span a wide range of years (say, 50 years), which is a likely necessity in some cases to have an adequate volume of data, this data could be marred by this evolving attitude towards women and their treatment. Because medicine (both practically and, somewhat cynically commercially), is designed around the majority – or referent – group healthcare inevitably falls short for other groups.[301] And this then leads to inaccurate statistics or healthcare information for both groups over time. This inaccurate data then feeds into further research and developments, and the problem is exacerbated and looped into other areas. Many algorithmic fairness solutions, in effect, go on to replicate this problem by attempting to fit the non-referent group (i.e. women of colour) into the statistics of the referent group (men) which overlooks the reality of their differences and the impacts this may have.

Additionally, it is possible that data is collected by someone with an explicit bias and been long embedded into a system or process before this bias was identified. Correcting the data to erase this taint is a monumental task. Indirect discrimination in the past is also complicated in this fashion, although easier to identify than intentionally discriminatory studies. Any corrective measures that alter data collected historically, particularly involving different racial groups or gender groups, would be legally, politically, and socially difficult ground to tread.

---

[299] Gabrielle Jackson "The female problem: how male bias in medical trials ruined women's health" *The Guardian* (International Edition, 17 Nov 2019) at 4

[300] Kate Young, Jane Fisher and Maggie Kirkman, ""Do mad people get endo or does endo make you mad?": Clinicians' discursive constructions of Medicine and women with endometriosis" (2018) 29(3) Feminism & Psychology. Young's article not only discusses the diagnosis and effective identification of endometriosis, but also the effect of historical gender roles and biases towards attitude which influenced decisions to research, or even investigate claims by women, of their conditions, at 3-5; Michelle A. Rodrigues and Kathryn B.H. Clancy, "A comparative examination of research on why women are more underrepresented in some STEMM disciplines compared to others, with a particular focus on computer science, engineering, physics, mathematics, medicine, chemistry, and biology" (2014) University of Illinois Women and Gender Global Perspective, Vol. 2, at 15

[301] This is an example of a feedback loop bias that is discussed in Podgurski, above n 267 at 15

## 5.4 Application of the Three Scenarios

The application of discrimination is a highly contextual discussion which can result in the conclusions reached varying due to the slightest changes. As a result, this application is not intended to represent a comprehensive analysis of discrimination in practice, but simply to provide an exemplar of the difficulties of applying conventional non-discrimination principles to these hypotheticals, but possible, scenarios. This application will serve to inform the subsequent discussion of reform in Chapter Nine by illustrating the areas where application is most difficult.

For simplicity, the discussed occurrence of alleged discrimination between each scenario will remain the same, based on an unfortunately common occurrence within healthcare. This scenario is:

> The patient (referred to from hereafter as "B"), who is of Pacific Island descent, seeks health advice [either from a doctor at a hospital or from a mobile application] about an unknown, concerning illness and is denied emergency treatment. B develops serious symptoms shortly after, requiring hospitalisation. B learns that a second patient, "C", who is Caucasian, previously sought health advice for the same initial symptoms, and was admitted immediately. B alleges this disparity in care is because of a racial stereotype, after reading a study in *Frontiers in Pediatrics* that states black and latinx children are less likely to be classified as requiring emergency care compared to white or Asian children.[302] B brings a complaint to the Health and Disability Commissioner alleging a violation of Right 2 of the Code.

There is a practical complication to this discussion when comparing AI systems to human agents. It is possible for an AI system to make a connection that a human doctor would not identify and to therefore come to a different conclusion. This conclusion may at the outset appear discriminatory but would actually be both justified and correct. While a real possibility, this is more an evidentiary concern and would in theory be identified during dispute. This would constitute a "fair" use of biased information and would likely not amount

[302] Xingyu Zhang and others, "Racial and Ethnic Disparities in Emergency Department Care and Health Outcomes Among Children in the United States" (2019) Vol.10 Frontier Paediatrics

to real harm, especially in instances where this connection provides greater help to the patient.

Following the test laid out in [6.2.3], B would allege there has been a distinction made based on race or ethnic background, and the likely comparator would be C.[303] The harm is real and more than trivial as it required hospitalisation and B is now suffering from new, serious symptoms which arguably developed because of not receiving immediate care. From the outset, it is clear that discrimination has occurred. As NZ's common law interpretation of discrimination is concerned with the actual occurrence, in instances where the law applies there is likely to be liability. The specific discussion of the three scenarios will therefore discuss two matters: firstly, whether the law applies in that situation, and secondly, to whom is its application likely to affect and whether this is fair.

### 5.4.1 SN1: A doctor utilises an AI

For SN1 there are two potential versions of events, which require distinct discussion. Firstly, the AI comes to a determination and recommends B is given emergency care, but the doctor acting on their own pre-conceived bias overrules this decision. In this instance, the doctor appears to have acted with racial prejudice and discriminated against B. It would be appropriate to apply Right 2 in its current form and hold them accountable.

Conversely, if the AI determines not to treat B, based on a prejudicially curated dataset and the doctor is unaware of this when actioning the decision, the situation may be different. This is where potential issues with NZ's common law application of discrimination as a matter of outcome becomes apparent. The doctor in this situation not only did not intend to discriminate but was also acting on what they believe to be reliable information for the benefit of the patient. The fact that this information was in some way corrupted by an inherent bias is unlikely to be known to them[304] and they are presumably acting on their interpretation of the information provided. This would be especially the case if they had acted with appropriate diligence and attempted to verify the results given by the AI. If the AI was utilised as one of many diagnostic tools, and the conclusions were able to reasonably be

---

[303] *McAlister v Air New Zealand Ltd* [2009] NZSC 78

[304] Doctors are rarely to be involved in the development, management and implementation of the systems they utilise within their practice as a matter of both practicality and expertise. This does however illustrate the importance of a more diverse and professionally integrated development regime, which is discussed in Chapter Nine.

interpreted as consistent, relying on the output of the AI (which will often be considered "more accurate") becomes increasingly understandable. However, if a doctor made use solely of an AI for diagnostics, where other historically accurate alternatives exist that may have made this disparity in treatment apparent, the doctor could be said to have not acted with due diligence and demonstrated an insufficient standard of care.

While it is fair to recognise that discrimination has occurred, it would perhaps be unfair to hold the doctor individually liable for this on the common law application of discrimination. Whether it is better to apply vicarious liability to hold the hosting institution (the hospital) liable is unclear and warrants further investigation in a more focused context.

### 5.4.2 SN2: A lone AI interacts with a patient

This scenario is largely the same as for SN1; if discrimination has occurred, it is largely irrelevant how it came to be. In this scenario there is the added simplicity of no human agent being involved and thus the concern of fairness against a single individual is null. If NZ were to require intent as a component of discrimination, then this situation would cause a grave problem for the protection of patients. However, in treating discrimination as a matter of outcome, this is succinctly avoided.

In the event that a doctor referred a patient to the process that resulted in a lone AI treating and discriminating against them, it is unclear what the outcome would be. Under negligence, if a manufacturer gives warning to a doctor of the dangerous properties of a device, the doctor may be liable for failure to warn a patient of this (but not the manufacturer).[305] If a doctor was in some way aware of the fact an AI had a propensity for discriminatory outcomes, and they referred a patient that may be discriminated against to that AI, it is reasonable to suggest they have a duty to inform of that risk. If they have no awareness of the risks of the AI, and are led to believe that this is not a likely risk, then holding them liable for the referral is likely to be considered too harsh in practice.

For SN2, if the actions of the AI were discriminatory, and able to be identified, it would result in the controlling authority over the AI being held responsible. Holding a hospital liable for discrimination in this case is less problematic as it reasonably has the resources,

[305] *Buchan v Ortho Pharmaceutical (Canada) Ltd* (1986) 54 OR (2d) 92, 25 DLR (4th) 658 (ONCA)

expertise, and access to required information to properly consider the risks of implementing the technology.

### 5.4.3 SN3: A patient utilises a mobile application

When discussing mobile applications, it is very important to identify where the application was sourced from, and who operates or is responsible for it. If a consumer downloaded a publicly available, and privately developed, application from their mobile phones' app store, it is likely that the Code, and associated rights under the NZBORA, will not apply.[306] This would then be an issue of a defective product or service, and dispute would occur under consumer protection legislation such as the Consumer Guarantees Act 1993 (CGA) or Fair-Trading Act 1986 (FTA).[307]

If the mobile application was in some way associated with the main healthcare system, then the situation would be different. While not within the conventional environment of healthcare provision, this would still constitute care under the Code, and also meet the necessary requirements of associated legislation such as the NZBORA.[308] It would also not matter if the application was developed and maintained by a third party private company, so long as it did so for use within the regulated healthcare system.[309] In this case, the same discussion as with SN2 applies. Discrimination has occurred irrespective of the lack of a human agent being involved, and liability would fall onto the providers of the mobile application (and by extension those who commissioned and permitted its use, likely the Ministry of Health).

---

[306] The Code, 4 "Definitions": "provider" means a health care provider or disability services provider; New Zealand Bill of Rights Act 1990, section 3

[307] There is no explicit protection against discrimination in either piece of these legislations; one could suggest an argument that a healthcare application that is discriminatory is not "reasonably fit for any particular purpose" under section 29 of the CGA.

[308] NZBORA s3 states the Act applies to acts done "by any person or body in the performance of any public function", a service being provided as a component of the healthcare sector would likely fit this definition and be applicable. However, see *Ransfield v Radio network Limited* (2005) 1 NZLR 233 for the current law on the relationship of s3 and doctor-patient relationships which may alter this conclusion. This decision has been heavily criticised, most notably in Butler A and Butler P "The New Zealand Bill of Rights Act: A Commentary" (LexisNexis NZ, Wellington, 2015) at [**add**]

[309] This would be, as discussed in fn.83 the defining question: whether the application was executing a public function or a private, commercial function.

It is of note that this situation would be considerably more difficult to identify. It is unlikely that patients utilising a private mobile application for health purposes would be aware of what is being done comparatively with other patients. It is also highly unlikely that an application in this circumstance would engage in explicit discrimination. While Right 2 could be applied directly in this situation, it is possible that discrimination in these instances would go both unnoticed, and unremedied, without some form of external auditing or oversight.

## 5.5 Conclusion

Many of the problems identified with medical AI in respect of discrimination will be echoed continually throughout the remaining chapters of this thesis. It is apparent that the nature of AI itself is difficult to cohesively marry established legal doctrines and applications in the current era. On a practical level, the primary problem is simply one of capability; an AI can discriminate against a massive population of people, very quickly, and in potentially subtle, novel ways. Whether the legal or health sectors can be capable of both identifying and responding to this challenge is difficult to predict, however awareness of the problem at least allows for attempts at pre-emptive regulation to mitigate the risk. In terms of applying the current legal doctrine to AI, there is no issue due to discrimination's lack of concern about intent or malice. The way in which discrimination is determined objectively as having occurred or not, irrespective of the defendant's thoughts, allows it to be applicable to scenarios regardless of whether an AI is involved within the treatment. However, this does lead to a concern as to whether this application is fair. Holding doctors (or even hospitals) liable in every instance where discrimination could be arguably proven could lead to extensive and unduly harsh burdens on the healthcare practice.

Discriminations focus on outcome, as opposed to intention or design, is helpful in one regard: whether the output of an AI as discriminatory is both easy to measure and acknowledge as harmful, irrespective of the AI's non-cognitive nature. This means recognising discrimination as occurring is not the problem, but the conflict between this risk and the benefits of involving an AI become more apparent.

# Chapter 6: Privacy

## 6.1 Introduction

This chapter aims to outline the privacy issues related to the use of AI within medicine. The key question is simple: can the benefits of AI within healthcare be balanced with society's current conceptions of patient privacy? Several related issues arise when following this inquiry, some of which will be omitted for ease of focus. When discussing personal data of any kind, the conversation invariably arrives at a discussion of property rights and data ownership[310]; this is too broad of a discussion for this thesis and will only be mentioned in respect of indigenous data sovereignty, in order to acknowledge NZ's obligations to tikanga Māori under the Treaty of Waitangi.

Privacy is an intrinsic component of the medical landscape. Healthcare deals with a wide array of personal, intimate and often embarrassing information relating to the person. To maintain trust in the system, respect for the information collected and the process around its storage is critical. Patient faith in the intentions and purposes of collection of information will have widespread effects on the use and effectiveness of AI technologies. Gavaghan has noted that questions of privacy are at the fore when considering AI in health, particularly due to the international nature of privacy and the limitations on enforcement that this imposes.[311] The HoL stated that a number of their witnesses believed "AI provided added impetus to the need to better educate the public on the use of their data and the implications for their privacy" more generally.[312] Recognition of the importance of data privacy has increased as the world has entered a data-driven economy, and major corporate conglomerates have begun to enter not only AI development, but also the public health sphere, through their technological

[310] Cameron F. Kerry and John B. Morris Jr. provide an overview of the "data ownership vs. privacy" debate, as well as their view why data ownership is the wrong approach for such a debate, in "Why data ownership is the wrong approach to protecting privacy" (June 26 2019) Brookings, available at https://www.brookings.edu/blog/techtank/2019/06/26/why-data-ownership-is-the-wrong-approach-to-protecting-privacy/

[311] Media Release from eHealthNews.NZ, New Zealand Doctor (2018) Available at <https://www.nzdoctor.co.nz/article/undoctored/ai-health-raises-privacy-concerns>

[312] House of Lords - Select Committee on Artificial Intelligence, above n56, at [49]

developments. In this landscape, there are two competing interests to keep in mind: the societal interest in innovation and betterment of technology, and the individual interests of the people whose data will be used to achieve this development. The Prime Minister's Business Advisory Council (BAC) report *The Future of Work*[313] noted that "outdated personal data legislation […] may be preventing innovative solutions and business models for the healthcare […] sectors."[314] This will provide the framing question of the chapter: is the current privacy regime in healthcare compatible with progress of AI, and if not, where do the conflicts lie? Following on from this, a secondary question will become clear: is patient privacy, in its current formulation, able to be sufficiently protected while enabling the specific form of progress that AI enables to develop?

Discussions of privacy and consent share considerable overlap within medical scenarios. Consent is often the determining factor for whether a breach of privacy has occurred. A number of these overlaps will be simply stated and discussed in greater length within the discussion of consent in Chapter Seven, because the nature of the problem is more in line with that discussion. Highlighting these overlaps does not serve to be repetitive, but instead illustrate the interconnectivity of issues that arise from the inclusion of AI systems. The potentially conflicted, complicated, and disparate legal frameworks that might work for or against plaintiffs/complainants in the event of a harm creates more tedious and ultimately drawn-out legal proceedings.

This chapter will begin with an overview of the operation of NZ privacy law generally, illustrating both its protections in statute and its developing place within the common law. Following this, the more specific healthcare formulation of privacy will be outlined, along with what is meant by "patient data". The issues associated with AI for privacy will then be outlined, largely focusing on the issues of use and access of information. Selected issues that are broader than just the practice and application of privacy will be considered under their own headings after this, including: Māori data sovereignty, jurisdictional complications of the technology, and the incompatibility of privacy with AI technology. The chapter will close with an application of privacy to the three scenarios from [1.3.2.5], to illustrate some of these issues in action.

---

[313] The Prime Minister's Business Advisory Council The Future of Work report (2019) available at < https://bac-staging.beingbui.lt/the-future-of-work-report >
[314] The Prime Minister's Business Advisory Council, above n313

There are two different conceptions of privacy to make note of first. The reason for this is that privacy is sometimes considered a wrong with no harm; the patient or person does not know their information has been breached, and nothing malicious is done with this – what is the harm? This is a consequentialist view: the negative consequences that affect the person whose privacy have been violated are the concern. For example, a consequence might be an individual's insurance premium being increased because they are known to suffer a heart defect. Contrary to this is the deontological view that privacy is breached, and a wrong is committed, irrespective of any measurable harm. In this view, privacy is an innate barrier that should not be breached. A simple example would be that when an internet service provider downloads your internet browser history, with the intention of targeting advertisements towards you. They discover no discernible information in the search history and discard it. It is difficult to identify harm in this case, as would be required by consequentialism, but deontologists would consider the invasion itself to be the harm. Instead of attempting to promote one view over another, this chapter will consider both as valid approaches and treat concerns of privacy with a broad, generous interpretation. Breaches of privacy can either result in actual change towards the patient in healthcare, or simply result in an exposure of information that they would otherwise not wish to be exposed (a quality which is reasonable to ascribe to most health data). The harm can be abstract, personal, and intangible, but assumed to have occurred purely through its potential. In this case, inadvertent and automatic actions associated with an AI's nature then are of concern, which adds considerable complications.

## 6.2 Privacy in NZ legislation

Privacy is not found as an explicit right in either the NZBORA, or as its own right in the Code. Its omission from these two does not invalidate it within the law,[315] or limit its seriousness, and has been argued to be because of the difficulty with defining privacy as a concept.[316] This prominent view associates privacy with context and considers that privacy's

[315] New Zealand Bill of Rights Act 1990, section 28: Other rights and freedoms not affected: "An existing right or freedom shall not be held to be abrogated or restricted by reason only that the right or freedom is not included in this Bill of Rights or is included only in part."
[316] For a discussion of this, see *Hosking v Runting* [2005] 2 NZLR 1 at [92]-[95]

validity or protection is dependent on circumstance. This is due to contextual rules about the access, frequency of use, and purpose of use associated with the use of things like personal data of an individual. Privacy acts as a holistic concept within the law; it is a recognition of a number of intersecting desires, ideals and needs and is contextual in both its identification and application. This section will provide a brief overview of privacy law generally within NZ, and then outline the specific protections or recognitions of the concept in healthcare.

### 6.2.1 General Provisions on Privacy

NZ is a signatory to the Universal Declaration of Human Rights (UNDHR) and has ratified the associated International Covenant on Civil and Political Rights (ICCPR). Both documents contain an explicit right to privacy, with identical wording. They state[317]:

> Universal Declaration of Human Rights, Article 12:
>
> No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks.
>
> International Covenant on Civil and Political Rights, Article 17:
>
> 1. No one shall be subjected to arbitrary or unlawful interference with his privacy, family, home or correspondence, nor to unlawful attacks on his honour and reputation.
>
> 2. Everyone has the right to the protection of the law against such interference or attacks.

This right, or a comparable transformation of it, is not mimicked within the associated rights documents of NZ despite the fact that the NZBORA is intended to bring the ICCPR into domestic law. Instead, privacy is protected in a sort of two-limb approach: statutory protections that originate from the Privacy Act 1993 (PA), and tortious protection through a

[317] Universal Declaration of Human Rights 1948 Article 12; International Covenant on Civil and Political Rights 1966 Article 16

common law tort of privacy.[318] It has been argued that this tort of privacy has developed because the social environment in which the NZBORA and PA were written was not an appropriate time to try to establish a somewhat fluid and vague right.[319]

#### 6.2.1.1 Statute and guidelines

The PA covers information-based privacy concerns.[320] The impetus for the Act was the need to combat the rapid technological advances seen in the 1990s and early 2000s; [321] this makes the Act aptly suited for discussion in the new technological revolution of AI. Unlike many Acts, the PA does not prescribe a list of strict rules or offences. Instead, it outlines 12 "Information Privacy Principles" which govern the handling of "private information" by agencies.[322] An agency, widely defined, is "any person or body of persons, whether public or private, and whether corporate or unincorporated" that is not specifically exempted anywhere in the Act.[323] There are no exemptions for hospitals or medical facilities within the Act. Health information is explicitly included within the scope of the Act, relating to patients whether living or dead[324] and any professional Code of practice issues shall be given effect through this Act.[325] Under the authority of the PA, the Health Information Privacy Code 1994 (HIPC) was made (now 2020). This collated the provisions specific to health information and goes into greater detail than the PA or Health Act. To avoid redundancy, the specific principles of the PA will not be discussed here, but will be discussed instead under the discussion of the HIPC in [6.2.2.]. To avoid disparities in terminology, any individual whose privacy is of discussion will simply be referred to as a "patient".

If an individual feels the principles of privacy have been breached, the Act permits them to lodge a complaint with the Privacy Commissioner (PC) or through any codified right to

---

[318] The tortious right to privacy was formally recognised in *Hosking v Runting* [2003] 3 NZLR 385 and later in *C v Holland* [2012] NZHC 2155.
[319] *Hosking v Runting* [2005] 2 NZLR 1 at [93]
[320] Steven Penk and others Privacy Law in New Zealand (2nd Edition, Thomas Reuters, Wellington, 2016) at 50-51
[321] At 51-52
[322] Privacy Act 1993 Part 2, sections 6-11
[323] Section 2
[324] Section 46(6)(a)
[325] Section 46(6) specifically, however this is also mentioned at numerous other points throughout the Act.

complain in both the HIPC and the Code.[326] This complaint is then investigated by either the PC, or the HDC (if the complaint was brought under the Code), and the process undertaken is focused on conciliation as opposed to punitive measures..[327] In the event this complaint cannot be conciliated, it may be referred to the Human Rights Review Tribunal, although this is a rarity in relation to healthcare. Importantly, with one exception, none of the Information Privacy Principles within the PA or HIPC are enforceable within the court system.[328] The principles are instead handled within the specific realm of the complaint, such as healthcare. However, under the Health Act, the sharing of specific kinds of health information is a criminal offence.[329] These are contextually very specific and therefore a comparative rarity.[330]

If the consequentialist view is favoured, for conduct to amount to an interference with a person's privacy, it must cause some form of adverse consequence or harm.[331] This can be interpreted quite broadly, as harm has been found to include emotional distress, humiliation, embarrassment, and financial loss. The harm suffered is determined by the PC or HDC, in assessing an allegation of a privacy breach.[332]

The PA recognises that privacy is not an absolute or paramount concept; there are competing interests and factors which must be weighed against the patient's privacy to decide an outcome. Broader human rights, societal interests, and even international obligations or guidelines must be considered when determining privacy concerns within NZ. The PC can even go so far as to grant authorisations regarding private information use which would otherwise contravene the Act or guidelines, if it is satisfied public interest or benefit warrants it.[333]

---

[326] Section 66 & 66; Health Information Privacy Code 1994, Section 7, The Code of Patient's Rights, Right 10

[327] See Section 69(1)(b); Tortious privacy actions can result in damages, injunctions or other common law remedies, see *Hosking v Runting* [2003] 3 NZLR 385

[328] Section 11

[329] Ss 112J and 112Y-112Z.

[330] These specific provisions relate to information collected in respect of "NCSP" (Natonal Cervical Screening Programme) and associated screening programmes.

[331] Section 66

[332] The case notes and determinations made by the Privacy Commissioner are publicly disclosed within Privacy Law and Practice intermittently. The last review of these disclosures occurred in May 2018, accessible via LexisNexis.

[333] Section 54; this is in line with a common theme of NZ's private rights and their balancing act. The NZBORA includes a component of "public benefit" and societal value when discussing justifications or limitations of rights.

#### 6.2.1.2 The common law

Privacy has also, historically, been protected by proxy under the common law. No specific tort of privacy existed until recently, and instead other torts were relied upon to provide remedies for situations akin to privacy breaches. With respect to health records and data, this has occurred under the tort of breach of confidence.[334] In this instance, a patient discloses information in confidence to another (i.e. their doctor). There is an obligation on that doctor (or hospital in control of the information) to not breach the patient's confidence by publishing or disclosing that information for a purpose other than its original communication. A breach of confidence is a referenced component of medical ethics in NZ, and is therefore important to keep in mind when discussing privacy under the Code. However, a general application of the tort of breach of confidence will not be discussed in-depth.

Since the NZBORA was enacted, NZ has developed a tort of invasion of privacy. First recognised in the case of *Bradley v Wingnut Films Ltd*[335] and in *P v D*[336] and later affirmed by the majority of the Court of Appeal in *Hosking v Runting,*[337] this tort was recognised to operate alongside the developing statutory controls for privacy and NZs international obligations.[338] The Court of Appeal formulated a two-step inquiry:

1. The existence of facts in respect of which there is a reasonable expectation of privacy; and
2. Publicity given to those private facts that would be considered highly offensive to an objective, reasonable person.[339]

The earlier HC judgment of *P v* D included the qualifying requirement that "there is insufficient legitimate public concern in having the facts made public."[340] This however was not referred to in *Hosking*.

[334] *T v Attorney-General* (1988) 4 FRNZ 582 (HC)
[335] [1993] 1 NZLR 415
[336] [2000] 2 NZLR 591 at [34]-[35]
[337] [2003] 3 NZLR 385
[338] At [115]
[339] At [76]
[340] At [35]

The requirement of "highly offensive" has seen some disagreement in later judgments[341] however will be considered as current law for the purposes of this chapter. The nature of health information is such that any public disclosure is likely to be highly offensive regardless, so whether this is the necessary standard is relatively insignificant. There is also no doubt that information about one's health is of a sufficiently personal nature to result in an expectation of privacy.[342]

When applying the law of privacy in [6.6] the tortious test outlined in *Hosking* will provide the basis of discussion. The principles and ideals of the Code and HIPC discussed next will be used to inform the application of this test.

### 6.2.2 Healthcare

Within healthcare, the two Codes (of Patient's Rights and Health Information Privacy) act together to provide robust coverage of a patient's right to privacy. The two do not cover the same grounds however, and most of the applicable regulation is contained within the HIPC. The Code provides simply for the right to have one's privacy respected with no additional requirements or explanation, whereas the HIPC provides specific rules, guidelines, and exemptions for how this is achieved or managed. The specific details for both mechanisms are detailed below, first with the Code's generalised right, and then with the specific protections of the HIPC.

#### 6.2.2.1 The Code of Patient's Rights (The Code)

There is no distinct right to privacy in the Code and instead it is presented as a sub-section of Right 1, the right to be treated with respect.[343] Right 1(2) says:

> Right to be treated with respect
>
> […]

---

[341] *Television New Zealand Ltd v Rogers* [2007] NZSC 91, [2008] 2 NZLR 277 by Elias CJ at [23]-[25]

[342] Stephen Todd and others, Todd on Torts (8th ed, Thomson Reuters New Zealand, Wellington, 2019) at [17.4.03]; see *Campbell v MGN Ltd* [2004] UKHL 22, [2004] 2 AC 457 at [94]

[343] Discussed in Chapter 4 at [4.3]

(2) Every consumer has the right to have his or her privacy respected.[344]

Later within the Code, privacy itself is defined as:

> Privacy means all matters of privacy in respect of a consumer, other than matters of privacy that may be the subject of a complaint under Part 5 of the Privacy Act 2020 or matters to which subpart 4 of Part 6 of that Act relates.[345]

This means that while the overarching principles of the Code apply to complaints related to privacy, specifics are determined by the governing legislative regime in play – the Privacy Act, except where it is excluded. The PA then empowers the HIPC to act as the contextual mechanism for healthcare. This means in practice, the Code establishes the purpose, and the HIPC facilitates the execution of that purpose.

#### 6.2.2.2 The Health Information Privacy Code (HIPC)

The HIPC is authorised under the PA as the health information specific Code, to allow for more flexible and specific interpretation than the broader PA. The HIPC governs the specific principles of privacy within healthcare insofar as they relate to patient data or "health information."[346]

The HIPC outlines a series of rules, in a similar structure to the PA, as to how patient information (or health data) is to be collected, stored, and utilised. For readability, this information will homogenously be referred to as "data". Interestingly the issues discussed most commonly (both domestically and internationally), in respect of healthcare and AI coincide with a number of these rules and their intentions providing an apt framework for the proceeding discussion. The eleven rules within the HIPC provide a broad range of obligations, protections and duties in respect of patient data.[347] The rules deal with the collection,[348] storage and access of data,[349] changes to and private use,[350] and its disclosure

[344] Health and Disability Commissioner (Code of Health and Disability Services Consumers' Rights) Regulations 1996, Right 1
[345] Part 4 "Definitions"
[346] Health Act 1956, Section 22B
[347] Health Information Privacy Code 1994, Part 2 Section 5
[348] Rules 1 – 4
[349] Rule 5
[350] Rules 6 and 7

and anonymization.[351] These rules mirror the principles contained with the PA, simply with a more focused drafting for their application to healthcare.

### 6.2.3 Patient Data

Before discussing the issues associated with it, it is important to establish the parameters of what is being referred to as "data." As a general term, data refers to almost any piece of information that one could collect about a thing, such as: size, shape, age, origin, internal make-up and so on. Personal data is this same concept but is specific to information relating to any specific feature or characteristic of a person. Under the PA, this is referred to as personal information and defined as:

> "information about an identifiable individual; and includes information relating to a death that is maintained by the Registrar-General pursuant to the Births, Deaths, Marriages, and Relationships Registration Act 1995, or any former Act (as defined by the Births, Deaths, Marriages, and Relationships Registration Act 1995)"[352]

This is a broad definition, referring to any information that is specific to that individual. Following this, "health data" is data collected by the health system about a person's health, treatment, or information that in some way impacts this (such as employment, incidence history, and causes). In the HIPC this is defined as "information to which this Code applies under clause 4(1)"[353] which refers to[354]:

> (a) information about the health of that individual, including his or her medical history;
>
> (b) information about any disabilities that individual has, or has had;
>
> (c) information about any health services or disability services that are being provided, or have been provided, to that individual;

[351] Rules 8 – 11
[352] Health Information Privacy Code 1994, Section 3
[353] Section 3, "health information"
[354] Clause 4(1)

> (d) information provided by that individual in connection with the donation, by that individual, of any body part or any bodily substance of that individual or derived from the testing or examination of any body part, or any bodily substance of that individual; or
>
> (e) information about that individual, which is collected before or in the course of, and incidental to, the provision of any health service or disability service to that individual.

This therefore covers a wide range of information about a person, not just about physical characteristics but also about their personal circumstances such as family, workplace, or personal activities (such as weekend occurrences like parties or drinking).[355] For the remainder of this thesis, precision will be sacrificed in lieu of readability; any piece of information which may warrant care or control within healthcare will be referred to simply as "data" without distinguishing whether it is health, personal, or another category.

#### 6.2.4 Conclusion

The protections afforded to privacy in NZ healthcare are designed to be as flexible as possible. As opposed to rigid rules and requirements, the codified protections afforded to patient data are based on principles and ideals, which allow for a more purposive and generous application to innovative situations. What is clear from both these principles, and the common law formulation of the tort of privacy, is that what is of concern is data which a patient should reasonably expect to be private.

With these rules in mind, the next section will outline where privacy issues arise in the use of an AI, and the technical challenges presented for these principles and their continued recognition. Due to the principled guidelines approach of privacy in NZ, this chapter will primarily focus on areas of concern for a patient's privacy, as opposed to actual application problems that arise from rigid tests like within discrimination or negligence.

---

[355] One of the unique components of healthcare is that even non-physical or mental aspects of a person's life are still of relevance. As highlighted in discrimination previously, healthcare has the most wide-reaching, justified need for information about a person and their life. This creates an immense variety of data that needs to be protected.

## 6.3 AI and Patient Data:

This section will now outline a select few of the privacy issues that arise from the use of AI. This is not an exhaustive list, nor will the section engage with each of these issues in depth.[356] Instead, the intention here is to highlight the wide array of privacy complications that an AI's use generates, and how these are often intrinsically linked to their purported benefits. The issues to be discussed can be divided into two broad categories: firstly, issues arising out of the capabilities of an AI, and the complexity of managing this, and secondly, issues arising from the by-products of AI and their control (principally the immense volume of data collected).[357]

Modern technical systems are capable of storing a tremendous amount of data. The use of "cloud computing" systems to permanently store data outside of the machine, and allow access to this data through networks, creates an immense landscape of potential issues for discussion. A common theme throughout this section is that these concerns are not unique to AI; oftentimes there are already principles or controls in place to manage these exact problems. However, the nature of AI and the scale of its capability, shifts these issues into the realm of the unmanageable. As discussed in Chapter Five, scale is a problem when discussing privacy. Unlike consent, or negligence, the wrongful conduct is often not something that can be identified at the time of occurrence. Privacy violations are identified later, through either leaks, public disclosures, or investigations. By the time this occurs, potentially thousands of potential claimants exist, which means singular HIPC-style complaints become not only an insufficient, but also an ineffectual, method of resolution.

### 6.3.1 The issue of re-identification

When storing and curating data that involves a number of pieces of information about an individual, or when multiple datasets are available to the same party, a common form of

[356] Many of these issues intersect with different components of the law, such as consent, contract, technological regulation and more. This results in many of these issues requiring a singular focused inquiry into each to appropriately investigate each. The intention of this thesis is instead to focus on overarching themes between different components of the Code.
[357] The AI revolution is often colloquially connected to the "Big Data" movement, heralding the use of large-scale data driven metrics for a variety of fields and uses.

protection is called de-identification (DI) or anonymisation.[358] This is the process of removing identifying features (that are not necessary for the purpose the data was collected for) from data to anonymise the individual that the data is based on. DI aims to turn personal data into purely statistically relevant information that cannot be used to directly identify any one person.[359] However, the ability for AI to recognise patterns from wider datasets creates the issue of "re-identification" (RI). Here, a patient can be reconstructed from a dataset, to undo the effect of their anonymisation. A recent example occurred in Chicago. In December 2016, The University of Chicago Medical Center (UCMC) entered into a research partnership with Google, which would allow the latter to use de-identified data from the Centre's electronic patient health records (anonymised to the standards required by HIPAA[360]) to improve on its predictive analytics. In June 2019, a class-action suit was filed in the Northern District of Illinois District Court against The University of Chicago, UCMC and Google, arguing that the information provided was not sufficiently anonymised because of the wealth of information Google independently had access to. This suit was dismissed due to the patient who initiated the suit failing to demonstrate the damages they had suffered as a result of Google's partnership.[361]

The HIPC does not contain specific guidelines or rules about the standard of anonymisation for NZ data, a feature that is common in other jurisdictions. Under a number of rules relating to the use of, or access to, collected information the HIPC states that the information:

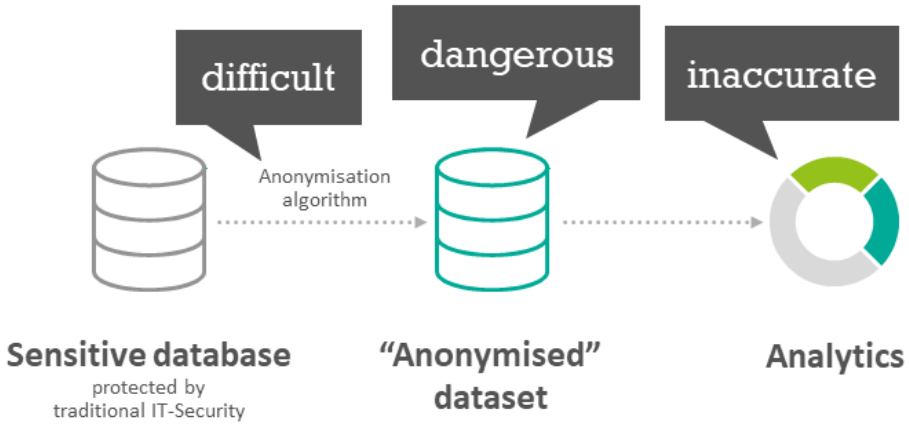
> Will not be used in a form in which the individual concerned is identified; or
>
> […] will not be published in a form that could reasonably be expected to identify the individual concerned."[362]

The rapid technological development of AI however presents a situation in which "reasonably be expected to identify" is a much more difficult standard to reach. RI often

---

[358] University College London maintains a comprehensive guide on anonymisation and how it is best achieved. Available at <https://www.ucl.ac.uk/data-protection/guidance-staff-students-and-researchers/practical-data-protection-guidance-notices/anonymisation-and>
[359] What information is included is highly contextual depending on the use of the AI.
[360] The Health Insurance Portability and Accountability Act of 1996 (HIPAA)
[361] *Dinerstein v Google LLC* (2016) No. 19 C 4311; the Illinois system for privacy breaches is a consequentialist system, without sufficient damages able to be illustrated, the supposed breach or ability to do so is insufficient.
[362] Health Information Privacy Code 2020, Rule 2(g), Rule 10

happens due to the involvement of a third party, such as Google, who has access to different datasets. The combination of their own data (such as publicly available search engine information) and the supposedly de-identified patient data means they can then "re-identify" the patients involved. The degree of anonymisation or "siloeing" (the separation of datapoints into distinct sets to eliminate connections between them) necessary to avoid this is extremely difficult, especially if trying to avoid ruining the quality of the data.[363] Nicholas Sartor explains that the analytical value of data is forfeited in exchange for true anonymity, and any analytics generated will be considerably less accurate than "open" data counterparts.[364] A simple example of anonymisation, and its shortcomings, is shown below, in a diagram created by Data Science Central.

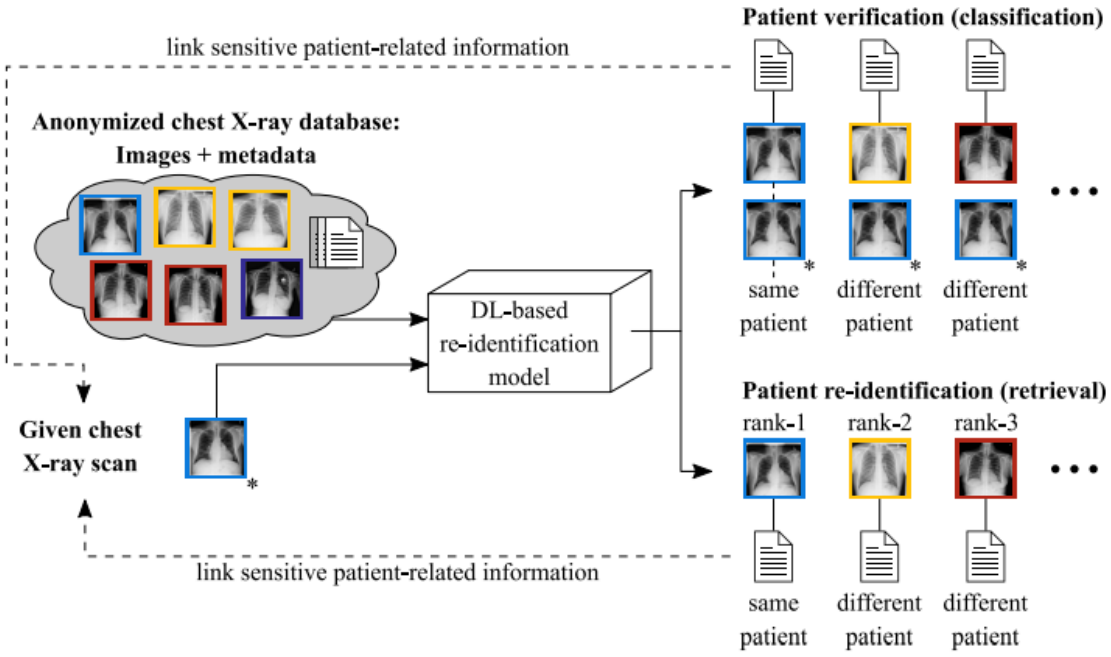



In the Chicago situation, HIPAA permits disclosures of individually identifiable health information if "reasonable" to believe the data cannot be used to identify the patient, the same as the HIPC. This means, if a patient cannot be identified via the given data, that data is not then subject to HIPAA rules, whereas for the HIPC it would engage the exceptions to barred use under Rule 10 and allow the data to be shared. Similarly, under Recital 26 of the General Data Protection Regulation 2016 (GDPR), anonymous information is defined as "information

[363] See for discussion Luc Rocher, Julien M. Hendrickx and Yves-Alexandre de Montyoe, "Estimating the success of re-identifications in incomplete datasets using generative models" (2019) Vol.10 Nature Communications, their conclusion is that "reasonable de-identification" is nigh impossible to achieve without jeopardising data usefulness, at [15].
[364] "Data Anonymisation Software – Differences Between Static and Interactive Anonymisation" (2019), available at <https://www.datasciencecentral.com/profiles/blogs/data-anonymisation-software-differences-between-static-and>

which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable". The issue with all of these similar protections is that the standard of what is "reasonably" anonymous, or "no longer identifiable" is a constantly moving (and perhaps already unfeasible) standard. It also relies on foreknowledge of what other data the third party has access to and how advanced their own AI systems are. Without knowing what data and skills Google possesses, it is not possible to determine if they can re-identify patients from the data they are gaining access to.

An example of re-identification created by VentureBeat when discussing the ability of an AI to reveal patient identities in the context of chest X-rays is shown below. In this instance, the study was intending to re-identify patients, simply to highlight that this was possible from "anonymised" data sets. Most re-identification would happen either inadvertently, or at a later point in time once the data has been used or transformed beyond its first purpose.




This details a situation in which a patients' X-ray scan is taken and then shared between a database (host) and a model that intends to re-identify the patient. The two columns on the right detail the different versions of the person that arise based on the scaling of

verification.[365] There are already algorithms in place to re-identify patients with prescription drug information within the USA.[366] And this process has been done manually by numerous groups throughout the last 20 years, even before the advent of many machine-learning systems.[367] Peter McOwan has said that AI systems have become noticeably better at automatically combining information from different datasets, to piece the puzzle together much better than people may have realised,[368] and Thereaux of the Open Data Institute (ODI) said, "AI is particularly problematic, because it can be used extremely efficiently to re-identify people [even when competing against] pretty good de-identification methods in places."[369] This is conceptually similar to the idea of proxies, discussed in Chapter Five; the process of using a mixture of information to replace the excluded or removed information and reconstruct a full picture. A by-product of AI's efficiency and large-scale datasets is that re-identification is both easy (in some contexts) and sometimes inadvertent.

Every jurisdiction has its own requirements for de-identification, and its own standards to which a patient must be made anonymous within datasets used for purposes beyond the treatment of that patient. NZ has received criticism for its perceived lack of attention in this area.[370] The concern is that a single AI systems' learning may be based off data sets from numerous sources, and of numerous types, to create the most in-depth learning environment for its algorithm. All of a patient's data, across a number of different fields, disciplines and specialities is collated and available within a singular system.[371] As a result, the niche

---

[365] Kyle Wiggers, "'Anonymized' X-ray datasets can reveal patient identities" (2021), available at <https://venturebeat.com/2021/03/16/anonymized-x-ray-datasets-can-reveal-patient-identities/> The purpose of this study was to highlight that this is possible; the intention to do so is not necessarily going to be present in many real-life situations, but the possibility of doing so should warrant concern.
[366] Christine Porter, "Constitutional and Regulatory: De-Identified Data and Third-Party Data Mining: The Risk of Re-Identification of Personal Information" (2008) 5 University of Washington Shidler, Journal of Law, Commerce & Technology, at 5; For a detail of an effort in the mid-1990s, see Paul Ohm, "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymisation" (2010) UCLA Law Review

[367] For a detail of an effort in the mid-1990s, see Paul Ohm, "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymisation" (2010) UCLA Law Review
[368] House of Lords Select Committee on Artificial Intelligence, above n56, at [49]
[369] At [61]
[370] The New Zealand Privacy Foundation has provided a criticism of NZ's regime and ignorance towards data anonymisation. See "The Ignorance of Anonymisation to Protect Privacy" (2015) at <https://www.privacyfoundation.nz/wp-content/uploads/2020/09/The-Ignorance-of-Anonymisation-to-Protect-Privacy.pdf>
[371] A centralised system is a common feature of health systems globally, the NHS is a notable example of a system which was the focus of the House of Lord's discussion on healthcare too.

anonymised markers within patient data could be pieced together, as a sort of patient jigsaw, to recreate a profile of a specific patient. One of the most pressing problems in the healthcare context is that often the data about patients is of immutable, and often unique, features of an individual and their history. The problems of DI are less prevalent when contained exclusively to healthcare; doctors in different departments having this information pieced together is of little effect. The risks arise when those responsible for, or who have access to, the AI system beyond healthcare receive these mosaics of patient information.

DI is a complex balancing act. While it may seem a simple solution to remove all identifying characteristics from a dataset, and to isolate datasets which require those characteristics from the others, this creates its own issues. The potential usefulness and flexibility of the datasets become stifled in response, and the interconnected benefits of learning systems is hampered in the trade-off for personal privacy. Allowing the association of distinct datasets to identify patterns unbeknownst to human doctors is one of the driving benefits of these systems; to do so would require a wilful concession that individualised patient privacy is not the paramount concern anymore. The theory that AI would make de-identification impossible is one that is not yet proven and has its own functional challenges to be addressed. But the rapid development of the technology thus far appears to suggest that requirements of "reasonable anonymity" amongst data are no longer possible to guarantee, and therefore makes the requirements of regulations like the HIPC unable to be met.

However, RI, and similar pattern-recognition processes, present a hypothetical situation in which modern conceptions of privacy are no longer feasible with the technology being developed and a judgment call must be made. This of course does not suggest that privacy as a concept is null and void, but instead that some of the established boundaries or requirements to privacy need to be either retired or modified as a result. Or alternatively, that instead of attempting to create stronger methods of anonymisation, the focus should be shifted towards different forms of control.[372]

[372] This will be discussed further in Chapter Nine, during the discussion of reforms.

### 6.3.2 Predictive Medicine

Oftentimes touted as one of the goals of globalised, smart medicine, an AI system presents a real prospect of "predictive medicine".[373] Predictive medicine, simply defined, is making decisions about healthcare pre-emptively, based on identified patterns or connections that do not yet illustrate the complete picture.[374] It is the statistical analysis of the likelihood of diseases or illnesses in order to attempt to act earlier to lessen their likelihood or impacts. As an example, the American Journal of Psychiatry featured a study in which researchers attempted to predict suicide attempts and deaths following an outpatient visit, based on electronic health records.[375] Their predictive model accurately predicted suicides within 90 days of the outpatient visit and identified the strongest correlated predictors of harm.[376] Predictive medicine is possible without the use of an AI based on conventional medical novel and skills, but AI offers the ability to identify connections faster, more accurately, and in areas previously not considered.

As an example, an AI system being used within radiology could feasibly identify a patient that has Parkinson's disease by detecting minute tremors in the patient while they are laying still for their scan.[377] This prediction could be made long before Parkinson's would be normally diagnosed, or could even lead to further discoveries in respect of the patient's deterioration. Two scenarios in which privacy issues could reasonably be predicted to arise are: firstly, that the patient was aware of the tremors and had chosen not to discover why they were occurring, or secondly, they were not aware of the tremors at all but want the ability to make choices about what they undergo or learn in their healthcare. In either case, the AI system has taken away the patient's ability to make the determination for themselves,[378] and has potentially violated their privacy by isolating and investigating this detail automatically.

---

[373] Christine Javid, "Predictive Medicine: Genes indicate diseases before symptoms do" (2004)
[374] This can be done either through surface level information, or through statistical analysis of surrounding factors ranging from physical attributes, to socio-economic status and work.
[375] Gregory E. Simon and others, "Predicting Suicide Attempts and Suicide Deaths Following Outpatient Visits Using Electronic Health Records" (2018) The American Journal of Psychiatry
[376] At 6
[377] Ekaterina Kovalenko and others, "Distinguishing between Parkinson's Disease and essential tremor through video analytics using machine learning: A pilot study" (2020) Vol 21(10) IEEE Sensors Journal,
[378] The implications for this in respect of informed consent will be discussed in the following chapter, Chapter Six.

The scenarios described above raise the difficult problem of whether privacy can be breached when others make an inference about a patient, as opposed to when information is explicitly known about that patient. The point at which these inferences become an invasion of privacy is unclear, however. As Nicholson Price II says[379]:

> If I were to believe that you are pregnant by stealing your OB/GYN's records or tapping your phone, that would clearly represent a privacy violation. However, if we are friends and I reach a belief that you are pregnant by seeing that you stop drinking when we go out for dinner, change your diet, and have put on some weight, it is hard to argue I have violated your privacy. The question is whether big data analysis is more like the former or more like the latter.[380]

Big data creates the ability to make more of these inferences and make more educated guesses. At some point, there is a reasonable argument to suggest an AI is peeking behind the curtain of an individual's life, simply because the standard to which it can do these things is new. This issue also overlaps with the discussion of consent that occurs next chapter, therefore, to avoid redundancy a more in-depth discussion of this will occur in Chapter Seven when discussing consent.

A further concern with predictive medicine is compounded by another issue discussed next in [6.3.4]; who has access to this information, and the awareness of patients in relation to this access. Access as a general concern will be discussed later, but here the issue is access to information that is potentially unknown to the patient entirely, that the patient has not chosen to find out about, and that is possibly not even true yet. If insurers or banks have access to this information, diagnoses could be seen to have discriminatory effects on the patient at a later point in time and be a violation of their control over their own data.

### 6.3.3 Use, Access and Security

Data collected about a patient, or from a patient, is conventionally to be used for their own immediate/ongoing treatment. A patient's blood is drawn to identify an illness, their

---

[379] Nicholson Price II is largely responding to the claim, cited from Jeff Scopek in Price and Cohen (see below, 380), that Big Datas value largely originates from its ability to draw inferences, as opposed to its explicit revelations.

[380] W. Nicholson Price II and Glenn Cohen "Privacy in the Age of Medical Big Data" 2019 25(1) Nat Med, at 9

DNA is sequenced to identify defects and so on. This section will discuss the concerns that arise from how this data is then used, stored and secured. This will include discussing who has access to the information, when they have access to it, and how far the use of this information might go beyond the original purpose of its collection. This section will also briefly make mention of the idea of "value". Personal data is a lucrative commodity in the age of AI development, and there are often times when public bodies want to capitalise on this value, but ensure respect of rights such as privacy in the process.[381] This is already a reality in a number of different fields and disciplines[382] and growth of a data-driven economy will continue to exacerbate this.

Chapter Four described the empirical research carried out to ascertain participants' understandings and views of AI technologies, and their place within healthcare. In Question 9 of the survey, participants were asked what information they would consider important to know about the AI, or its functionality, to allow for it to be involved in treatment. One of the most common responses was how their data was being utilised by the system both during and after their treatment. While there was an implied recognition of the fact the data needed to be stored and processed to aid in their care, what happened with this data after their illness was resolved was another matter. 43% of participants referenced their concern over data privacy and access, highlighting both a strong concern about one's own data and an awareness of the realities of AI's interconnectedness.

Any discussion of the use of data invariably results in a discussion of property and ownership rights.[383] Within this inquiry however the property element will be omitted. Whilst privacy and consent are both linked to ownership and rights of control, this leads to a far wider discussion than this thesis is attempting to engage with.

Patient information which becomes available to another business or entity can cause harm in the form of embarrassment, paranoia, or other mental pain. Additionally, it can result in financial or personal harms through the consequences of the information being revealed. Due to re-identification, this information could also result in harms that the patient is not even

[381] House of Lords Select Committee on Artificial Intelligence, above n 56, at [86]
[382] Beth Allen, "Information as an Economic Commodity" (1990) Vol 80(2), The American Economic Review, 268-263
[383] For one such proposal, see Ivan Stepanov, "Introducing a property right over data in the EU: the data producer's right – an evaluation" (2019) 34(1) International Review of Law, Computers & Technology, at 65-86

aware of, or based on information they themselves did not yet know.[384] Alongside this, if we consider that a patient's right to be informed is a continuing right, or even that a patient has the right to alter or control their data, this creates an immensely complex logistical problem.[385] While dataset protection and access is a major component of privacy laws, AI does present a unique landscape of parties and sharing to try manage. Many of the major parties involved in AI development are international conglomerates, and oftentimes these companies have financial stakes in banks, insurance firms, and even competing development firms.[386] These parties have financial incentives to use the information they access to the detriment of the patient (such as adjusting their insurance premiums). While the HIPC, Rule 6, details the access to information by a patient, once it has moved beyond the health system's immediate control this is no longer applicable. The inherent nature of AI means this will almost always be the case, as international businesses are often the only parties capable of facilitating the amounts of data needed, and the component control of it. This issue is representative of a wider problem, and another parallel with consent, which is how data is used after its original intention.

Often the concepts of privacy and transparency are inconsistent with one another, especially in discussions involving AI systems and sensitive data. "Open data" is a position which argues in favour of making all data-sets and the mechanisms behind their curation or management, open-source. However, as the HoL highlights, open data "cannot be the last word in making data more widely available and usable" and is often too "blunt an instrument" when sensible or valuable data is involved, which healthcare is most fundamentally.[387]

The conflict between development interests and social rights was recently highlighted in the UK, in what has been dubbed the "Royal Free Hospital/Deep Mind fiasco."[388] This was a deal between the National Health Service (NHS) and Deep Mind (a subsidiary of Google) to exchange patient data for technology access, the logic being that the NHS reasonably

---

[384] In the event of predictive medicine, decisions could be made about information that a patient has yet to have the chance to determine consent over, or how they feel about it being known.
[385] If a dataset contains over one million distinct individuals to analyse a particular health concern, notifying each of them and receiving consent, then modifying the set per reply, each transformation would be a laborious process.
[386] The largest known percentage investment by a firm currently is SoftBank, a Japanese based technology conglomerate dedicating nearly 98% of its investment into "AI ventures."
[387] House of Lords Select Committee on Artificial Intelligence, above n 56, at [8]
[388] At [286]-[290]

expected favourable (if not free) access to any AI-based products or developments that came from their patient data. Deep Mind agreed to this, and with the development of Deep Mind's Streams, the Royal Free London NHS Foundation Trust was given "five years" of free use of the system in exchange for testing the application.[389] The Trust provided the personal data of over 1.6 million patients as part of the trial test. When discovered, the Information Commissioner's Office (ICO) investigated and found that the Trust had failed to comply with the Data Protection Act 1998. Nicole Perrin further added that this instance shows a lack of cohesion and uniform negotiation and protections is highlighted by the NHS' fragmented Caldicott guardian system. [390] A system mirrored in the current (although soon to be reformed) DHB system. This highlighted the problems of data access and value when the Royal Free Hospital made public data assets available to a private corporation for free, in exchange for efficient technologies. As is often the case, public bodies lack the expertise or financing to develop technologies that major corporate bodies like Google are interested in, and the competing commercial and public values are a web that will be difficult to untangle.

One approach that has seen favour is that taken by Transport for London, whose data is used to identify traffic and public transport patterns of the city's residents.[391] Their data has been made available through a single point of access – only one method of retrieval and approval is therefore possible. They have made a publicly available list of terms and conditions for accessing, using, and sharing the data after receiving it which focuses almost exclusively on the privacy of individuals.[392] Another form of control, dubbed "secure systems" is an attempt at making a form of one-way data transfer "vault." The principle behind this is that data can be transferred into the system to curate, modify or supplement datasets but no data can then be returned. This creates a one-way mechanism for data and is a potential solution for privacy and misuse concerns as discussed by van Rysewyk.[393]

Finally, to speak of value, there is an incentive to gain every benefit out of data once it is collected. If properly anonymised, it follows to reason that a health system should be able to

[389] Oral evidence provided to the House of Lords, above n 56, by Dr Julian Huppert, at [286]
[390] Oral evidence provided to the House of Lords, by Nicola Perrin, at [287]-[292]; this fragmentation could be seen as mirrored in the now-defunct DHB system in New Zealand, which has now be consolidated into the singular HNZ.
[391] At [66]
[392] At [6]
[393] Simon Peter van Rysewyk "Machine Medical Ethics" (Matthijs Pontier (ed), Springer, Intelligent, Systems, Control and Automation: Science and Engineering, (2015)) at 318

sell this data to researchers or AI developers for their use. Nicola Perrin commented on the reality that NHS trusts were separately making distinct financial arrangements, with different corporate bodies, to use their datasets.[394] A problem highlighted here was the fractured nature of the NHS, which resulted in ineffectual negotiation of the data's value. As a result, this potentially violated or, at least, challenged patient privacy in exchange for inadequate value. This resulted in data being misvalued in terms of not only financial value, but also social worth, suitability for sale and application of purchase. Additionally, some districts were incentivized to "over-sell" their data, and further erode the standard of privacy available by not taking appropriate care or consideration in the process.[395]

This is a problem that may be avoided, in part due to NZ's merging of DHBs into the singular HNZ recently. Now, negotiations or value judgments made will be consistent, as opposed to fragmented.[396] While this may lead to less disparate instances of patient privacy invasions, the concern remains. This leads to the concern that those who control sensitive health data may be allowing access to a wide array of parties, oftentimes without proper consideration of value or protection. Additionally, those responsible for such negotiations (either in HNZ or the Ministry at large) are unlikely to possess the requisite knowledge of how AI utilise this data, or its safe curation, to ensure patient privacy is maintained throughout this process.

An obvious reaction to these concerns is to sharply limit access to health data. As with any solution, this comes with its own drawbacks, notably that limiting data access or control results in hindering innovation and development of further AI. This results in a somewhat paradoxical issue: why is the health system investing in AI systems, if it is going to harm the AI systems that it might also want to invest in? Additionally, controlling data too tightly can result in patchy care and coverage when a patient interacts with different providers or forms of medical institutions, if there is a risk their data overlaps with another dataset or institution. The consequent result of this is a regime that is privacy-focused, but method-secret. Big Data institutions could use the excuse of patient privacy as a justification for being more secretive of their methodologies, in an already notoriously opaque industry. This would only further

[394] House of Lords Select Committee on Artificial Intelligence, above n56, at [288]-[289]
[395] At [285]
[396] However, the exact interaction between HNZ and the Māori Health Authority is unclear, there may still be some disparity between these two organisations.

serve to hinder trust in the healthcare system, and its providers, particularly when innovative technologies with seemingly wondrous results are being employed.

There is no simple solution to this problem, however it is important to recognise the conflicting interests at play; patient privacy and innovation, better health outcomes and personal protection. Identifying what is the priority will be an important step in the coming years.

#### 6.3.4 Conclusion

The issues of patient data when an AI is involved are varied, although largely reflect the privacy issues of the past. The primary difference stemming from AI involvement is the scale involved, and the new capabilities that previously did not warrant inquiry. As discussed, a number of these issues are also issues within the context of informed consent, and this can (in theory) serve as a solution to much discussed here. In the next chapter, Chapter 7, this will be discussed in more depth. The remainder of this chapter is focused on some broader systemic issues associated with privacy, as well as a discussion in [6.6] about the possibility of managing many of these concerns.

## 6.4 Systemic concerns: Identity and Locality

The following sections will discuss two specific issues which do not fall specifically into the realm of patient data, but instead into wider concerns of the systemic application of privacy rights. These are included primarily for completeness, and to highlight that privacy and data issues are not one-dimensional or local, and instead engage with a wide variety of discussions happening simultaneously, and in different contexts.

The first issue is in respect of Māori data sovereignty, and how NZ's obligations under the Treaty of Waitangi may come into play with AI issues. The second is about the jurisdictional complications that arise from discussions of privacy and Big Data. This thesis is primarily concerned with NZ healthcare, but the reality is that Big Data is a global, and heavily interconnected undertaking. This means that violations of one's rights may not occur

domestically, or even be actioned by domestic bodies, further complicating the practical realities of enforcement.

### 6.4.1 Māori data sovereignty

New Zealand's commitment to tikanga Māori and being an inclusive, respectful space of Māori and Pasifika values adds an additional layer of complexity to privacy discussions. NZ healthcare is obligated to ensure that responsibilities under Te Tiriti o Waitangi (The Treaty of Waitangi) are fulfilled, and to respect the values that stem within.[397] However, the already existing complications of patient control of their own data, or of access to their data, is further complicated by recognition of Māori data as distinct, or in some way warranting unique input.[398] Māori data is sometimes discussed as a distinct, and unique, category of data within the NZ populace.[399] This is known as Māori Data Sovereignty (MDS) and is one example of recent movements in indigenous data sovereignty (IDS) seen around the world. State of Open Data, an activist group that amongst other things campaigns for IDS, argues that not recognising IDS is akin to "digital colonialism" and "co-opts indigenous knowledge and removes indigenous people from the discussion of data, and self-governance."[400] This is occurring alongside a shift of attitude towards data collection and privacy, particularly in the wake of revelations involving social media giants Facebook and Google, and data manipulation situations like Cambridge Analytica.[401] Cambridge Analytica utilised the data of 87 million Facebook users to provide analytical assistance to a number of political movements, including Brexit and the campaigns of Ted Cruz and Donald Trump.[402] DS

---

[397] Unique in the sense not of recognising indigenous values or cultures, as this is done for First Nations and Native Americans for example. Instead, this uniqueness is simply related to the fact Māori is an endemic cultural group, and its values are distinct and therefore result in its own application.

[398] Stephanie Carroll and others, "Data as a Strategic Resource: Self-Determination, Governance, and the Data Challenge for Indigenous Nations in the United States" (2016) Vol 8(2) The International Indigenous Policy Journal at 5-6

[399] AI Forum, Towards Our Intelligence Future: An AI Roadmap for New Zealand (2020), available at <https://aiforum.org.nz/wp-content/uploads/2019/09/Towards-our-Intelligent-Future_v1.01.pdf>, at 109

[400] State of Open Data, "Indigenous Data Sovereignty" (Accessed on 05/06/2021 at https://www.stateofopendata.od4d.net/chapters/issues/indigenous-data.html)

[401] One such example that resulted in legal action was in 2016, resulting in *Microsoft v. the United States* (2016) in which Microsoft was held accountable for its methods of international data sharing.

[402] Nicholas Confessore, "Cambridge Analytica and Facebook: The Scandal and the Fallout so far" (8 April 2018) New York Times

typically means that any data is subject to the laws of the country in which it is stored, which is problematic for any jurisdictional issues that arise as noted later in this chapter. Indigenous Data Sovereignty (IDS) is an extension of this concept which recognises that data is subject to the laws of the nation in which it was collected or originates.

While this form of data sovereignty is not necessarily a legally recognised concept, it's ideals and purpose has gained continued support across the developed world in recent years, especially within NZ.[403] Māori data sovereignty aims to respect iwi sovereignty and the role of iwi in their own health and autonomy. Article II of the Treaty of Waitangi states:

> "Queen of England agrees to protect the chiefs, the subtribes and all the people of New Zealand in the unqualified exercise of their chieftainship over their lands, villages and all their treasures."[404]

Māori argue that "all their treasures" includes all things which have tikanga or are *taonga* (precious treasures which should be given reverence) and this includes data. For many Māori, the collection and use of data from Māori communities or peoples, without their inputs, ignores that their data is a form of *taonga.* The Waitangi Tribunal has shown support for this, when it advised that Māori data should be included in its heritage, and thus should be included in the government's approach to safeguarding Māori cultural heritage and well-being.[405] In 2021, the High Court ruled in favour of Whānau Ora Commissioning Agency (WOCA). WOCA had requested health information on unvaccinated Māori in the North Island, which the MOH had declined. The HC acknowledged that this data was likely tikanga, but also necessary for the proper governance and care of Māori.[406] In order to abide by the obligations of the treaty, it was stated in the Royal Commission on Social Policy[407] that three principles are necessary: co-governance, co-design and co-innovation.[408] It is argued by prominent Māori scholars, such as Karaitiana Taiuru, that for this to be fulfilled, Māori require an equal say in, and control over, their data and information about their persons.

---

[403] AI Forum, above n399, at 105-106; Stephanie Carroll and others, above n398
[404] Treaty of Waitangi, Article II (English)
[405] Waitangi Tribunal WAI 2275 (2015)
[406] *Te Pou Matakana Ltd and Whanau Tahi Ltd v Attorney-General* [2021] NZHC 3319
[407] Report of the Royal Commission on Social Policy (Wellington, 1988)
[408] At 55

Māori data is defined by Te Mana Raraunga, The Māori Data Sovereignty Network, as:

> Māori Data refers to digital or digitizable information or knowledge that is about or from Māori people, our language, culture, resources or environments.[409]

Māori Data Sovereignty (MDS) is the idea that Māori have inherent rights and interests in respect of their data, and that this warrants some control or authority over its collection, ownership and application. And this sovereignty is actualised through Māori Data Governance which are the "principles, structures, accountability mechanisms, legal instruments and policies through which Māori exercise control over Māori data".[410]

How MDS will function is still being debated. For example, there is the question of whether MDS would be governed as an individual action, or through collective control by iwi or larger Māori groups. How the representative interests and views of *taonga* are managed amongst a wide variety of iwi and Māori represents a broad practical barrier at this time. Andrew Sporle (Ngati Apa) discussed Māori data issues in his "Hack Aotearoa" address[411] and provided a broad framework for what he considers appropriate guidelines for Māori DS. Firstly, he believes that health data should be aimed at promoting improvement and facilitating advancement of healthcare, rather than what he sees as a model "of highlighting deficits within a group."[412] Sporle argued this, as well as wider Te Tiriti o Waitangi values, could be instilled by the inclusion of Māori researchers in both development and implementation of these systems, or by including Māori in more leadership roles in research teams to ensure that these values are appropriately embedded from conception. Secondly, Sporle states that explicit awareness and maximisation of informed consent when data serves a commercial purpose (either at the time, or at a later stage of use) is critical.[413] This position has since been echoed by the AI Forum, that autonomy of data sharing is paramount – explicitly agency over one's own data and its uses.[414] Karaitiana Taiuru has echoed similar positions by arguing for a Māori data governance board to be established, which with

---

[409] Te Mana Raraunga "Principles of Māori Data Sovereignty: Brief #1" (October 2018), at 1
[410] At 1
[411] Hack Aoteraroa, NZ's 1st Artificial Intelligence in Healthcare Conference (Auckland: January 2019)
[412] A summary of Sporle's address can found in: AI Forum and Precision Driven Health, Artificial Intelligence for Health in New Zealand, (2020) available at < https://aiforum.org.nz/wp-content/uploads/2019/10/AI-For-Health-in-New-Zealand.pdf>, at 56-56
[413] At 56
[414] At 56

appropriate standing and approval by the Māori community would be responsible for achieving the goals of MDS.[415]

AI Forum has discussed the question of MDS in its report *Artificial Intelligence for Health in New Zealand*[416] calling for data collection to recognise Te Ao Māori perspectives in developing methodology, purpose and storage frameworks, as well as including the input of Te Ao Māori in the leadership of these decisions.[417] IDS aims to refocus the purpose of health development from benefitting private interests, or institutional interests, and instead benefiting the collective health of Māori and their communities.[418] Karaitiana Taiuru has argued in favour of localised data-storage within NZ, to ensure that Māori maintain access and the ability to exercise control over their data.[419] The practicalities of this are something Taiuru acknowledges are difficult, and this proposal is also discussed later in [6.4.2].

The purpose of discussing Māori and indigenous DS is to recognise not that it is a problem in and of itself, but instead that it adds another layer of complexity. Recognising IDS would require that the development, and management, of AI systems responsible for large-scale datasets or decision-making, to include Māori. It would require RI (and de-identification) processes to identify who is Māori and who is not, to ensure this information is not curated and circumvented. Additionally, it would result in an additional process for curation or correction of data and cause potential harm when international parties not subject to the Treaty of Waitangi are in control of or accessing their otherwise sovereign data. While it is crucial to respect tikanga Māori and its place within the country, ensuring that data is useable on scale without potentially excluding Māori data will be difficult when attempting to ensure equitable rules or regulation is enforced. How Māori may view their data as being appropriately curated may not align with Pakeha or NZ's wider societies', views. As a result, segmented data, or skewed and incomparable data, may result in differential treatment and the exacerbation of cultural lines. It is important to identify these risks in advance, and to

---

[415] Kairitiana Taiuru, Māori Data Sovereignty (2019), Available at <https://www.taiuru.maori.nz/maori-data-sovereignty-and-digital-colonisation/#Treaty_of_Waitangi_overview>

[416] Stephanie Carroll and others, above n398
[417] At 98; AI Forum, above n399, at 103-104
[418] At 106
[419] Greg Noone, "How New Zealand's Māori people are fighting for their data sovereignty" (2021) available at < https://techmonitor.ai/policy/privacy-and-data-protection/maori-data-sovereignty-new-zealand-indigenous>

attempt to manage them at the stage of design so as to not emphasise historical biases and disparities when these technologies become more widespread.

### 6.4.2 Jurisdictional Complications

A pressing problem for patient privacy is the reality that the data used, and the storage of data being used, often will not occur within NZ, let alone the hospital it is being used by. While privacy in NZ has its own protections and rules, these are of no effect when dealing with privacy breaches on an international scale. AI requires immense datasets, which are not only expensive and difficult to procure but also costly and equally difficult to store and maintain. The bodies capable of undertaking this task are few and far between, and they fall into three primary categories: governments, universities, and private corporations (notably within the USA, such as Google or Facebook). As Gavaghan pointed out, foreign health providers are not subject to NZ law or its protections, and how data is used once it crosses the border is not only unclear, but ultimately beyond any reconciliation.[420]

Private corporations represent the driving force behind much of AI's development, not just within medicine but in all sectors. Whilst sometimes referred to the Fourth Industrial Revolution, the development of AI has also been called the silicone gold rush,[421] or the new economy. Corporate bodies such as Google,[422] OpenAI[423] and IBM are just three examples of major industry players involved in the development, testing, and implementation of early AI. This results in a two-fold concern for patient privacy; not only is their data being held by a private corporation as opposed to the healthcare system itself, but this corporation is most likely based in an overseas jurisdiction.[424] The first matter calls back to the issue of trust once more; while a patient may trust the health system with their personal data, they might not trust that same information to Google. This can also serve to exacerbate fears associated with

---

[420] Media Release from eHealthNews.NZ, New Zealand Doctor (2018) Accessed at: <https://www.nzdoctor.co.nz/article/undoctored/ai-health-raises-privacy-concerns>

[421] This term has been used as far back as 1999 by Karen Southwick in Silicon Gold Rush: The Next Generation of High tech Stars Rewrites the Rules of Business (California Press, Stanford, 1999)

[422] The DeepMind project is one of the focus areas of the House of Lord's report Ready, Willing, Able? (2018)

[423] OpenAI is an Elon Musk started project that is attempting to research both AI and neural augmentations.

[424] Even when a corporate body has a NZ office or registration, often this is a virtual office and is not actively manned or involved in their operation.

mistreatment or discrimination, or corporate duress through insurers. The second matter leads to the unfortunate reality that in the event of a breach, or measurable harm to a patient, there is little to no recourse available. Holding a large corporate body accountable for breaches of privacy is difficult for an individual person, but it becomes impossible when no jurisdiction is held over that corporate body. The enforcement of rules at a local level is of course possible, and attempts can be made to regulate how data is shared or transferred internationally.

One solution that has been proposed is to host the data of a particular country only within that country.[425] For example, all NZ health data that is collected in a system managed and created by IBM would be stored only on NZ's shores, and not be permitted to be transferred. This would require, for example, Google to create unique data centres in each country it wishes to implement its DeepMind system that only holds data from within that country.[426] While this could be a beneficial move for both regulation and enforcement it is ultimately self-defeating as it would eliminate many of the benefits of AI due to greatly reducing the scale of datasets utilised by an AI. By reducing the size of the data available, the effectiveness of an AI to draw conclusions and develop its understanding is hindered. Much like the issues of controlling or limiting access, this stifles the benefits of the technology, in this case not only domestically but even potentially internationally too. International privacy protections exist, but are ineffectual when dealing with limited groups of people or isolated incidents and result in an alien and unapproachable problem for people.

Another potential avenue is the establishment of an international governing agency of AI. Such an agency would function similar to the World Trade Organization (WTO) or World Intellectual Property Organization (WIPO) to facilitate congruent development and implementation of AI internationally. Such a proposal is far beyond the scope of this thesis, concerned only with AI's interaction with the domestic healthcare system. This idea will be mentioned again in Chapter Ten, during the recommendations for future research.

With these two systemic concerns in mind, as well as the more specific issues discussed in [6.3], the next section will approach the question of whether personal data privacy is compatible with AI at all.

---

[425] This is an extension of arguments stemming from the enactment of the EU GDPR, however it has not seen widespread support thus far.

[426] DeepMind is currently based in the United Kingdom, and therefore interacts with the NHS.

## 6.5 Is personal data privacy incompatible with modern AI?

A final privacy-related discussion is perhaps a somewhat cynical one; is personal data privacy an incompatible concept with modern AI's direction? Angela Ballantyne of the Department of General Practice at the University of Otago has said there is a danger of "underutilizing AI solutions" in healthcare.[427] The risk is that by being overly concerned with social or legal barriers, systems will persist with "avoidable levels of error" and their associated patient harms; the benefits of these new systems will be mitigated, making their development and investment a waste. This is not to suggest that Ballantyne believes the concepts are incompatible, however it does provide a starting point for the discussion. Calling back to the earlier discussion of value judgements, this is an area in which it may be necessary in the future to wilfully concede some privacy-related protections or established barriers in favour of progress and better results.[428]

The 2019-2020 coronavirus/COVID-19 outbreak has yielded some interesting points of discussion around community willingness to adapt for greater societal benefit. Whilst not universal, public acceptance and support for both widespread community lockdowns, contact tracing, and testing highlighted a strong recognition of the benefits of sacrificing some established freedoms in exchange for both personal and community wellbeing.[429] The Centre for Data Ethics and Innovation (CDEI) noted to the HoL that there is oftentimes a disagreement amongst the public about how and where AI or data-driven technologies should be used, and what safe-guards should be applied to them.[430] Oftentimes there is recognition that these innovations come with trade-offs to security, privacy, and free-speech in exchange for better results, safety, and progress.[431] As COVID-19 continues to be an active issue in the world, and wide-scale analysis of its impacts, nation's responses, and public sentiment have yet to be extensively carried out, it may be too early to pass judgment yet. However, this

---

[427] Media Release from eHealthNews.NZ, New Zealand Doctor (2018) Accessed at: <https://www.nzdoctor.co.nz/article/undoctored/ai-health-raises-privacy-concerns>
[428] For an in-depth explanation of "value judgements" see Chapter 2's discussion at [2.6]
[429] Michael D. Kokkoris and Bernadette Kamleitner, "Would you sacrifice your privacy to protect public health? Prosocial responsibility in a Pandemic paves the way for Digital Surveillance" (2020) 18 Front. Psychology, discusses the effects of social sacrifice and "prosocial" narratives of healthcare and the risks they pose for eroding social rights.
[430] House of Lords Select Committee for Artificial Intelligence, above n56, at [350]-[352]
[431] At [353]

event could serve as a benchmark to highlight the willingness of a society to adjust over time, in the face of necessity, and perhaps less clearly, for innovative methodologies.

As discussed by Nicholson Price and Cohen, and mentioned numerous times elsewhere in this thesis, the balance at issue is between benefits and conventional protections.[432] It is possible to protect a patient's privacy in almost all circumstances: the imposition of total control over access or security, the prohibition of the continued storage of data, and other hard-line measures would (at least on longer timescales) prevent privacy breaches. But in doing so, the benefits of the employed technologies are lost. And in the cases of transferrable or inter-connected datasets, the benefits are not only lost for that patient, that hospital, or that type of medical procedure. They can also be lost on a wider scale for similar kinds of technologies; the learned benefits of an image-based radiology AI can be applicable in facial recognition technology for search and rescue as an example. Where the protective line is drawn, and which concessions society are willing to make, is a decision that will eventually have to made. And while concessions in one area may be necessary, they could be offset by strengthening protections in others.

## 6.6 Application of Three Scenarios

There are numerous potential ways in which a patient's privacy may be violated, or a patient may allege their privacy has not been respected. The communication of sensitive information could happen at a variety of steps, under different circumstances, and for different reasons. Any application that involves features of all the different possible breaches, or rule violations, would be too decidedly complex for this thesis. As a result, this application will focus on one, albeit extreme, potential breach relating to two main privacy issues: disclosure and re-identification. The discussion of this will largely follow the rules of the HIPC and the HDC's interpretation of respecting privacy, as opposed to the common law conception under *Hosking.* However, the ideas behind the invasion of privacy tort will be used to inform the discussion.

[432] Price and Cohen, above n 380, at 15-16

This application will assume that relevant data is held and used in NZ. As discussed in [6.4.2] the reality of data leaving NZ because of the involvement of international conglomerates or server hosting is one that requires careful consideration in the future, but is beyond the scope of this thesis. If patient data was transferred from NZ to the USA, where it was then publicly distributed, it is likely that liability would fall on the NZ party if they were aware this was the intention. If the American party did this discreetly and without the NZ party's knowledge, a similar issue as discussed in Chapter Five about the fairness of application arises. The lack of reasonable enforcement for the NZ patient also becomes apparent in this situation.

The framing scenario for this discussion is:

> The patient, "H", has been regularly updating their doctor on their condition after being diagnosed as HIV-positive. This information is [inputted or directly connected to] an AI system ("System X") which collates and analyses the information to create an adjustable treatment plan for H. This information is added to a wider dataset about HIV patients but goes through a process of de-identification first.
>
> H then begins treatment with a therapist for depression. They [meet with a therapist who utilises an AI, begins consultations with a help "chat-bot", or utilise a private mobile application for advice]. These options all utilise a different AI ("System Y") owned and developed by the same manufacturer as System X. This system is used to identify patterns of concern in patients with depression. Its dataset is a considerably larger one, which includes the data from System X as part of a broad analysis of the connection between illness and depression. The system, when applied to H, re-identifies him as an HIV-positive individual. This information is used in the analysis and treatment of his depression. H believes his right to have his privacy respected has been breached. This scenario presumes both health practices are covered by the Code regime.

This is an example of datasets being not only interconnected, but also utilised in ways either knowingly or not, between different spheres of the healthcare sector. There is a clear competing interest to H's privacy in this situation: the benefit of medical professionals having access to, and a complete picture of, a patient's wellbeing for diagnosis.

### 6.6.1 SN1: A doctor utilises an AI

There are two human agents whose use of data requires consideration. The doctor is inputting H's data into a system which takes it to a third party (the company), and the therapist is relying on information provided by the system. The HIPC makes a number of concessions or exceptions for reasonable belief or expectations; what each practitioner is doing and why will inform the application of those rules.

#### 6.6.1.1 Doctor with an AI

By entering H's data into the AI, the doctor is passing this information on to those who host the data, and then future users of the dataset involved. However, whether this constitutes "disclosure" of private information is dependent on the understanding of the doctor. Data that has been sufficiently anonymised was, prior to the advent of AI, considered to be appropriate for sharing for secondary purposes like research, education, and re-use.[433] Health information can be disclosed either with the patient's consent[434] or in a way in which the individual is not identified.[435] Additionally, the fact the data is being disclosed for the purposes of further statistical use (by the dataset within the AI), would also meet the exception of Rule 10(1)(e)(ii) HIPC:

> **1** A health agency that holds health information that was obtained in connection with one purpose may not use the information for any other purpose unless the health agency believes on reasonable grounds,—
>
> **(e)** that the information—
>
> **(ii)** is to be used for statistical purposes and will not be published in a form that could reasonably be expected to identify the individual concerned;

However, this would be reliant on it having been explained to H that this was necessary for the functionality, and continued development of the system in question. Presuming the mechanism of input involves sufficient steps to anonymise the data, it is unlikely the doctor

[433] Health Information Privacy Code 2020, Right 10(2)
[434] Right 11(1)
[435] 11(2)(c)(i)

would be found to have violated H's privacy because the standard is "reasonable expectation." The problem as noted is that this standard is perhaps insufficient for anonymisation given the realities of the technology, as discussed in [6.3.1]. This raises the question: can disclosure of this nature be reasonably justified on the current rules? Arguably, it is more reasonable to suggest that there needs to be recognition of this impracticality in the rules by allowing disclosure which may be able to identify the patient under specific controls.

This is of course a discussion which can be resolved by the consideration of consent. If H consented to having his information collected and stored in this way, with knowledge of the AI systems functionality, they would have both no reasonable expectation of privacy in this respect and have authorised the disclosure under the HIPC.[436] However, H might not have expected this information to be utilised within an entirely distinct context (mental health) which would be relevant to a discussion of informed consent.

This discussion focused on the disclosure of information, but the same discussion applies to the information's use by the AI manufacturer under rule 10. The major difference is perhaps that those responsible for the use of the data (in this case the manufacturer) would be less likely to be able to rely on their "reasonable expectation" that H stays anonymous. They are more likely to understand the impracticalities of this due to their expertise.

#### 6.6.1.2 Therapist with an AI

The therapist who utilises the information provided by the AI is perhaps in a different position. Presuming that H has not himself disclosed they are HIV-positive, it is his right to choose whether this information is included and utilised in the discussion with his therapist. The therapist did not necessarily intend to (or was not necessarily even aware they could) re-identify H. The discovery they were HIV-positive may be inadvertent within the normal course of their discussions. The choice to utilise this information is most likely what H would take issue with.

The therapist would perhaps suggest that the use of this information is necessary to prevent or lessen a serious threat to "the life or health of the individual concerned" (being H).[437] By

[436] Rule 11(b)
[437] Rule 10(d)(ii)

more properly understanding a holistic view of his health and wellbeing, the therapist can more appropriately judge his mental state. By knowing this information, the therapist has more information to work with. This is the conflict that often arises from the use of AI within healthcare; it allows for a better, more complete picture of health, but comes at the cost of the patient's control of, and autonomy over, their information. By not allowing H to choose when to disclose this information, it is likely the therapist has not appropriately respected his privacy. However, the inadvertent and perhaps unavoidable nature of this is likely to make this comparatively minor as a breach.

### 6.6.2 SN2 A lone AI interacts with a patient & SN3: A patient utilises a mobile application

Both scenarios 2 and 3 involve largely the same considerations, so will be discussed in tandem. The mobile application or AI system lacks the ability to rely on the "reasonable expectation" component of the HIPC. It itself does not make judgment-based decisions like this in regard to patient respect and understanding. Instead it applies the data empirically, as instructed to do so. This largely means the issue here amounts to how the functionality and continued use of the data works was explained to H as a matter of informed consent.

In respect of privacy, it cannot be said that an AI or a mobile application itself violated H's privacy. There must be a person against whom to action a complaint. The potential breach instead happens at the manufacturer stage where H is not only re-identified, but also then is shared with another system, in a different context. This likely violates Rule 10(1)(b) which HIPC requires information may not be used for another purpose, unless believed that "the purpose for which the information is to be used is directly related to the purpose in connection with which the information was obtained."[438] It is difficult to suggest that H's mental health is a purpose directly related to the treatment of his HIV, and secondly this exemption relies on the health agencies "reasonable belief" again. The manufacturer that utilises H's data for more than one purpose is unaware that they will be exposed to the second AI system, and therefore could not rely on this argument.

[438] Rule 10(1)(b)

The use of data for a different purpose, and its disclosure to allow this development and possibility, is on its face permitted when data is appropriately anonymised. The reality is that this is both considerably more difficult, and often impossible to properly determine in advance. The nature of AI in requiring large-scale datasets is that they need to collect, share, and re-utilise information in this way. But AI's capability of re-identifying patients in the process means that the common protection afforded by the HIPC is not only insufficient, but also realistically no longer applicable.

## 6.8 Conclusion

Privacy is a vaguely defined and fluid concept, and its protection is a heavily contextual task in the current age. AI presents the opportunity to not only breach a patient's privacy in new ways, faster and with greater precision, but to also to do so on an immense scale affecting thousands, or even millions, of people at a time. The issues identified in this chapter are largely related in their inherent association with the nature of health AI; they require large datasets, shared often between numerous bodies or departments, and have the capability of connecting the dots between points contained within with remarkable precision. In doing so, established protections or guiding principles of privacy are thwarted simply by being not only logical insufficient, but credible protections being a technical hurdle as well.

Privacy represents another aspect of the risk AI poses to trust in the national healthcare system. Big data presents the opportunity for unparalleled quantities of data to be moved and manipulated in ways patients are not aware of, or properly able to recognise when occurring. While the response of the law to privacy problems is still applicable due to its flexible principle-based ruleset, the scale of potential breaches of privacy will likely mean the practicality of applying the conventional methods will be untenable.

The ability for localised person-to-person principles of medicine to be applied to a largescale tech framework is highly dubious. And while the tests can still be applied while human agents are involved in the process, as they become farther removed the ability to action effective responses for individuals is lost. By potentially requiring that patients concede a standard of protection they are accustomed or entitled to, patient trust in the system may falter, which in turn may diminish the benefits gained by the concession. It is likely NZ

will require a shift towards largescale forefront regulation and industrial controls similar to the EU's recent privacy developments. The proposed reforms in this area, guided by the EU's GDPR, will be discussed in Chapter Nine.

# Chapter 7: Informed Consent

## 7.1 Introduction

As a staple of the healthcare system, consent is the mechanism that enables patients and doctors to cross boundaries otherwise forbidden by law. When specific requirements are met, a doctor can act in ways that would normally be invasions of the patient's privacy, autonomy, and physical well-being. Informed consent is the standard which healthcare requires from a patient before a medical procedure or process can be undertaken. What this means varies greatly by jurisdiction and within context but broadly it requires the patient to understand, or at least be appropriately made aware of, the consequences, need, processes and relevant factual considerations involved with their treatment.

This chapter will begin with an explanation of the codification of informed consent within the Code. This involves the discussion of two inter-related rights within the Code; the right to be informed, and the right to give informed consent. After this, the section will include a discussion of the test applied when determining whether a patient did or did not provide informed consent.

Broadly speaking, the application of informed consent contains two main inquiries: whether the patient was given the appropriate information to base their consent on, and whether they are competent to understand this information when making their decision. The former, commonly known as disclosure, is a question of expectation. After determining if they are capable of giving the information, the amount and scope of information they should give in respect of an AI's involvement is one of purely doctrinal application. The latter limb, competency, focuses on the ability of the patient to understand the information they are told and then utilise it to make a decision. This chapter will briefly discuss the impact that AI has on the ability for a patient to understand, and if perhaps this standard is flawed going forward. These two limbs will be discussed largely as cautionary matters as opposed to fundamental issues which may impact the application of consent. It will be shown that informed consent as

it currently operates can still be applied, and largely without issue, but some cautionary points need to be highlighted.

Following this, there will be a discussion of the issue of dynamic consent. This idea was mentioned briefly in the previous chapter when discussing data use after-the-fact in respect of privacy. How patient data is used after its initial use, and how the requirement of informed consent facilitates this, will be analysed with the question: is the current requirements of disclosure and understanding sufficient for the potential breadth of uses and manipulation that data is likely to undergo outside the patient's involvement?

Finally, the chapter will close with an application of the three scenarios, outlined in [1.3.2.5], to a situation to determine whether the application of informed consent is reasonable in such situations, or if the test is inadequate for the realities of AI medicine.

How these issues may affect the viability of the test for informed consent is important to determining whether the current formulation of the right is realistic going forward. Consent also relates strongly to the issues raised in the next chapter, negligence and accident compensation, because failure to obtain sufficient consent is evidence of an inappropriate standard of care being provided.

## 7.2 Informed Consent: Overview of Current Law

Informed consent is a central component not only of healthcare, but of law and regulation at large. Within different fields and disciplines, the requirements of consent can vary, but the central premise remains the same: is this something that the person can consent to,[439] and do they appropriately understand the decision they are making? The second question is made up of several smaller issues, each of which will be discussed in turn in the following sections.

Firstly, how informed consent is codified within the Code, and what requirements are specified there will be outlined. Then the requirements outlined within the common law that

[439] The issue of what can be consented to is common within the criminal law, for example: an individual is considered incapable of consenting to their own murder (Crimes Act 1961, s63), or an individual cannot consent to undergo "female genital mutilation" (Crimes Act 1961, s204A). Within healthcare this is largely only relevant when discussing the competency of minors or those suffering from mental incapacities; these are both largely alterations of the baseline test of informed consent so will not be addressed within the context of this inquiry.

inform this will be explained. How this works in practice will be discussed, outlining the steps involved and where potential for dispute can occur within medicine. And finally, whether this test is relevant at all during the coming years, or if a new standard is inherently necessary, will be briefly discussed.

### 7.2.1 Consent in Application

Consent is both a legal and ethical issue, and accordingly the determination of whether consent has been obtained varies greatly depending on circumstance and individual. The requirements for capacity to give consent is not equal across all people; it will differ between adults versus children, and between the mentally sound versus those with mental illnesses or disabilities. For simplicity, this inquiry will not engage with this degree of variety, and will instead discuss consent in regard to an adult, with no immediately apparent features that would warrant concern over competence. The test for ascertaining consent is made up of two core features: what information must be provided, and is the recipient of this information competent to evaluate that information to form a decision?

Set out in the Australian case of *Rogers v Whitaker*[440], the standard of information needed for informed consent is information that is sufficient to "make a meaningful decision." This requires that information meet the "materiality" test: the information provided must be that information which would reasonably be expected to in some way alter the patient's decision-making process.[441] This requirement is reflected in the wording of both Rights 6 and 7 of the Code, which refer to what a "reasonable" patient would expect to be informed of. In *Bolam v Friern Hospital Management Committee*[442], the court held that a doctor is required to act in accordance with "the standard of care that would be expected by a responsible body of his or her peers."[443] The test under *Rogers* was qualified by the court, deciding that this standard of "appropriate care" should not be driven by professional standards, but instead by the specific needs of that patient, as would be expected in the opinion of their professional peers.[444] This

[440] *Rogers v Whitaker* (1992) 175 CLR 479
[441] At [30]
[442] [1957] 1 WLR 582
[443] At [56]; This case and its significance will be discussed in more depth in the following chapter, Chapter Eight, when discussing the "Standard of care" under breach of duty.
[444] At [65]

was further clarified in *Bolitho v City and Hackney Health Authority*[445] where this professional expectation was required to hold "a logical basis", be seen as a "responsible action", and, given the circumstances, be seen as a "reasonable" choice. This removed the option for a doctor to claim that their actions were common or ordinary, and therefore justified. There must now be some reasonable evidence as to *why* this is worthy of continuing to apply. The most recent decision to alter or clarify this test is *Montgomery v Lanarkshire Health Board*[446] where the reasonableness and appropriateness of an action is a case-by-case inquiry, considering not only the specific circumstances involved, but also the views, needs and understandings of the patient too. This shifts patients from passive recipients of care to active and respected stakeholders in the healthcare scenario.

In respect of competence, without contrary evidence, the law presumes that an adult is competent to make decisions about their health and wellbeing in healthcare. This presumption might be rebutted by evidence of temporary incapacity or reduced capacity. Evidence can be provided by either the medical Professional ("this individual is not competent to make X decision") or by the patient (or their family) ("my consent to perform this treatment was not valid, I was incompetent at the time."). Examples of temporary incapacity to make competent decisions are the influence of debilitating substances such as some medications,[447] or intoxication from drugs or alcohol. Examples of reduced capacity could be being underage,[448] suffering from a mental illness,[449] or suffering from an intellectual disability.[450] A patient who has capacity to give informed consent is expected to:

1. Comprehend and retain necessary information about the procedure or treatment;
2. Be able to believe the information; and
3. Be able to weigh the information, balancing the risks and needs involved.[451]

---

[445] [1988] AC 232 (HL)
[446] [2015] USKC 11, [2015] AC 1430
[447] *Re T (Adult: Refusal of Treatment)* [1992] 3 WLR 782
[448] NZ grants the statutory capacity to consent to medical procedures to those aged 16 years or up, under the Care of Children Act 2004, s36. NZ has supported the Gillick competency test (Gillick v West Nortfolk and Wisbech Area Health Authority (1986) AC 112) in *Hawthorne v Cox* (2008) 1 NZLR 409 to allow those under 16 to provide consent when able to display the necessary capacity.
[449] *Re C (Adult: Refusal of Medical Treatment)* [1994] 1 All ER 819
[450] *Secretary, Department of Health and Community Services v JWB and SMB [Marions Case]* (1992) 175 CLR 218 (HCA) 250
[451] *Re C (Adult: Refusal of Treatment)* [1994] 1 WLR 290, at [50]

In doing so, the patient is able to arrive at a choice once considering and executing those three factors.

### 7.2.2 Relevant Code Provisions[452]

Following Herbert Green's "unfortunate experiment"[453] with cervical cancer testing, one of the principal focuses of the Code was to centre the role of the patient in their healthcare experiences. This meant that patients are to be afforded a much more involved and important position within decision-making, ultimately having the final say on decisions. This however excludes scenarios in which consent cannot be obtained or cannot be considered to be sufficiently "capable". Informed consent is afforded by two rights, which set out the requirements of first being informed, and second being able to utilise that information in making a decision.

Right 6 of the Code is the right to be "fully informed". This outlines that a patient has the right to information that a "reasonable consumer" would expect to receive.[454] This is focused largely on communication of the relevant information throughout the healthcare process.[455] Right 6(2) contains the pertinent piece for consent:[456]

> (2) Before making a choice or giving consent, every [patient] has the right to the information that a reasonable [patient], in that [patient's] circumstances, needs to make an informed choice or give informed consent.

This provision largely serves as an introduction to the succeeding right where the requirements involved in obtaining "informed consent" are specified. This right also importantly establishes the standard associated with this inquiry, that of a "reasonable" patient. Importantly, this is different from the rule in *Bolam*. In that case, the rule is what information the reasonable doctor would provide in respect of that specific patient's needs in that circumstance, whereas the Code is concerned with what the reasonable patient in that circumstance would need (or expect) to know. This results in a difference in what evidence is

---

[452] Refer to Appendix I for a complete copy of the Code.
[453] Discussed in Chapter 3 at [3.2.3.2] and Chapter 4 at [4.2]
[454] Right 6(1), 6(2)
[455] Proceeded by Right 5 ("The right to effective communication") which is a largely procedural right about the way in which patients are to be informed (such as the form of language taken, or need for an interpreter).
[456] 6(2)

required; what would the reasonable doctor have told the patient in the same circumstance, or what would the reasonable patient expect to know, which may have changed their decision? This necessary material is referred to as the information that is "material" to the decision-making process. It will often be the case that the information required is the same under both Bolam and the Code, and in practice informed consent is a matter of shared, communal decision-making between a patient and their trusted health professional. For this chapter, the focus will be on what patient can reasonably expect to be told, following the formulation under the Code.

Right 7 establishes the "right to make an informed choice and give informed consent." Whilst the previous right establishes that a patient has the right to have certain information communicated to them, this right then establishes their right to use that information and either accept or decline to proceed with treatment as a result.

The longest right within the Code, informed consent contains ten sub-sections, detailing specific requirements or situational considerations. Some of these sub-sections will be omitted from this inquiry because they contextually do not apply to the scenarios this inquiry is concerned with.[457] Those which are being focused on relate to the action of giving or withdrawing consent, and the requisite competency (and associated exceptions) to this process.

Right 7(1) first establishes informed consent as:

> Services may be provided to a [patient] only if that [patient] makes an informed choice and gives informed consent, except where any enactment, or the common law, or any other provision of this Code provides otherwise.[458]

Violation of this provision results in a medical practitioner acting in breach of the Code and potentially committing a number of crimes and/or tortious breaches.[459] A medical procedure without consent can amount to both the tort of negligence[460] and criminal battery or gross

[457] For example, Right 7(5) is the right for a Patient to use an advance directive, which is largely a matter of competency and execution. There are some potential avenues of discussion in respect of changing technologies, although this is a broader inquiry that warrants further research. Those omitted are: 7(5) and (6) which dictate advanced directives and the requirements of form, and 7(9) and (10) which are concerned with the removal, storage and use of bodily parts or substances (such as organs, plasma or blood). See Appendix I for the full Rights contained within Right 7.
[458] Right 7(1)
[459] Such breaches, like the tort of negligence, are the focus of the following Chapter, Chapter Eight.
[460] *Abel v Brownlee* [2002] DCR 407

negligence.[461] Thus the burden of ensuring that informed consent has been obtained can be considered quite high, due to the harsh penalties available for misconduct.

Sub-sections 7(2)-(4) all outline the requirements of, and exceptions to, competency. Informed consent obtained from a non-competent patient will often be considered invalid due to their inability to properly consider and evaluate the decision being made. The presumption is that a patient is competent to give informed consent, unless there are "reasonable grounds" for believing otherwise.[462] A patient with diminished competence can only give informed consent to the degree appropriate for their level of competence; the more serious the decision to be made, the higher degree of competency required.[463] Right 7(4) expands the rules on how and when a doctor can provide treatment if a patient is incapable of providing informed consent and no suitable guardian or proxy is available.[464]

As well as outlining the right to give consent, the associated counter-action is available; "every [patient] has the right to refuse services and to withdraw consent to services."[465] This means that patients can refuse treatment they do not wish to undergo, even if it will result in their subsequent death.[466] Additionally, patients can remove consent during the process at any point (necessity can allow a proxy to withdraw consent, such as if the patient is undergoing surgery and is unconscious). Both of these are important for discussions of AI when considering the overarching principle of trust; the choice not to engage, or cease engagement when discovering its inclusion, is a real possibility.

The final provision of relevance is right 7(8): "every [patient] has the right to express a preference as to who will provide services and have that preference met where practicable."[467] The "practical" requirement limits this right to one of reasonable accommodation, however, does raise interesting questions about what is included in "who". When an AI reaches the

[461] Crimes Act 1962 section 151, 157
[462] Right 7(2)
[463] Right 7(3)
[464] Right 7(4)
[465] 7(7); This is a right enshrined within the NZBORA also, as the "right to refuse to undergo medical treatment" in section 11
[466] This is sometimes represented as the "right to make a mistake" within the common law, where the right to make decisions about one's own autonomy exists regardless of the consequences. For example, see *Airedale NHS Trust v Bland* [1993] 1 All ER 821, 860 per Lord Keith. *Re T (Adult: Refusal of Treatment)* [1992] 3 WLR 782, 786 goes further still with Lord Donaldson suggesting this right (for a competent adult) to make decisions about their own health is absolute.
[467] 7(8)

capacity to act independently, whether patients will be able to express a preference for the AI to do so over human agents is uncertain.

These sub-sections together form the wide net of situations, requirements and applications of informed consent within the Code. How these fit together in application will be discussed next, so that the points at which issues may arise from the involvement of AI can be illustrated. The idea of requiring doctors to facilitate the approval of patients is in itself not controversial; the issues arise when considering how these rules alter the implementation of new technologies and their nature.

## 7.3 Informed Consent: Overview of AI Issues

The previous section provided a brief overview of consent under the common law and the Code. The remainder of this chapter will focus on the requirements for informed consent associated with information and those of patient capacity. The three issues to discuss are: (1) what information is necessary to disclose, or what might the patient expect to be disclosed for there to be informed consent, (2) the capacity to, and ability to, understand an AI, and (3) dynamic consent.[468]

### 7.3.1 Issue One: Disclosures and necessary information

As stated, a patient has the right to information they would "reasonably" expect to know before deciding. This information is what they consider material to the decision, and they therefore require to make it. When discussing information that is material to the patient, it is important to recognise that this is a largely personal inquiry. Different patients may place increased value on different factors; the reasonable patient hypothesized is one in the same circumstances as the patient, meaning they can vary greatly based on situation, personal experience and process. This section will briefly discuss different components of an AI's involvement that may be considered relevant in a person's decision-making process, and

[468] Another way to think of dynamic consent is "fluid" consent; where the consent given or required changes, or needs to be updated, as the purpose and use changes over time.

whether they warrant specific disclosure to achieve "informed" consent. The different pieces of information which will be discussed are:

1. The involvement of an AI;
2. Risks and outcomes associated with the AI; and
3. The expertise of the person operating or utilising the AI.

These issues are focused on the disclosures involved in a medical scenario, so as to highlight both the information required to be disclosed, and the patient's capacity to understand and utilise this information.

To understand the potential responses of the hypothetical "reasonable patient" to the use of AI, this section will utilise the results from my survey discussed in Chapter Four.[469] While not definitive, the results of the conducted survey will be used to illustrate what sort of consensus can be inferred amongst participants; what they themselves considered "material", and whether this aligns with argumentation from elsewhere. A study performed by Richards and Hutchison into how the "innovative" nature of a surgical process impacts informed consent will also be used. Within this study, healthcare professionals were asked what they considered to be the impacts, risks and difficulties of achieving informed consent for innovative surgical processes. Many of these concerns are applicable to AI, at both its early implementation stage and beyond.

#### 7.3.1.1 The involvement of an AI

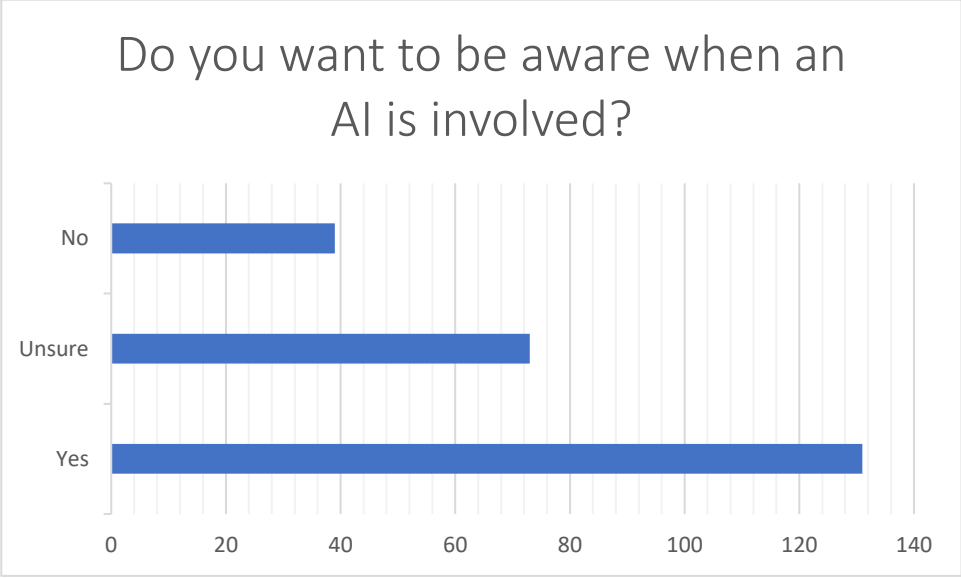
The first, and perhaps most critical, piece of information to discuss is the involvement of an AI in the treatment process. While typically patients are given explanations as to the mechanisms that are involved in their treatment, they are not necessarily given an explanation as to the functionality of that device. For example, a patient is told they require an MRI scan. A patient may know about or have a vague understanding of what an MRI is, but the doctor does not explain the actual technical processes involved. If this same MRI was integrated with an AI system that analysed the data collected, and generated a diagnosis, treatment plan, or any other output, would that feature of the system need to be disclosed? If so, would the form of AI also warrant disclosure? While this inquiry is treating AI as a monolith, it should

[469] A full explanation of the survey and prominent results from it occurs in Chapter 4 at [4.5]. The survey questions are available in Appendix II.C.

be remembered that there is a stark difference in functionality and mechanism between systems that utilise a neural network from one that uses decision-trees.

If the involvement of the AI is relevant to the decision making, the reason for this might needs to be identified. On one hand, if the determination that the involvement of an AI should be disclosed was due to its innovative nature, a requirement of disclosure is only relevant for the near future where it is still "innovative." This would mean as time progressed, and AI became more common place, their involvement would become progressively less important to disclose. Alternatively, it may be that the determining element is not its innovation, but its inherent nature. In this case, it would mean that regardless of time, integration or commonality, the AI's involvement would be expected to still be disclosed. Identifying which of these (or any other reasons) for why people hold their opinions of AI would require more robust research on public perception, and would likely be reliant on better understanding of AI as a concept amongst participants.[470]

In this thesis' survey, participants were asked if they expected to know that an AI was involved with, or responsible for, their treatment in any capacity (Q11). Over half of participants said they did want to know outright.




Those who were unsure had widely varying answers so little pattern could be identified,

[470] It would be difficult to verify that the latter reason, of the AI's inherent nature, was the cause of their position when the participants understanding of what AI is varies so greatly.

however a few answers that appeared commonly were: transparency for early adopters, and for situations considered especially serious (such as cancer) so they could trust the results. Some made distinctions based on what the AI was involved in, during their treatment. If an AI was being used to diagnose an issue, they expected to know; if an AI was being used to develop a treatment plan or something more logistics focused, they were not concerned. The broad conclusion appears to be those who view medicine as inherently intimate (or those who consider some procedures more personal than others) expected to know that an AI was involved. Those who were not concerned about such things, and saw no special status to these kinds of interactions, did not expect it. Of those who answered "yes", many cited their trust in the system relying on this, and that disclosure of process being undertaken was inherent to their personal understanding of healthcare. It was suggested by a small number of participants that this would likely change over time, highlighting that if the AI systems became known to be more reliable or commonplace they would be less concerned. If AI becomes commonplace, it is possible that the automatic expectation is that an AI is involved in one's care. At least in the event of steps like diagnosis, utilising an AI (at least as a part of the broader process) could be considered not only the norm, but also expected.[471] A small number of "yes" participants answered that they expected it to always be disclosed, and that it is never okay to utilise AI without a patient's knowledge (there was no specified reason why AI was singled out in this way). One participant specified they would avoid any medical practice that utilised AI technology in any capacity. Those who answered "no" largely did not give specific reasons, the most common comment simply being that they were either not concerned, did not care, or did not understand why it mattered. These participants could be seen as representative of those who more generally trust healthcare practitioners, or do not concern themselves with specific details of their treatment.[472] The broad conclusion appears to be those who view medicine as inherently intimate (or those who consider some procedures more personal than others) expected to know that an AI was involved. Those who were not concerned about such things, and saw no special status to these kinds of interactions, did not expect it.

[471] This will be discussed in more depth in the following chapter, Negligence and Compensation, in [8.4.2.1]

[472] These participants were also common amongst those who said they considered "statistically better outcomes" the most important common of healthcare, discussed in [4.5.4]

These results appear to suggest that an AI itself is worthy of disclosure, irrespective of how commonplace its involvement is. This would mean that at any point, and with any degree of involvement, a reasonable patient would consider not knowing an AI was involved to be a violation of their informed consent. While it is possible this is a valid position to hold, it is also likely evidence of both a limitation of the survey conducted and also the understanding of the participants of AI. The survey questioned the involvement of an AI generally, with no reference to or guidance on time, commonality, or innovation as mitigating or exacerbating factors. Because of this, whether this difference was of sufficient importance to patients was not ascertained. Additionally, because participants were self-admittedly limited in their understanding of, and perhaps reserved about, AI systems this position may be one that changes over time. As AI systems become more commonplace and integrated this expectation may recede, illustrating that this position was largely concerned with the innovative nature itself. While the survey is therefore useful, more detailed empirical work will be needed to explore answers in more depth.

#### 7.3.1.2 Risks and outcomes

Medical procedures come with a wide array of different risks, prospective outcomes and expectations. It is perhaps difficult to specify all of these risks, especially those which are particularly rare or difficult to isolate. The question then is, what is risk, and which are important to disclose? In *Sidaway v Governors of Bethlam Royal Hospital*[473] Lord Scarman adopted the test in *Canterbury v Spence*[474] that stated:

> "a risk is… material when a reasonable person, in what the physician knows or should know to be the patient's position, would be likely to attach significance to the risk or cluster of risks in deciding whether or not to forego the proposed therapy."[475]

Lord Diplock concurred with this position, stating that one would need to be "fully informed of any risks"[476] to properly exercise their self-determination. The caveat to this position was given by Scarman, by soft endorsement of the *Bolam* position, that the risk need *not* be disclosed where the doctor "reasonably believes" that disclosure would harm the patient's

[473] [1985] 1 AC 871, 882
[474] 464 F 2d 772, 780 (1972)
[475] *Sidaway*, above n 463, at 787
[476] At 895

health.[477] Lord Templeton argued that too much information can potentially be just as harmful as too little information. The obligation on doctors was to inform of "general" risks, and specific risks required to be inquired by directly by the patient.[478]

In Question 9 of the survey, participants were asked what information they would consider important to know about the AI, or its functionality, to allow for it to be involved in treatment. Largely participants were interested purely in a systems' effectiveness; was it better or worse than the regular alternative, and if worse, what benefits did its use incur? These answers were mirrored by another portion of participants who questioned the dangers, or oversight protocol, of the AI system. Fifteen participants highlighted their need to understand the misdiagnosis rate (as opposed to the rate of better results), and nine others wanted to know what degree of oversight a human doctor had over the system and its outcomes. The largest number of concerned respondents, twenty participants (all of whom gave some indication of computer science background[479]), said data storage protocols and any data sharing in place was their main concern. One participant referred to what "biases may be present in the system, particularly if the patient is like me and a POC" (person of colour). Four participants dismissed the question by simply stating "it doesn't matter, I wouldn't allow it". While coming from different angles, these answers could be recognition of the importance of understanding the risks and outcomes of a system that they may be exposed to. Generally speaking, risks and potential outcomes from a treatment procedure are material information. If a surgery has the potential to render a patient blind, the reasonable patient would expect to be informed of this when deciding whether to undertake the surgery. Richards and Hutchison highlighted that the disclosure of *unknown* risks was a particular area of concern for health practitioners surveyed. Especially for innovative, early adoption (but also true for AI generally to a degree) is the risk that unknown, and entirely unforeseeable, outcomes arise. This concern was in part connected to the aforementioned loss of authoritative status as "knowers" but also based on the potential harm it can create.

In Richards and Hutchison's survey, practitioners highlighted practical concerns in respect of innovative processes, namely that there is an impossibility of disclosure for things not yet

---

[477] At 889-890
[478] 902; Frustratingly, Templeton did not provide a definition on what "general" or "specific" risks are.
[479] Discussed in Chapter 4, at [4.5.2]

recorded or discovered.[480] However multiple participants mentioned that discussion of unknown risks may not only be difficult but even immoral, or detrimental to the overall care of a patient. "Participant 8" stated "you don't want to… you know… scare them."[481] Disclosing that something is unforeseeable or decidedly more complex than previously experienced may result in increased fear, diminished trust or confidence, and a reticence to engage from patients. This hesitancy may result in decisions to decline treatment or engage with lower-effectiveness options, detrimental to their health. Patients have the right to do this, but doing so purely because of an emotional response from "over disclosing" is a concern. This position is supported by *Bolam* where the "specific needs" of a patient are the determining factor for disclosure. The omission of information to prevent greater harm could be reasonably argued as a necessary concession, especially in instances where the use of the AI results in a greater standard of care generally.

Overall, the risks and outcomes associated with the use of an AI, even when they are unknowable, should be disclosed as part of the comprehensive pre-treatment discussions. However, *Bolam* and earlier jurisprudence both support the position that some information can be withheld, or at least dismissed, in the event that it would lead to greater harm. Importantly, there is no expectation in the law that all possible risks are identified and communicated to a patient. The reality that AI treatment involves a plethora of potentially unknowable risks is arguably no different than the situation with conventional healthcare. Scientific understanding of the body and the complex interactions of healthcare interventions are an ever-evolving field of study and investigation. Disclosure of the fact that some things may be unknown is a case-by-case determination, but the risk of not identifying potential risks is largely inconsequential.[482]

#### 7.3.1.3 Providers' expertise

It is perhaps taken for granted by patients that a doctor who recommends a particular procedure or treatment has experience with that line of action. When a patient who requires

[480] Bernadette Richards and Katrina Hutchison, "Consent to Innovative Treatment: No Need for a New Legal Test" (2016) 23:938 Journal of Law and Medicine, at 942
[481] At 942
[482] At 947

heart surgery is met by a heart surgeon prior, they would reasonably presume that this heart surgeon has performed the procedure before. However, with the involvement of any new medical development, there is always a first time. The inclusion of new AI systems raises the question of expertise or understanding on the part of those utilising the systems. New technologies all come with an associated period of training and skill-development. The question is whether a doctor is required to disclose this relative inexperience, or the process of upskilling, during their recommendation or proposal of a systems' use? This is largely a concern for early adoption, but the principle can be applied broadly to any new development within healthcare. A significant conflict identified by Richards and Hutchison was the effort to maintain the roles of both patient and doctor within the informed consent process. They found that there were significant fears from professionals about maintaining their authoritative positions as experts, while minimising patient burdens of knowledge.[483]

As previously mentioned, Question 9 of the survey asked "what information would you consider most important to know when determining whether to allow your treatment to be performed by an AI system?" 19% of participants said they expected to know who was operating the AI (or was responsible for its oversight) and what experience they had doing so. This appears to be centred on the fact the technology was new, as opposed to a component of AI treatment generally. It is fair to suggest that once a mode of treatment is commonplace, an inherent trust is formed by the patient that those operating it are adequately trained to do so. When undergoing a CT scan, for example, a patient is unlikely to quiz the radiologist's credentials. In respect of new forms of treatment, participants appeared reticent to offer this same trust without it being verified.

In a decision of the Health Disability Commissioner (HDC), it was concluded that the "relative inexperience" of the surgeon involved was something a "reasonable person" in the plaintiff's position would want to know.[484] In this case, the new element was the use of a robotic assistant for prostate surgery, which was described as "relatively new" within NZ at the time of the complaint. The doctor was found liable for not having disclosed that the procedure was both new, and that they themselves were on the "learning curve".

This position is likely directly associated with the obligation to disclose risks; greater experience should correlate to reduction of some risks, and an increased ability to respond to

[483] Richards and Hutchison, above n470, at 942
[484] Health and Disability Commissioner, Decision 08HDC20258 (11 November 2009)

unforeseen variables. For new methods and processes, the patient is exposing themselves to a situation in which the practitioner may be unaware of how to proceed, or to efficiently conduct the treatment, thus increasing risks and burdens on the patient.

#### 7.3.1.4 Conclusion

Within NZ there is no bright line test in place to determine what information is, or is not, material to the consent process. While the "reasonable patient" provides at least a theoretical starting ground, this is qualified by the recognition of personal and circumstantial factors which may alter, or at least warrant consideration of, the information provided.

In respect of early adoption, it is most likely the case that a patient would find the fact that a procedure or mechanism is new a materially important piece of information. The HDC has concluded previously[485] that this is information that a reasonable person would want to know, and this is consistent with the logic applied in healthcare generally for the kinds of things patients considered important to them.

While reservations exist amongst professionals as to the effect that disclosures will have on healthcare, it is suggested that practitioners err on the side of caution by disclosing information even if they themselves may think it will in some way detriment their position. The concerns of patient fear and stress are valid, but the potential for patients to feel violated or deceived post factum is perhaps the more egregious concern. Additionally, as pointed out by Richards and Hutchison, much of the concerns about the perceived "authoritative position" are engaging a standard that far exceeds the legal one and appears to be largely a concern of professional status.[486] What is important to emphasise though is that none of these provide any particular challenge for the application of informed consent doctrine and is also not a practical barrier to medical innovation. The disclosures required are commonplace within healthcare and are easily adaptable under the current NZ regime.

---

[485] Health and Disability Commissioner, Decision 08HDC20258 (11 November 2009)
[486] Richards and Hutchison, above n 470, at 946

### 7.3.2 Issue two: Artificial Intelligence and understanding

The second issue with informed consent is that of understanding; does the patient properly understand the technology in question, and can they use the information disclosed to them effectively? Consent requires, as discussed in [7.2], that the patient: be able to believe the information, and be able to weigh the information to make their decision. While the presumption is that a patient is competent in the absence of evidence to the contrary, this section will discuss whether this assumption is fair in respect of AI. As discussed already in this thesis, participant understanding of AI is often both flawed and negative in its connotations. This section will briefly outline the concern around patient understanding, and then discuss whether or not this concern is warranted in light of the current application of the law.

#### 7.3.2.1 The concern

Richards and Hutchison noted that a common comment by their participants about patients was that "they just haven't had the formal medical education, that they have no understanding and no chance of understanding."[487] It was a common concern expressed by Richards and Hutchison's participants that innovative processes would impact the patient's ability to make both a "rational" choice[488] or understand the process being described to them.[489] This is intimately linked to the earlier issue of both practitioner expertise, and unknowable risks, in [7.3]. If a doctor themselves does not understand the process completely, how can they communicate the process to a patient sufficiently? This would be especially applicable to treatment involving an AI, where the actual functionality of a system is unlikely to be understood by a utilising doctor.

An additional concern noted was that the prospect of new processes or treatments could limit or impair patients' understanding, or even their desire to understand, their options appropriately. If a patient is particularly stressed about a condition, informing them that this treatment involves an AI (either as a new process, or commonplace) could exacerbate their stress and result in them not properly weighing the information given. A related challenge identified by their participants was patient's willingness or ability to take information into

[487] At 944
[488] At 945
[489] At 944

account because of their own self-interests. The authors recount two forms of "indifference" to information, noting:[490]

> "… the apparent abrogating of personal authority and deferring to the professional skills of the surgeon, describing patients who "don't want to know what he [the surgeon] is doing, as long as he fixes the problem" (Participant 2)." and
>
> "A somewhat different issue was associated with patients who were desperate for treatment, where innovation offered new hope. In these situations, patients' indifference to information was not due to passivity, but more appropriately labelled as an active rejection of information, with patients preferring to focus on potential positive outcomes. One participant described receiving "more than 1,000 phone calls within the next week […] saying they'd pay anything" (Participant 12) to receive a new treatment described in a news article, and worried that these patients would not even read the patient information."

Both concerns contain an implicit suggestion that a patient's competency is somehow related to a rational, or "best interests" decision. The perception is that the involvement of complicating elements (in their case innovation, but equally as applicable to AI) mean that patients will not appropriately follow the reasoned path expected of them, and that this somehow implies their decision is incompetent.

#### 7.3.2.2 The degree of understanding needed

Once a patient has been given the necessary information, they are required to be able to understand and then apply that information to make an informed choice about that treatment. They must take it in, retain it, believe it, and weigh the risks and needs that this information provided.[491] What this means can be divided into two steps: how are patients expected to go about this process of consideration, and what level of "understanding" is necessary for them to make the decision? The short answer is that this is a matter resolved by the requirement of "materiality." In practice, discussions of informed consent are approached from the angle of whether the information required *was* provided, not what was done with this information afterwards.[492] Patients are highly likely, in almost any healthcare scenario, to be distracted by

---

[490] At 941-942
[491] *Re C (Refusal of Medical Treatment)* [1994] 1 All ER 819
[492] *Rogers*, above n 430

their own anxieties, ailments, and misunderstandings of healthcare; the law already allows for these disparities in consideration as an acceptable, or at least unavoidable, component of the process.

While the process or "rationality" of a patient's choice, is a valid concern on the part of health professionals, it appears to have little effect on the actual application of consent. If a patient is considered capable of giving consent, it is not relevant whether their decision is seen as "rational" or appropriately reasoned in the end. The fact that a patient's own understanding or belief system may be "unusual"[493] is not sufficient to decry their chosen path. As discussed in *Re B (Adult: Refusal of Medical Treatment)*[494] the law should aim to avoid "benevolent paternalism"[495] and to respect the autonomy and right of the patient to not only choose, but potentially choose poorly. Reasonable steps must be taken to inform the patient of the necessary information and assist them in understanding it. Whether they choose to reject or disregard this does not invalidate that reasonable steps were taken in the circumstances. The issue of rationality is ever-present in healthcare and is not necessarily exacerbated in a unique way by the inclusion of an AI system.[496]

In regard to how much the patient needs to understand, the law makes no expectation that patients are able to understand completely, or to a degree equal to that of the health practitioner. The doctor-patient relationship is an inherently uneven one, irrespective of the mode of treatment being applied, or the standard of information being relayed to the patient. Within healthcare a patient is unlikely to understand the technical details of any treatment they receive (except those with personal expertise in relevant areas). Whether a patient is being treated by an AI system, or is being prescribed a routine medication, there is no expectation that the patient can explain the actual scientific processes involved.[497] Instead, the expectation is only that patients who are considered by the law capable of making determinations about themselves are given the information material to do so. Doctors merely have to ensure the information they provide does not mislead the patient. Whether a patient

---

[493] *Re C*, above n481
[494] *(Adult: Refusal of Medical Treatment)* [2002] 2 All ER 449
[495] At 56
[496] For example, the decision to deny a blood transfusion on the grounds of religious belief as Jehovah's Witnesses may be seen as irrational to those who do not share those beliefs, but this does not invalidate their decision. See *Auckland District Health Board v W & W* [2012] NZHC 1563 for one such discussion.
[497] *Rogers*, above n 430

being told their treatment utilises an AI will be lead to a conclusion that harms them is a risk that needs to be weighed under the *Bolam* test, but does not affect whether the patient is competent to consider that information. The individual characteristics of a patient, including their own intelligence, expertise, fears or anxieties, are components of the discussion of what to tell them in respect of their needs, not their abilities.[498]

A way of rephrasing competence considering this discussion is that the requirement is not that patients *actually* understand what they are told. Instead it is that they are fairly given the *opportunity* to reach their own conclusion for their well-being, without being misled or deceived. Whether the mental process they undertake mentally is rational, or even properly considers the information, is irrelevant to whether they were afforded the opportunity. So long as the patient is not suffering from a recognised vulnerability or limitation, their ability to actually reason the information provided is secondary to their right to do so.

### 7.3.3 Issue Three: Dynamic consent

The third issue pertinent to informed consent when an AI system is involved is what will be referred to as "dynamic consent." Typically, informed consent is "static" – it is a specific permission granted for a specific action. This permission can be expansive in scope, and allow for specific deviations, but it is for that singular broad purpose. An example would be a patient who grants consent to undergo heart surgery. Their consent may include allowing surgeons to respond to any unexpected occurrences (like internal bleeding), or even to remove something else they discover during the procedure (a cyst near the heart for example). Once the surgery is over, this consent has ended and for future actions further consent is required. With patient data, AI presents the opportunity for continuous, sometimes unexpected, and highly expansive uses to continue long-after (or well-beyond) the planned procedure. This is the issue of dynamic consent: how does the process and requirements of informed consent adapt to a new form of medicine that involves reusing, manipulating, and altering data in entirely new ways.

As discussed in relation to privacy in [6.3.4], AI's collection, aggregation and utility of data leads to a complicating problem for consent. The consent given by a patient during

[498] *Rosenberg v Percival* (2001) 205 CLR 434; [2001] HCA 18, at 55

treatment is generally for that specific event; a doctor seeks consent to give the patient an MRI scan, the patient provides consent for that MRI scan. The patient also generally gives consent for that information to be stored by the hospital, and then perhaps to be utilised for research or other purposes later. This conventionally occurs through individual acts of consent, within the same discussion. Within the given consent there is often included an emergency consent for situations that may arise during, particularly in the case of surgery. If a patient agrees to open heart surgery and then begins to suffer internal bleeding, their consent extends (either explicitly or otherwise) to remedying this as well.

AI creates a situation in which this otherwise static consent may be simply too restrictive for situations in which unforeseeable, or distinctly remote, occurrences occur involving patient data. This need for more dynamic consent comes into play for two main types of situations: (1) the AI acts in an unforeseen way and diagnoses or utilises something the patient did not wish for, that was not within the consented purpose[499], and (2) that the patient's data is used after the fact for a variety of purposes to which the patient did not consent or is not necessarily even aware.[500]

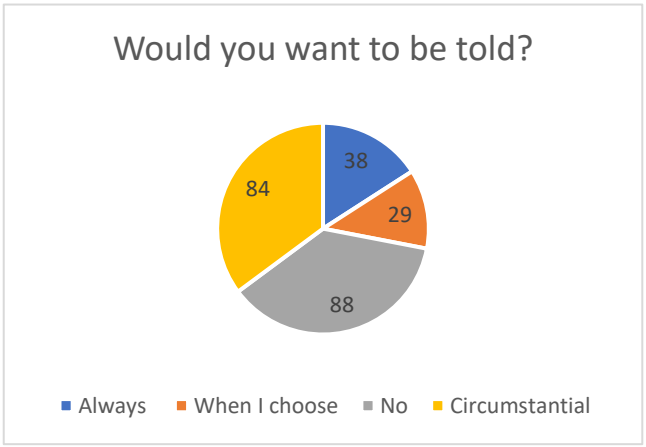
#### 7.3.3.1 Anticipatory medicine & unforeseen actions

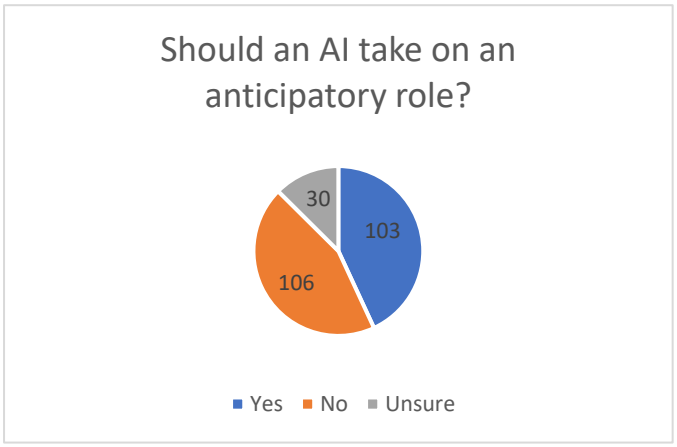
AI's ability to process immense amounts of data and form previously unknown connections has a number of benefits beyond simply faster decision-making; most notably of which is its ability to perform a large number of tasks simultaneously, and use its generated conclusions to facilitate more decisions concurrently. Where a single radiologist might require outside opinions, consultants, and further investigation to draw different conclusions, AI have been shown to be able to "bridge gaps" and identify markers of illnesses that otherwise elude detection. This however presents a problem; what if an AI comes to a determination that involved utilising data in a way not consented to or beyond the purpose of the initial request? For example, a patient goes to a doctor who is making use of an AI system to identify complications with pregnancies. After inputting any number of scans, and other medical data, the AI performs an analysis of the pregnancy, and in the process also identifies

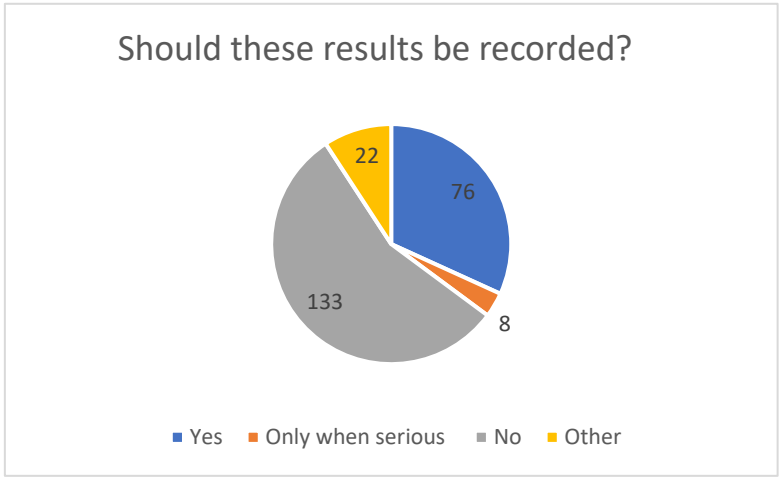
[499] An example of something unforeseen that is related to the main purpose would be unknown internal bleeding being found, during a different procedure. However, the discovery that a patient has Huntington's disease during an MRI for a bone injury is entirely unrelated, but in theory possible.
[500] This situation is intimately related to the issues of privacy discussed in the previous chapter, particularly in respect of "use" discussed in [6.3.4.]

that a patient is at an increased risk of cervical cancer. While it is not unheard of for doctors to identify a secondary problem when investigating another, this discovery by an AI came about because of capabilities that a human doctor lacks. The issues here are therefore that the information which is utilised, and *how* it is utilised, are not able to be conveyed during the pre-treatment consent process.

When asked if they were comfortable with an AI taking on an anticipatory medicine role, my survey participant's answers were at their most varied. Anticipatory healthcare was defined as when an AI system makes judgements or identifies connections beyond what a patient has necessarily sought treatment for without consciously consulting on it. Participants were first asked whether they wished for an AI to do this, and then secondly whether they would like the subsequently identified diagnosis to be recorded or discarded. Participants were not given binary options of yes and no, and instead were left to describe their opinions generally. Below are three graphs outlining the common responses given, focusing only on the dominant reasoning given by participants.



Interestingly, those who answered in the affirmative for AI taking on this role were not all interested in this information being recorded or communicated to them. The common ground across the three questions was largely for serious conditions; in the event of conditions like cancer, or early on-set dementia, participants would like to be informed. Those in the "when I choose" category for being told largely responded that they would want this information recorded, but only to be brought up or utilised when they themselves opted for a deeper look into their health. The logic here appears to be about efficiency of ongoing care, as opposed to actual benefit of the recorded information at that specific time.

The law has a clear position on emergency detours within medicine. In the event that a patient requires urgent care during the process of the consented care (i.e. internal bleeding during a surgery) then this is both permitted and is even a common component of health guidance.[501] There is an expectation to not proceed and to wait until the patient's consent can be obtained if reasonably safe to do so.[502] This same reasoning cannot really be applied to situations in which inadvertent, incidental, or perhaps careless detours occur however. It is most likely the case that things identified in these instances are not immediately life-threatening (otherwise they would generally be detectable normally) and while disclosure may benefit the original purpose of the patient's consent, it is difficult to suggest they would be necessary. One context in which this has the potential to arise most often, at least on a single patient-doctor scale, is in instances of hereditary illnesses that a patient is choosing not to investigate. Huntington's is a common example where a person may know there is a chance they have the illness, due to their own parents having had it, but they choose not to get tested.[503]

In the case of self-service AI systems, it would likely be the case that use of the system involved the agreement that these situations were a recognised, and acceptable, risk. It would be difficult for a patient to suggest that because they utilised a self-service mobile application and were told something they otherwise would have not wished to know in the process of a

[501] Right 7(4) allows for a doctor to operate on the "best interests of the patient" in the event of the patient being incompetent to make a decision (which being unconscious would amount to)
[502] Emergency situations that allow a doctor to proceed without obtaining consent need be "for the immediate goal of saving the patient's life or wellbeing". In the event that the harm is not sufficiently serious, it is likely the case a doctor would be required to wait.
[503] A brief discussion of the duty to inform patient's in respect of a diagnosis like Huntington's occurs in [8.4.1.2]

different diagnosis, they could blame anyone but themselves. This is of course reliant on the whether the risk of this occurring is disclosed at the outset.

#### 7.3.3.2 Data use after the fact

By design, AI systems require large, curated, datasets which continue to grow and be developed over time as they are implemented. The reality is then that a patient's data continues to be used, analysed, and manipulated long-past the patients' diagnosis and treatment. Generally, when a patient provides consent for their data to be used for research, it is for that specific research. A patient may participate in a study, or be made aware that their test results are relevant to a study and opt to allow the results to be shared. There is a strong incentive in the age of Big Data for corporations to then buy large amounts of data utilised by these studies for training or development of AI systems later. Presumably, a patient can opt to provide consent for their data to be used in this way also. Utilising patient data for research purposes usually requires one of three criteria to be met: consent of the person; that the person is dead; or that the person is unidentifiable.[504] Often data of this kind is anonymised already, as required under the HIPC, and this is an integral component of the patient's willingness to allow it.[505] As discussed in [6.3.1] the impracticalities of this are becoming more and more apparent. What is likely to happen is a scenario like the following:

> Patient X consents for their anonymous data to be used in a research study by ScienceInc. ScienceInc uses this data, and when their study is done, they sell their curated dataset, still anonymous, to BigTech. BigTech is developing new AI systems and has several purchased and combined datasets. BigTech re-identifies X within the data set and continues to utilise their data for training.

In this scenario BigTech has not only violated X's privacy but is also complicit in the breaching of their informed consent *and* for failing to obtain their consent for its continued use. Once the data has been de-anonymised, the reasonable conclusion is that the now identifiable person's consent is required to continue to utilise it. This is difficult, if not

---

[504] HIPC, rule 8; Kalra D, Gertz R, Singleton P, Inskip "Confidentiality of personal health information used for research" (2006) 333(7560) HM BMJ at 196-8

[505] Peter Singleton and Michael Wadsworth, "Consent for the use of personal medical data for research" (2006) 333(7561) BMJ 333, at 255-258 provides an overview of the limitation on use of data for research in the UK context. This is largely similar to the HIPC situation as well, with the Medical Research Council of the UK providing their specific rules. Accessible here <www.mrc.ac.uk/pdf-pimr.pdf> accessed on 01/02/2021

impossible, to do as a matter of scale in these situations,[506] and would result in almost entirely discarding the benefits of the data in the first place.[507] As Singleton and Wadsworth discussed, it is likely the case that sufficiently removed data (that is data which is "removed" by either time or distance from its original collection and purpose)[508] does not warrant consent after its initial use.[509] However they recognise that different rules might be required in regards to access, and who can purpose or utilise information as a method of control because the requirement of consent likely becomes less practical the further removed data is from its original source.[510]

The issue of consent after-the-fact, and dynamic consent generally, is not unique to AI. Discussions of the impracticalities of consent in this area have occurred for decades in respect of organ donation,[511] developing genetic illnesses, and other such evolving conditions. The easiest solution is to have the patient consent to their data being used for transferrable purposes indefinitely. This may be tenable in the case of research-only uses, but the transfer from research to the public sector will likely cause hesitancy in patients. Additionally, the requirement that their data remains anonymous is likely unrealistic and would act as a practical barrier to ensuring patient trust and approval. A study into public views on clinical data use was conducted in Europe and found that a willingness to give broad consent for data use in healthcare-specific data banking was very high at 93%.[512] This willingness was shown to be closely tied to the protection of their own interests[513] and when the process of doing so was in compliance with the GDPR (involving anonymisation of data).[514]

---

[506] The time and administrative investment necessary to contact and verify the consent of each participant likely outweighs any financial benefit or internal metrics associated with the use of the purchased dataset.
[507] Singleton, above n 495, at 256 where they analyse the cost-per-case of increased informed consent
[508] An example of this would be patient data, which has been sold to a third party and is now being used for socio-economic analysis.
[509] Singleton, above n 495, at 256
[510] At 257
[511] O. O'Neill "Some limits on informed consent" (2003) Symposium on consent and confidentiality (Newnham College, Cambridge, UK)
[512] Gesine Richter and others, "Patient views on research use of clinical data without consent: Legal, but also acceptable?" (2019) 27 European Journal of Human Genetics, p841-847, at 4
[513] Patient data use could not be done in a way that conflicted with their own interest, in this respect they expected to be able to broadly consent to specific uses and categories in advance, at 9
[514] At 10; This meant that if the procedure to obtain consent before storing health data including some evidence that the standards of the GDPR were considered and met (i.e. an explanation of how they would anonymise the data, and some form of audited approval that this was acceptable), participants almost universally agreed.

The additional complication for this issue is one that has been remarked upon numerous times throughout this thesis; the scale of impact. In instances where a large group of people become aware that their data has been utilised in a way they did not consent to (and most likely violates their privacy also), the conventional mechanisms for resolution under the Code and tort are likely unable to manage. Scale and its impact on the enforcement and protection of rights and obligations in healthcare will be discussed in Chapter Nine "From Intimate to Industrial" in [9.5].

### 7.3.2 Conclusion

Like in the previous chapter, which considered privacy, the issues identified with consent can largely be seen as issues of scale. AI technologies allow for data to be collected and utilised on a scale unprecedented in healthcare, and this creates both practical and reasonableness issues when trying to enforce when consent is needed, or how consent can be sufficiently obtained. Similarly, data in this context can be manipulated and transformed in ways that mean the original consent may be insufficient, but now impractical to update or facilitate.

It appears at this stage that the public perception of AI leads to the conclusion that the involvement of an AI, and any AI-specific risks or criterion, is material information to the consent process. What this conclusion is rooted in, whether it be innovation or the technology itself, is still difficult to determine. What is clear however is that AI attracts a unique perception amongst people, and this will no doubt colour and complicate its implementation and acceptance in the coming years. With these conclusions in mind, the next section will apply the three scenarios set out in [1.3.2.5] to illustrate the practical application of informed consent in these scenarios, and the possible shortcomings of this.

## 7.4 Application of Three Scenarios

This section will illustrate the application of informed consent, discussed in [7.2] to the three scenarios set out in Chapter One. As in the previous chapters, this application will utilise a singular base scenario to which the three scenarios are applied, allowing for consistent discussion and analysis.

Any associated issues of privacy that could arise within the scenario will not be addressed within this application to avoid repetition of discussed issues from Chapter Six. This application will aim to illustrate the transformative nature of consent and principally the practical effect of time on the potential issues discussed within this chapter.

The framing situation will be the following:

> The patient, T, develops some concerning symptoms over the weekend. On Monday, they [either visit a hospital, or utilise their mobile application] to seek diagnosis of the symptoms. T consents to their patient data being used to diagnose their current illness, which is later identified as a complication of their known diabetes. They also consent to their data being collected for research into AI diagnostics; their consent is reliant on them remaining anonymous. The AI utilised has access to T's wider medical records, and whilst diagnosing this, identifies a connection between their current symptoms and their recent blood tests, which suggests they may also have undiagnosed Huntington's Disease (HD). T knew there was a possibility they had the disease but had elected not to investigate previously. Because HD is exacerbating his current symptoms, it becomes known to him he has the condition once his current status is explained to him. He believes he did not give informed consent to the diagnosis of Huntington's.
>
> Later, his patient data has been anonymised by the health system, and is sold to GeneticInc for research into latent Huntington's disease. This data is used for research and is then used to develop a new diagnostic tool that utilises the data of patient's known to have HD, to identify it faster in younger patients. The development of this system involves combining the HD dataset, with an age-illness correlated data set which, in doing so, makes T reasonably re-identifiable. T believes this use of their data goes beyond the collection they consented to.

T's situation is more complex than the previous two chapters, and this has been designed simply to highlight the two major steps in which informed consent issues arise. Firstly, consent to the treatment being provided, and then continued consent when their collected data is transformed and re-used later.

As with the other applications within this thesis, this is not intended to provide a comprehensive overview of the issues involved, or attempt to generate a solution, or

authoritative application of the law. This serves to provide a realistic representation of the practical application of the law to these scenarios, which could occur in the near future, to establish parameters for the later discussion of reform, if necessary.

### 7.4.1 SN1: A doctor utilises an AI

With discrimination, the distinction was made between the AI being used as the sole diagnostic tool, or one of many. When discussing consent, this makes no difference as consent should in theory be obtained for each individual procedure and action; five tools used results in five needed instances of consent. In this situation, the question is: did the doctor explain to the patient that the tool being used was an AI?

Regarding the HD diagnosis, it could be reasonably said that if properly communicated that this form of connection-based diagnosis was possible, that T had factored this risk into their decision to proceed. This voluntary assumption of the risk means T has put themselves willingly into a situation in which harm may arise (in this case the "harm" is their loss of choice over the HD diagnosis). This would be dependent on T's consent given freely given and voluntary, and with full awareness of the risks involved (in terms of both nature and extent). [515]

The difficulty here would be whether the doctor could have properly communicated this risk or even themselves been aware of it. It is likely that those utilising AI systems will be made aware of the capabilities of an AI, and the possibility of connections being drawn or unforeseeable results arising. Whether communicating this possibility to T is sufficient to amount to a proper description of the nature and extent of the risk is unclear. In *Unknown Unknowns: Surgical Consent During the COVID-19 Pandemic*[516]it is suggested that simply explaining the unknown nature of outcomes, and the possible forms of ways in which these can affect a patient, would suffice. A patient being made aware of the fact that things are unknowable, or at least difficult to predict, should still be able to expose themselves to that risk. In the case of diagnostics, the harms are comparatively minor (as opposed to during surgery or using an unproven pharmaceutical) so the extent of the risk is similarly minor. T would need to be told there is a risk of being diagnosed with something they might not wish

[515] *Sidaway v Board of Governors of the Bethlem Royal Hospital* [1985] A.C. 871
[516] Ava Ferguson Bryan and others, (2020) 272(2) Surgical Perspective

to know – T could then consider this risk, and their knowledge of the risk of HD, and proceed as they wish.

In respect of T's data, provided the doctor communicated that the data was being used in this way, there is no issue at the first stage of transfer.[517] The sale to Genetic Inc is also itself not an issue. At the point of sale, T's data is still (likely) to be anonymised. While T did not expressly consent to this continued use, if T remains anonymous, express consent from T is unlikely to be needed.[518] Once re-identified however T's condition of their consent, and the requirements under NZ law are both no longer met.

### 7.4.2 SN2: A lone AI interacts with a patient

The first question in this situation would be how was the information communicated to the patient? For both the Lone AI and the mobile application, it is most likely the case that the pre-treatment discussion occurs via a static, pre-prepared user agreement or disclaimer. There has been some discussion as to the impact of AI on these forms of agreement in respect of consumer guarantees[519] and digital privacy. In regard to informed consent, the issues raised are largely the same as when a doctor is involved: was the nature of the risk to T (in diagnosing something they did not wish to know) appropriately explained? The practical problem here is that an automated diagnostic mechanism is unlikely to be able to generate situational disclosures, and is reliant on a sufficiently broad, but still appropriately detailed, "general" agreement.

By removing the treatment from a situation involving a human verifying agent, there is the element of competence to note. Patients are, as discussed, presumed to be competent in the absence of contrary evidence. In the case of the Lone AI, the patient in some way has to be referred to, or permitted to utilise, the AI system. This would be the stage at which issues of competence need to be identified. This is not a live issue in this instance.

---

[517] Under the HIPC, Patient data can also be shared for research purposes without their explicit consent, so long as they are "unlikely" to be identified within the data; HIPC Rule 10(e)-(f)

[518] Under either the HIPC or Privacy Act, depending on whether the organisation is public, private, or in some way contracted to the health system at large. The principles largely operate on the same logic irrespective of source.

[519] Sara Gerke, Timo Minssen and Glenn Cohen, "Ethical and legal challenges of artificial intelligence-driven healthcare" (2020) Artificial Intelligence in Healthcare, p295-336, at 9

In respect of the transfer of T's data, the problems are largely the same as the first scenario. A user agreement is likely to include reference to the sharing or sale of data after collection, but would still likely be required to maintain the anonymity of the user. The moment at which T is re-identified, their consent is most likely required to be re-obtained before proceeding with their data (especially for an ongoing purpose like a dataset for an AI system).

### 7.4.3 SN3: A patient utilises a mobile application

The issues in respect of this scenario are the same as for the Lone AI. If the mobile application is not associated with the public health system (in other words, an entirely private enterprise) the only major difference is that the Code and HIPC will not apply to the situation, but the Privacy Act and associated protections still would. While they are generally less specific than the HIPC, the protections are largely the same in principle.

## 7.5 Conclusion

Consent is the integral underpinning of any action taken within healthcare, and its role in emerging smart-medicine is vital in ensuring a smooth, effective transition. Whilst AI's involvement complicates the application of a number of established rights within NZ's healthcare system, consent does not appear to be in this category. The requirements of informed consent are, while somewhat strained, still applicable and readily able to be fulfilled and executed within healthcare situations involving AI. The problems associated with AI, particularly in respect of complexity, unclear risks, and the disparate expertise of involved parties are of concern, are not unique but simply exacerbated by these systems involvement. The law has already developed allowances and interpretative mechanisms for these issues, and their application is sufficient for the near future. While caution should be exercised in regard to information given to patients, and how the uncertainties associated are conveyed, they do not represent critical failures for the application of informed consent. The potential for unexpected or unforeseeable actions being undertaken by an AI system results in a similar conclusion as for discrimination; unduly harsh burdens on those responsible for operation needs to be balanced against the protections being afforded to patients.

The primary problem that arises from consent is a recurring theme throughout these chapters; scale and transformation. While discussions of informed consent in conventional healthcare are largely a matter of one-on-one discussions and relationships, the inclusion of AI creates the landscape for this intimate setting to be overshadowed by the immense volume of choices and parties involved. AI exacerbates this concern by reaching new, unparalleled scales of Big Data, and also resulting in improper use often being revealed on an immense scale. Mechanisms for resolution are simply not prepared for the realities of the scale involved and this is an area of concern.

While dynamic consent is the principal issue that will arise for patients, it does not necessarily impair the ability to apply consent doctrine. The new capabilities and requisite transfers of data result in the need for caution on the part of those communicating information to patients, to ensure proper disclosure and consent has been obtained. However, little can be done both practically and reasonably on the part of patients when their data is transferred or transformed to multiple parties, especially if they lack awareness of when or how these changes are occurring.

# Chapter 8: Harm and Compensation

## 8.1 Introduction

A common question in discussions of AI is how liability is applied in situations involving AI systems. Who will be liable for harms caused by an AI, can an AI itself be liable, and can conventional tests for concepts like negligence be applied to AI systems or their use? This chapter will explore these questions in the context of medical negligence. This chapter will first analyse how NZ's unique no-fault accident compensation scheme, commonly known as "ACC", may mitigate the issues that arise when discussing negligence in relation to AI systems. Following this, there will be a discussion of the available actions in negligence, in light of the application of accident compensation, and an analysis of any issues that arise for its application.

Within common law jurisdictions, when an individual is harmed by a negligent action, they may bring a civil action against the person or body that causes the harm seeking compensation for their injuries. However, since 1974, NZ has operated an alternative mechanism, now colloquially known as "ACC", which can provide the necessary compensation without the need for legal action. Now given effect as the Accident Compensation Act 2001 (ACA),[520] the scheme provides coverage for personal injuries, although mental injury is covered in certain circumstances,[521] suffered within NZ.[522] Compensation under the Act can take the form of covering the costs of medical treatments, loss of income, social rehabilitation, vocational rehabilitation, and lump sum payments for

[520] First established in the Accident Compensation Act 1972, this was the result of the 1966 Royal Commission on *Compensation for Injury* which investigated expanding existing coverage, which was often very specific or short-timed. The first "no-fault" principle of compensation was introduced in the Workers' Compensation for Accidents Act 1900.
[521] Sections 21, 21A and 21B detail circumstances in which mental injuries are covered: mental injuries arising from certain criminal acts, work-related mental injuries. Also, mental injuries that arise from an otherwise covered physical injury are also covered (Section 26(1)(c)).
[522] Despite this legislation and its outputs commonly referred to as "ACC", it will be referred to in this chapter simply as "AC." This is to not create confusion during this analysis between the accident compensation (AC) provided and the Accident Compensation Commission (ACC) which is responsible for administration of the scheme.

permanent impairments.[523] In the event that the harm suffered is covered by the Act, the victim is barred from bringing another action against the person responsible for the harm, for compensation. As a result, negligence actions against medical practitioners are comparatively rare in NZ due to their narrowed availability. If the personal injury caused by medical treatment involving an AI is covered by AC, then the application issues for negligence that may arise from the AI's involvement are largely nullified.

In light of this unique interaction, this chapter will try to identify where potential issues caused by the utilisation of AI systems may arise in respect of medical negligence within NZ. To do this, there will first be an overview of the operation of AC and the requirements for coverage under the Act; this will include which potential actions are barred, to illustrate the areas in which wider discussion of liability may still be useful. This discussion of AC will conclude with an application of the three scenarios discussed within this thesis.

Following the discussion of AC, the chapter will discuss two alternate actions that a patient may wish to bring in situations involving an AI. The first of these actions, consumer protections, will only be mentioned briefly, to highlight that it is available. As consumer protections is a broad area of the law, which involves a number of different areas of discussion, it is simply too broad for analysis within this thesis. The second course of action, is an action in negligence; either under the Code, or under the common law. Within this chapter, the analysis will focus on the common law application of negligence, however both will be established in [8.4].

This chapter will illustrate potential gaps in liability or compensatory coverage for patients in the age of "smart medicine". These gaps will then be discussed in the following chapter to evaluate possible mechanisms for their mitigation or elimination.

It is often presumed in the literature that at an indeterminate point in the future, AI will be able to be held independently liable. This idea is linked to a much broader discussion on the legal status afforded to AI, and how this might shift with time. This chapter will not engage with this idea of suing an AI directly as this prospect would require substantial reform to the law, and will instead be left to the following chapter.

[523] Accident Compensation Act 2001, section 69

## 8.2 Accident Compensation

This section will provide an overview of the functionality and role AC plays within discussions of medical negligence. First, there will be an explanation of how coverage under the Act operates, and then the two main forms of coverage likely to be relevant to medical situations (treatment injury and personal injury by accident). Following this, there will be an explanation of the consequences of coverage, namely the effect of the statutory bar on further actions for compensation. With this technical detail provided, there will be a discussion of the questions raised by the involvement of an AI system; what issues may be caused when attempting to apply AC to such situations? To illustrate these questions, this section will close with an application of the three scenarios, outlined in [1.3.2.5], to highlight the potential interpretation of coverage in circumstances where a patient suffers harm as the result of treatment involving an AI.

### 8.2.1 Coverage under the Act

To receive compensation under the Act, all claimants must suffer "personal injury" (PI) as defined in s26, with exceptions or limitations imposed by s26(1A) – (5). The relevant subparagraphs for medical cases are found in s26(1)(a) to (c), which states that PI includes:[524]

> (a) the death of a person; or
>
> (b) physical injuries suffered by a person, including, for example, a strain or a sprain; or
>
> (c) mental injury suffered by a person because of physical injuries suffered by the person; …

As per Blanchard J in *Allenby v H*,[525] a "personal injury" means any injury suffered which has some "appreciable and not wholly transitory impact on the person".[526] This injury need not be long-lasting or serious bodily harm. This can range from broken bones or sprains from falling, to diseases contracted.[527]

---

[524] Section 26(1)(e) is relevant to healthcare at large, but will less likely to be relevant to the circumstances discussed within this thesis. 26(1)(e) states: damage (other than wear and tear) to dentures or prostheses that replace a part of the human body.
[525] *Allenby v H* [2012] NZSC 33, [2021] 3 NZLR 425
[526] At [56]
[527] *Allenby*, above n515, per Blanchard J

These personal injuries must be caused in the circumstances outlined in section 20(2), which lists a wide variety of situations, causes and exceptions to apply to s26(1). The list, from (a) to (j), covers a variety of circumstances which can be summarised as three main categories:

1. Personal injuries caused by an accident to the person (PIBANA);
2. Personal injuries that is consequential to, or considered, a "treatment injury"; and
3. Personal injuries caused by gradual processes, diseases, or infections that are consequential to a treatment injury, or other personal injury, for which the person already has cover under the Act.

This third category is largely an extension of the other two; when an injury which is already covered by the Act, creates circumstances in which further injuries arise, those injuries are also covered. This category will not itself be discussed in-depth, as the important considerations involved, such as what is a "gradual process" or "disease", arise during discussion of PIBANA.[528] Importantly, a person who is covered for a treatment injury, cannot also be covered for a PIBANA.[529]

The next sections will provide an overview of the requisite components of PIBANA and treatment injuries, as well as the exceptions to their application. They will be discussed in the order of treatment injury, and then PIBANA, as the former is most likely to apply in healthcare situations, and the latter is therefore relevant only in instances in which treatment injury cannot be met.

#### 8.2.1.1 Treatment injury

Treatment injury is defined in section 32. A treatment injury is a personal injury that was caused by treatment received[530] but is not a necessary component of, or expected consequence of, the treatment itself.[531] This treatment must be suffered by a person either[532]:

(i) seeking treatment from 1 or more registered health professionals; or

[528] The issues and considerations involved in (3), such as what a "gradual process" or "disease" is for the purposes of the Act, will be discussed when discussing (1): a personal injury caused by an accident to the persons.
[529] Accident Compensation Act 2001 Section 25(2)
[530] S 32(1)(b)
[531] S 32(1)(c), this takes into account both the patients underlying health at the time, and the clinical knowledge available at the time of treatment under (i) and (ii).
[532] S s32(1)(a)

(ii) receiving treatment from, or at the direction of, 1 or more registered health professionals; […]

"Treatment" is defined broadly under section 33, covering essentially all situations involved in the provision of healthcare. Treatment includes[533]:

(a) the giving of treatment:

(b) a diagnosis of a person's medical condition:

(c) a decision on the treatment to be provided (including a decision not to provide treatment):

(d) a failure to provide treatment, or to provide treatment in a timely manner:

(e) obtaining, or failing to obtain, a person's consent to undergo treatment, including any information provided to the person (or other person legally entitled to consent on their behalf if the person does not have legal capacity) to enable the person to make an informed decision on whether to accept treatment:

(f) the provision of prophylaxis:

(g) the failure of any equipment, device, or tool used as part of the treatment process, including the failure of any implant or prosthesis (except where the failure of the implant or prosthesis is caused by an intervening act or by fair wear and tear), whether at the time of giving treatment or subsequently:

(h) the application of any support systems, including policies, processes, practices, and administrative systems, that—

(i) are used by the organisation or person providing the treatment; and

(ii) directly support the treatment.

Generally speaking, this covers all conventional medical scenarios. A gap worth noting is treatment that results in a mental injury, but not a physical injury as well.[534] There are also a number of exceptional medical situations that fall outside the scope of this coverage, notably cases involving the costs of bringing up an unplanned child, born as a result of failed treatment,[535] or a child born with congenital disabilities due to pre-natal negligence for failure to warn the parents.[536]

If the procedure meets the requirements of s32(1) and s33, there are exceptions to treatment injury, outlined in s32(2) and (3). Treatment injuries do not include personal injuries brought about wholly or substantially by the patient's underlying health.[537] Similarly

[533] Section 33
[534] *L v Robinson* [2000] 3 NZLR 499
[535] *J v ACC* [2017] NZCA 441, [2017] 3 NZLR 804
[536] *Cumberland v Accident Compensation Corporation* [2014] 2 NZLR 373
[537] Section 32(1), (2)

if a patient unreasonably withholds or delays giving their consent for treatment, and this results in a personal injury, it is not a treatment injury.[538] Perhaps the most important for discussions of new, innovative treatments, is found under s32(2)(b): if the personal injury can be solely attributed to a resource allocation decision, it does not qualify as a treatment injury.[539] Failure to achieve the desired result of a treatment will also not be considered a treatment injury.[540] It was decided in *Adlam v Accident Compensation Commission,* that the treatment injury is required to result from a "departure from appropriate treatment choices and treatment actions."[541] This departure does not need to amount to the same standard of negligence, but is instead phrased as a failure to take a step required by an objective standard of care. The way this was conceived in *Adlam* is whether an "experienced specialist" in the relevant medical field, in the same circumstances as which caused the personal injury, would have done something differently. This judgment was applied in the later case of *Accident Compensation Corporation v* Shand, where it was held that a departure required an "alternative decision on treatment to not only have been possible, but to have been the decision that the treating medical practitioners should, in fact, have made." [542] If so, this omitted action was considered an "objective" necessity to the treatment, and its absence was a departure from the appropriate course of action.

The personal injury suffered must be "caused by" the treatment provided. This means it must be shown to have contributed to the outcome, not that it was a possible risk of the treatment and happened to occur. For example, if a particular method of treatment carries a risk of internal bleeding, it will not be enough to show that internal bleeding occurred after the treatment. It must be shown that the treatment itself caused the internal bleeding (the outcome) on the balance of probabilities; coincidental (even likely) injuries are not sufficiently causative. In respect of a failure to warn of a risk, which results in a treatment injury, it has been held that the standard of proof is on the balance of probabilities that the

[538] Section 32(2)(c)
[539] Section 32(2)(b); The realities of resource allocation is a common exception across healthcare situations. For example, failure to provide treatment because of funding or resource inadequacies does not violate the right to life under s8 of the New Zealand Bill of Rights Act 1990, see *Shortland v Northland Health Ltd* [1998] 1 NZLR 433 (CA)
[540] Section 32(3)
[541] *Adlam v Accident Compensation Corporation* [2017] NZCA 457, [2018] 2 NZLR 102
[542] *Accident Compensation Corporation v Shand* [2020] NZHC 2743, [2020] 3 NZLR 507 at [30]-[34]

failure to warn the patient caused the injury.[543]

#### 8.2.1.2 Personal injury by accident (PIBANA)

If a patient injured during the provision of healthcare services, or in another medical scenario, is somehow unable to meet the requirements of treatment injury, they may meet the requirements of PIBANA.

A PIBANA must meet the requirements of PI outlined in s26, discussed in [8.2.1], and occur in circumstances that meet the definition of "accident" in section 25. Section 25(3) states an important point for analysis purposes:

> The fact that a person has suffered a personal injury is not of itself to be construed as an indication or presumption that it was caused by accident.[544]

Accident is defined as "any of the following kinds of occurrences"[545]:

> (a) a specific event or a series of events, other than a gradual process, that—
>
> > (i) involves the application of a force (including gravity), or resistance, external to the human body; or
> >
> > (ii) involves the sudden movement of the body to avoid a force (including gravity), or resistance, external to the body; or
> >
> > (iii) involves a twisting movement of the body:
>
> (b) the inhalation of any solid, liquid, gas, or foreign object on a specific occasion, which kind of occurrence does not include the inhalation of a virus, bacterium, protozoan, or fungus, unless that inhalation is the result of the criminal act of a person other than the injured person:
>
> (ba) the oral ingestion of any solid, liquid, gas, fungus, or foreign object on a specific occasion, which kind of occurrence does not include the ingestion of a virus, bacterium, or protozoan, unless that ingestion is the result of the criminal act of a person other than the injured person:
>
> (c) a burn, or exposure to radiation or rays of any kind, on a specific occasion, which kind of occurrence does not include a burn or exposure caused by exposure to the elements:
>
> (d) the absorption of any chemical through the skin within a defined period of time not exceeding 1 month:
>
> (e) any exposure to the elements, or to extremes of temperature or environment, within a defined period of time not exceeding 1 month, that,—
>
> > (i) for a continuous period exceeding 1 month, results in any restriction or lack of ability that prevents the person from performing an activity in the manner or within the range considered normal for the person; or
> >
> > (ii) causes death

[543] *ACC v Ambros* [2008] 1 NZLR 340
[544] 25(3)
[545] Accident Compensation Act Section 25

In theory, any of the kinds of occurrences listed are possible in a healthcare situation, although whether they occur directly or at an AI's suggestion, is a matter of how advanced the relevant system is.[546] Of course, the exception provided by s25(2), mentioned in [8.2.1] applies; if the PI meets the requirements of treatment injury, it cannot meet the requirements of PIBANA also.

An exception to when something is a personal injury is found in s26(2); personal injury does not include[547]:

> personal injury caused wholly or substantially by a gradual process, disease, or infection unless it is personal injury of a kind described in section 20(2)(e) to (h).

This restriction is not important for treatment injury,[548] but is relevant to the application of PIBANA as none of the exemptions within s20(2)(e) to (h) apply. To clarify this from the earlier position mentioned in *Allenby v H* in [8.2.1]:

| **Personal injury** | **Cause** | **Covered** |
|---|---|---|
| Anything | Disease | No |
| Disease | Accident[549] | Yes |

This means PIBANA is not covered in if the cause is wholly or substantial one of the circumstances listed in 26(2). What these exceptions mean is discussed below.

A gradual process is, as opposed to a singular event like a car crash, where the injury suffered is caused by continued exposure to something which causes harm over time. Often relevant in workplace claims, examples might include where one's work involves prolonged

---

[546] For example, at current levels of advancement, it is more likely that an AI makes a decision which leads to one of these things occurring (i.e. an AI recommends a patient ingest a particular substance), as opposed to an AI being directly responsible (such as injecting the substance itself), although such systems are progressively being developed.

[547] Section 26(2)

[548] This is due to exemptions (f) and (h), which state that "personal injury caused wholly or substantially by a gradual process, disease or infection" which are consequential to a treatment injury, or consequential on treatment given to a person for another PI for which they are covered, is still covered by the Act.

[549] "Accident" as defined in s25(1)(b) and (ba) does not include the "inhalation or oral ingestion of a virus, bacterium, protozoan or fungus on a specific occasion unless this is the result of a criminal act of another person." How the disease itself was contracted would matter here, but this table is simply presuming that it was an accident. An example then might be: an individual is deliberately injected with a needle known to have been used by an individual with herpes. The individual contracts herpes via a criminal act (assault), via injection, and is then covered.

exposure to dangerous substances (like paint stripping materials) or constant loud noises which damage the ears. An example of this in a healthcare environment would be the developing of cancers from continued exposure to high intensity x-rays in radiology,[550] or the exposure to testing chemicals, such as those utilised in PET scans which have been linked to the development of cancers. Similarly, gradual processes can include illnesses developed during gestation such as spina bifida.[551] In *ACC v AZ*, the misreading of a pre-natal scan by the doctor led to AZ being born with spina bifida.[552] It was held that AZ's spina bifida was caused wholly or substantially by a gradual process.[553] In this instance, the misreading was carried out by an RHP, and was therefore covered under s20(2)(f). However, if AZ was otherwise reliant on coverage by PIBANA, s26(2) would deny them coverage.

Like a gradual process, a personal injury caused by disease (or infection) also excludes the sufferer from claiming compensation. When discussing disease, the law has divided diseases into two types: idiopathic and non-idiopathic disease. The former is where the disease is caused by an unknown source and is not consequential on other covered events. Idiopathic diseases are considered outside the scope of the Act.[554] A non-idiopathic disease is where an identifiable "source" is involved; for example, the inhalation of asbestos dust causing mesothelioma. These diseases, so long as interaction with this source is an accident, are covered under the ACA. Coverage for personal injuries involving diseases is therefore determined as follows:

| **Is there a known source?** | **Is the interaction with this source an accident?** | **Coverage** |
|---|---|---|
| Yes | Yes | Yes |
| No | N/A | No |
| Yes | No | No |

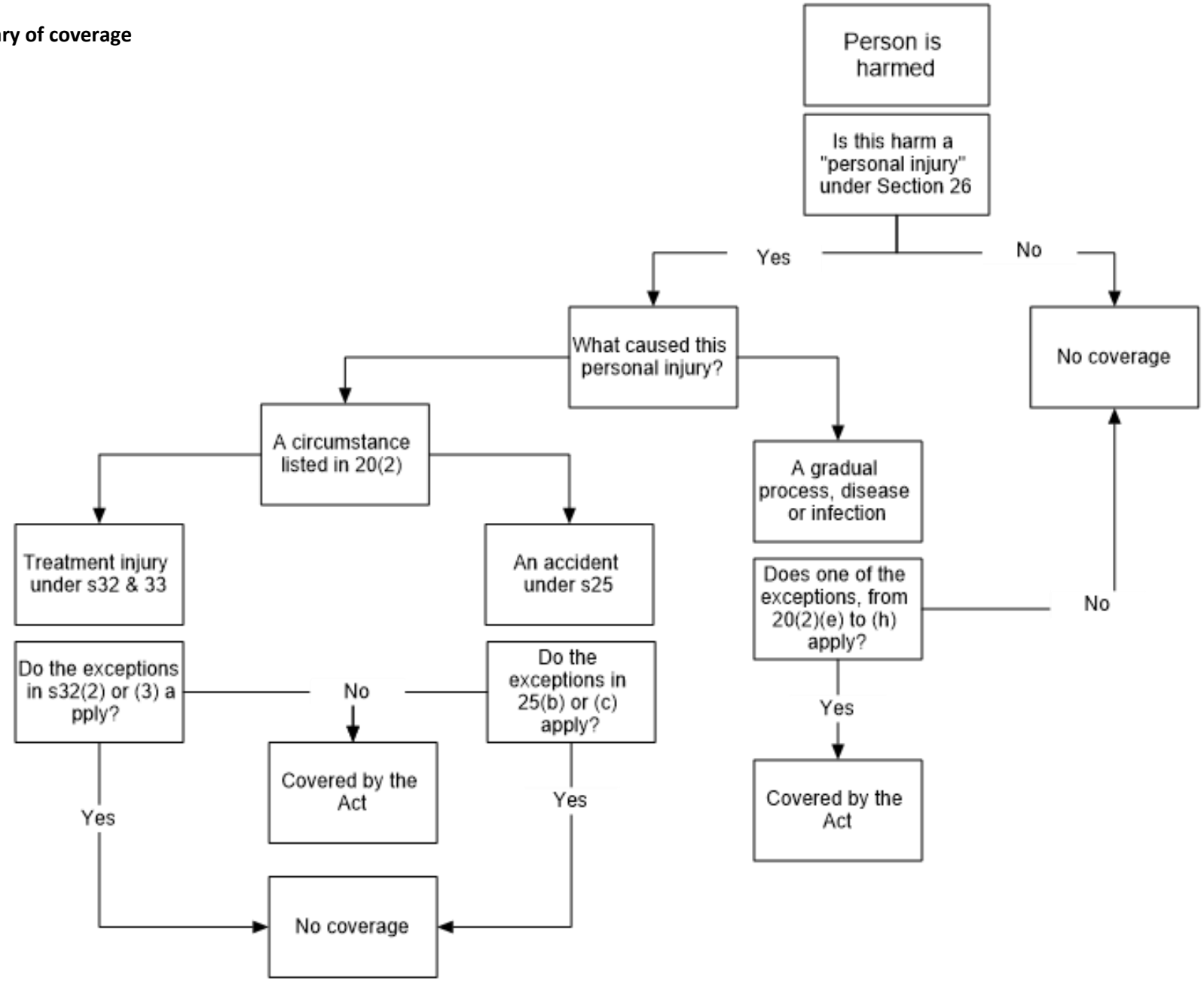
#### 8.2.1.3 Summary of coverage

A diagram has been provided below to consolidate the overview of coverage provided in the previous three sections. The next section will outline the consequences of coverage, or its lack, under the Act.

---

[550] Radiation and ionised radiation exposure is specified as an exposure risk in relation to both Accidents under s25, and work-related gradual processes under s30.
[551] *ACC v AZ* [2021] NZHC 2852
[552] This case relied heavily on the reasoning of *Allenby v H* [2012] NZSC 33, [2021] 3 NZLR 425
[553] *ACC v AZ*, above n541, at [61]
[554] *Accident Compensation Corporation v Calver* [2021] NZCA 211; [2021] NZLR 721

**Diagram: Summary of coverage**



#### 8.2.1.3 Consequence of coverage

This section will outline what the consequence of coverage under the Act is; namely, what actions are available to a patient who is entitled to coverage under the Act, and which are not?

If an injury is covered by AC, the injured party is barred from suing for damages to compensate that harm, which they may have done otherwise. This bar is provided by section 317(1) of the Act, and provides:

> No person may bring proceedings independently of this Act, whether under any rule of law or any enactment, in any court in New Zealand, for damages arising directly or indirectly out of –
>
> (a) Personal injury covered by this Act; or
> (b) Personal injury covered by the former Acts.[555]

It is important to note that s317 only comes into effect when there is "coverage" under the Act. Its intention is to prevent an individual from being compensated twice for the same injury,[556] not to prevent them from seeking compensation at all. *Queenstown Lakes District Council v Palmer*[557] describes the scope of the bar, Thomas J saying:

> "…the application of the Act and the corresponding scope for common law proceedings automatically adjust as and when the scope of the cover provided by the Act is extended or contracted. To the extent that the statutory cover is extended, the right to sue at common law is removed; to the extent that the cover is withdrawn or contracted, the right to sue at common law is revived."[558]

Simply put, if the damages sought arise from an injury that is covered by AC, the bar applies and prevents double-dipping by bringing a claim under the tort of negligence. If the damages arise from an injury that is not covered by AC, then the bar does not apply and an action in tort can arise. AC is concerned only with the provision of compensation and not with other

---

[555] Accident Compensation Act 201, Section317(1)
[556] *EM v Accident Compensation Corporation* [2016] NZHC 2535 at [33]
[557] [1988] 1 NZLR 546 (CA); this decision was made under the previous iteration of Accident Compensation, however the bar remained unchanged in its 2001 iteration.
[558] At [10]

functions of law that civil proceedings may afford a plaintiff. This means that claims seeking the vindication of rights[559] or exemplary damages[560] are not barred.

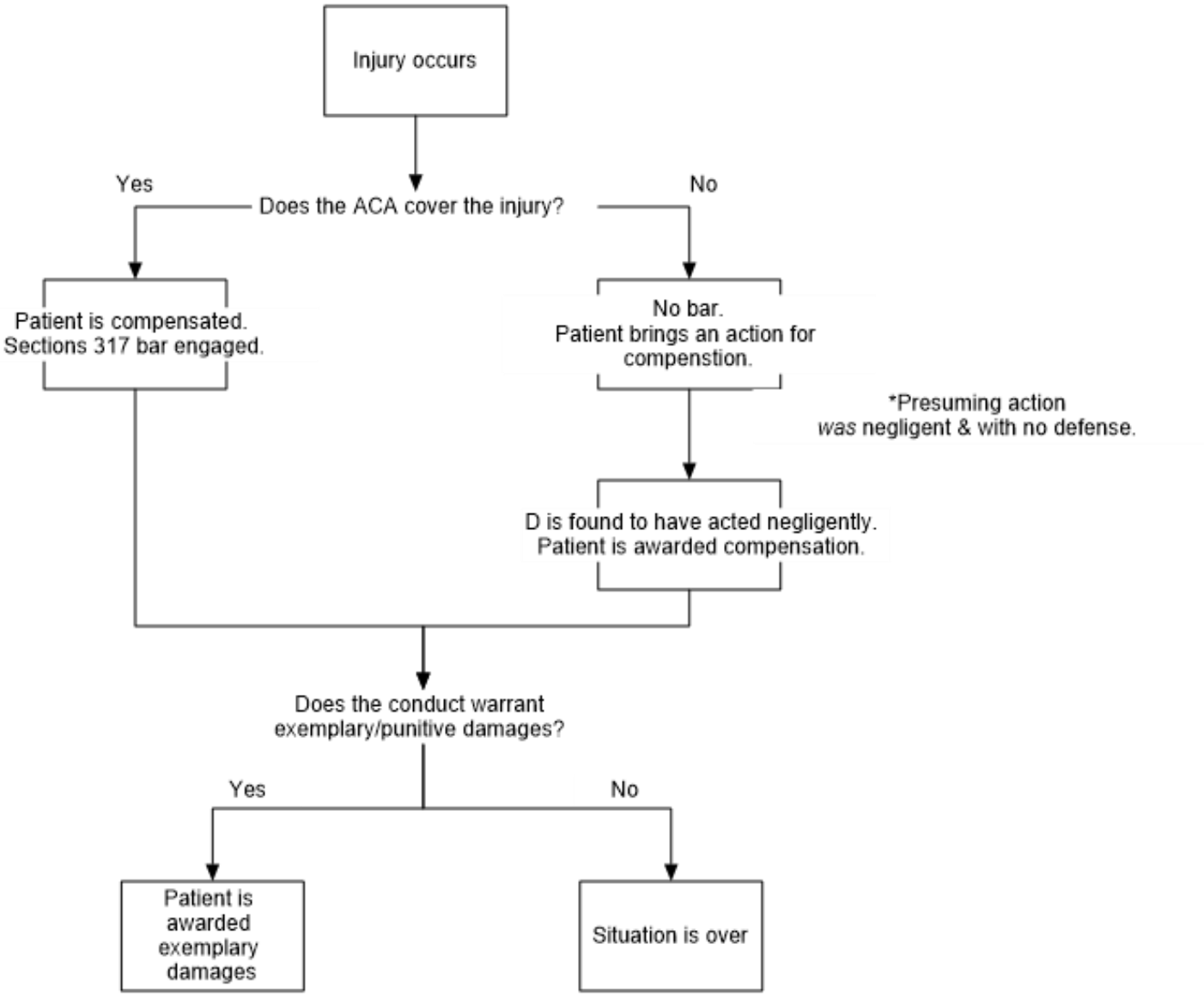
The awarding of exemplary damages in cases where the injury covered was first permitted in *Donselaar v Donselaar*.[561] Here the Donselaar family were engaged in a long-running, and often hostile, dispute. This resulted in J trespassing on his brother A's property. The resultant confrontation led to A hitting J over the head with a hammer, causing serious injury. J was entitled to compensation under the ACA, and wanted to sue his brother for damages relating to his personal injury. Section 5 (now s317) barred "suing for damages for physical injury covered by the Act", which led J to attempt to sue for exemplary damages. Quilliam J, in the High Court, held that separate damages for "a hurt to dignity and the like, if caused by the same conduct" [as the covered injury] were not permitted.[562] On appeal, the Court of Appeal held that exemplary damages were not prohibited by the ACA's statutory bar, and that the conduct of the defendant (here, the brother A) and not the injury suffered by J was the deciding factor on whether exemplary damages should be awarded, and not the injury suffered by J.[563] This outcome is now codified within section 319 of the Act. A claim for exemplary damages can be brought even when compensatory damages have been barred by s317, and irrespective of whether criminal charges have been filed.[564] However, it is important to note that when determining whether to award exemplary damages, or their amount, the court may have regard to any penalties that have already been imposed against a defendant.[565]

With the relationship between AC and the common law, NZ has a unique procedure for addressing injuries caused by allegedly negligent behaviour. Assuming a patient is only

---

[559] Stephen Todd and others, "Todd on Torts" (8th ed, Thomson Reuters New Zealand, Wellington, 2019), at 72. It should be noted that this point is somewhat arguable and there are no cases actually deciding that claims for vindicatory damages remain possible, notwithstanding the application of the ACA. This is of minor importance to this chapter, so will not be discussed in depth.
[560] *Donselaar v Donselaar* [1977] Wellington, A 454/76; Accident Compensation Act 2001, Section 319
[561] [1982] 1 NZLR 97 (CA)
[562] *Donselaar v Donselaar* [1977] Wellington, A 454/76
[563] This judgment is consistent with the general nature of exemplary damages as some form of punitive damages, and is stated here simply for explanation, and not to suggest *Donselaar* is some way reformulated exemplary damages at large.
[564] This was clarified in response to *Daniels v Thompson* [1988] 3 NZLR 22 (CA) that prevented exemplary damages being claimed against a defendant who had already been criminally punished. The formulation of s319 as allowing for "double punishment", and its overriding of the principle of *Daniels* has been criticised. See, for example, Todd and others, above n549, at [2.5(4)].
[565] Section 319(3)

concerned with compensating themselves and punishing the action involved, they would apply the following roadmap[566]:




8.2.1.4 Questions around the involvement of AI

With the requirements of coverage outlined above, this section will now consider two questions that arise from the involvement of an AI. The first is: What is a "registered health professional" (RHP) for the purpose of s32 and is this requirement compatible with the

[566] The final limb, relating to exemplary damages, is engaged with in the same inquiry as to whether the action was negligent. It has been separated here into a sequential step purely for visual aid.

changing landscape of smart healthcare? And the second is: if the failure of the RHP requirement results in "treatment injury" not being available, are there any potential issues with coverage under personal injury by an accident (PIBANA) caused by the use of an AI?

#### 8.2.1.4.1 The registered health professional

As discussed in [8.2.1], for a treatment injury to be covered by the ACA, the harm suffered must be towards a person who was:

> (i) seeking treatment from 1 or more registered health professionals; or
> (ii) receiving treatment from, or at the direction of, 1 or more registered health professionals; […][567]

This requirement of treatment given either by, or at the direction of, an RHP creates an interesting issue for prospective smart healthcare. Current uses of AI in healthcare are largely limited to AI systems being utilised by human medical professionals to supplement or augment their capabilities. In these instances, there is no issue as an RHP is involved. However, prospective developments for AI include entirely autonomous healthcare systems that can act independent of human agents, and could even be used by patients without first consulting with, or being referred by, a human agent. In these instances, it warrants a case-by-case analysis as to whether s32(1)(a) is still fulfilled.

Under the ACA, a registered health professional is defined as a "registered health professional of a type defined in regulations made under this Act."[568] The Accident Compensation (Definitions) Regulations 2019 provides a list of those considered RHPs under the ACA in section 7(a) and (b). These listed professions are all defined under s3 of the Regulations, and their definitions all refer to them as "a person who –".[569] This means that an AI cannot fulfil these requirements as it is not a person and therefore treatment provided solely by an AI system would not qualify as being provided by an RHP. However, prospective developments towards automated healthcare, in which an AI system would operate independently of a human agent are more unclear. An analysis of how this might be interpreted in practice occurs next in [8.2.3].

---

[567] Accident Compensation Act 2001, Section 32(1)(a)
[568] Section 6(1) registered health Professional
[569] The Accident Compensation (Definitions) Regulations 2019, Section 3

If the "treatment" is found to not be given by, or at the direction of, an RHP then it would mean the requirements for treatment injury cannot be fulfilled. Here the only available coverage would be personal injury by an accident (PIBANA) under s20(2)(a). While generally this will not be an issue (if an accident occurs, and harm occurs, the patient would still be covered), this coverage may also be unavailable if the s26(2) exception applies.

#### 8.2.1.4.2 The s26(2) exception

To re-iterate, if the personal injury was caused "wholly or substantially by a gradual process, disease, or infection" then it is not covered by the Act. If a patient was not covered by the treatment injury provision, and sought to rely on PIBANA, are any of the exceptions contained in s26(2) likely to be an issue? The two most likely to apply in medical situations, is where the personal injury was caused by a "gradual process" or "disease." Both are discussed in turn below.

##### A) Caused by a gradual process

If the personal injury suffered was caused by a gradual process, s26(2) prevents the sufferer from claiming compensation. A hypothetical instance where this would be relevant is where a patient interacted with an automated AI system as part of radiological testing, which could not be said to be at the direction of an RHP, and then developed cancer through continued exposure to radioactive material. This would not be a treatment injury and would be disqualified for coverage by s26(2) because the PI arose from a gradual process.

Alternatively, if a situation like *ACC v AZ* occurred, and the misreading of a scan was carried out by an independent AI, and was therefore only eligible for cover as PIBANA under s20(2)(a) (as no RHP was involved), the exception in s26(2) would disqualify the patient from coverage. This illustrates a potential gap in coverage for those engaging with innovative healthcare treatments. However, the difficulty here is accepting that an independently acting AI system would "misread" the scan in the way necessary for AZ's parent to be unaware of the spina bifida, and therefore for the gradual process to occur.[570]

[570] This question reaches levels of speculation about the capabilities of the hypothetical AI system, as well as the future uses of AI, which warrants future research.

B) Caused by disease

It is highly unlikely that the disease or infection exceptions are relevant to the involvement of an AI. An AI system on its own is essentially a processing computer. It does not emit or release particles, substances, or radioactive elements in a way that is known to cause disease (unlike say, dust particles causing mesothelioma). Diseases contracted while a patient was interacting with an AI would be (on current understanding at least) coincidental, and likely considered an idiopathic disease (a disease developed from an unknown cause). Such a disease would be outside the scope of the Act due to its unknown cause, and not being consequential on covered events. If an identifiable root cause exists, such as something emitted by an AI, then the disease would be non-idiopathic. If interaction with this source can be considered an "accident", then s26(2) would not apply and cover would be available.

Brown J, in *Accident Compensation Corporation v* Calver[571], states that personal injury is determined by the person's condition as a whole, and drawing distinctions based on the infliction of a disease, and the manifestations of that disease, was arbitrary. In this case, the personal injury was caused by an accident (the inhalation of asbestos), under s20(2)(a), and s26(2) did not prevent cover because it was *not* caused wholly or substantially by a disease. If such an occurrence did arise from the involvement of an AI, cover would therefore be possible, although as stated this is highly unlikely.

With these questions in mind, namely the issue of requiring treatment by or at the direction of an RHP, the next section will conduct an application of three scenarios, outlined in [1.3.2.5] to illustrate how this may be interpreted in practice.

### 8.2.3 Application of the three scenarios

For this analysis, the three scenarios will be discussed by applying the same medical situation, where an identifiable symptom was overlooked, leading to patient harm. This situation is:

[571] *Calver*, above n 544

> The patient, J, seeks advice about a concerning set of symptoms. J is diagnosed based on these symptoms only. Trusting the reliability of these results, no testing is done. Normal procedure for J's symptoms would involve a blood test, which would identify an additional, serious symptom. Without the blood test, J is prescribed a particular medication that poorly interacts with the undetected symptom, causing severe internal bleeding.

The question here is simply whether AC would cover the injury caused, the internal bleeding.

#### 8.2.4.1 SN1: A doctor utilises an AI

This requires no discussion. The "treatment" provided was a diagnosis under s33(b) and (c), and the harm caused can be causatively connected to the specific decision involved. The doctor who makes the judgment is a registered health professional (RHP) and their reliance on the AI does not impact the coverage provided by the Act. J would receive compensation under the ACA for their injury, and would be barred from pursuing other compensatory actions by s317. J still can pursue an action for exemplary damages if relevant.

#### 8.2.4.2 SN2: A lone AI interacts with a patient

In this scenario, when entering the GP practice J is sent through some form of "AI diagnostic room" in which no human practitioner engages with him. A personal injury results from the treatment, which would have been avoided had more traditional procedure been followed.

The only potential question is whether this treatment has been done "by" or "at the direction of" an RHP. It is likely this question would be interpreted broadly, with the intent of AC in mind. Had J been guided to "self-service" his care at the direction of an RHP this would reasonably be "at the direction of".

Presuming J is not directed on the day, the fact J knows to utilise the self-service AI system is likely the result of previously being informed by his GP practice. Given AC's intentions, it is probable (and likely) that an extended reading of what "at the direction of" means would be applied in this situation. The fact that the GP practice is allowing patients to utilise this

device, and allowing medical decisions to be determined by it, is an implicit form of endorsement. This could be argued as being treatment under the purview of the health professionals operating within the practice, as an extension of their own treatment. In this case, J would be covered by the ACA, and barred from pursuing compensatory actions by other means.

In the event that this process is entirely self-determined and automated, meaning J is not directed or instructed to utilise this device, the result may be different.[572] It could be argued that s32 is still satisfied if using the machine itself can be considered "treatment." Section 33 lists what treatment includes but does not define what it actually is. If the act of utilising this service was itself treatment, then J would most likely still be covered. Oxford's *Lexico* defines "treatment" as "Medical care given to the patient for an illness or injury."[573] The Collins English dictionary similarly defines it as "medical attention given to a sick or injured person".[574] The words "care" and "attention" would appear to suggest that some form of positive intention or association is necessary, and that autonomous processes, which are not overseen by a person, would not qualify as treatment. If this is correct then it is almost certainly not able to be considered treatment "at the direction of" an RHP either, thus disqualifying J from coverage under s32.

In respect of PIBANA, the question is whether J suffered his harm (internal bleeding) as a result of an accident. In this case, the cause was the interaction of his unidentified symptom with the prescribed medication. As the consumption of the medication was not the result of a criminal act by another,[575] this kind of occurrence does not meet any of the requirements under section 25 for an "accident." J is likely not covered for PIBANA under s20(2)(a) either in this scenario.

#### 8.2.4.3 SN3: A patient utilises a mobile application

[572] Although it should be noted, it is difficult to imagine a scenario in which this would apply, but the potential uses and integration of the technologies is still indeterminate.
[573] Lexico UK English Dictionary, available here <add link>
[574] Collins English Dictionary, available here <https://www.collinsdictionary.com/dictionary/english/treatment#:~:text=noun-,1.,2.
[575] Accident Compensation Act 2001, s25(b) or (ba)

This situation is very similar to that of the lone AI in SN2. The difference here is that the mobile application can be used when separated entirely from the conventional setting of healthcare.

The same conclusions can be drawn as in SN2: if J is told to utilise the application by a RHP, then treatment occurred at the direction of the RHP who recommended it. If the application was promoted by an institutional body, such as the Ministry of Health, as an official component of the healthcare regime, it could also be argued to be at the direction of an RHP. In this case, J would be covered by the ACA and barred from pursuing compensatory actions by other means.

If the application relied upon by J is not an official, or endorsed, avenue of treatment then it is almost certain s32 would not apply. It is also likely that PIBANA under s20(2)(a) would not cover J, for the same reasons as in SN2. However, this possibility is largely irrelevant for the circumstance being discussed. An unofficial application would not be in the position to generate useable prescriptions, resulting in the harm that J suffered. However, it may be that such an application "prescribes" J utilise a substance he does not need a prescription to acquire (over-the-counter medications). In this case, the same conclusion reached in SN2 would apply: this does not need the requirements of an accident under s25. It is more likely the case that J relies on a non-official source to self-diagnose, and then chooses either not to seek further treatment, or to self-treat with alternative means. In both cases, any injuries or harms suffered would not be covered by AC and would be the concern of J's own medical insurance to manage. With likely no coverage under the ACA, J would have to employ other potential courses of action.[576]

### 8.2.5 Conclusion

AC presents NZ with a unique benefit in the discussion of medical AI. Discussions of medical liability and the complications caused by AI for the conventional application of negligence rules have become common place in recent years, particularly within the USA.[577] In respect of the scenario discussed, SN1, ACA would certainly apply without issue. The

[576] Depending on *why* he acted the way he did, an action under costumer protections may be relevant to J, which are discussed next in [8.3]

[577] For example, see Matthew U. Scherer, "Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies" (2016) 29:2 Harvard Journal of Law & Technology

following two scenarios rely on perhaps a more stretched, but realistic interpretation of ACA's coverage and would likely also be covered as treatment injuries. In both SN2 and SN3, if the treatment injury provisions do not apply (due to the RHP requirement) then an individual would also not be covered by PIBANA.[578] This is a clear area of concern, that prospective automated medicine would not be covered by the ACA and would require patients to rely on other courses of action, such as negligence actions under the Code or common law, or even consumer protections (for the use of an automated service).

It is apparent that some clarity will be needed should AI become more common, and potential changes, to some of the requirements for AC coverage, namely the requirement of a "registered health professional (RHP)." It could be argued that this requirement simply reflects current realities; treatment is thus far provided by human agents who can be certified and disciplined professionally. In a prospective future in which the medical professional evolves to allow autonomous care, it is reasonable to assume that the consensus on what qualifies, and the circumstances in which it is considered safe, would be similarly expanded. Given that s33 provides a holistic list of every step of one's engagement with the system as "treatment", it is reasonable to assume that if an available (and presumably regulated) service is provided, it would qualify as treatment under ACA. If an option is then considered treatment under the Act, it should reasonably follow that any ill effects resulting from its provision to a patient is covered. A move from the requirement under s32 for treatment to be "by or at the direction of an RHP" to instead treatment that is regulated or approved by a governing authority (i.e. the Ministry of Health) would fill this coverage gap. In doing so, it would maintain the limit on covered health care to "official" care, and would not open the Act's coverage to potentially unwanted areas.

Whilst currently in the realm of testing, automated or digitised medicine presents the opportunity for human specific requirements to complicate discussions of application or dispute. However, the ability to apply the ACA not only provides a robust mechanism to simplify claims that may arise as part of AI's involvement in healthcare, but also a means to facilitate a smoother transition to smart-medicine.

In light of this conclusion, this chapter will now operate on the assumption that AC would apply to each of the three scenarios, and therefore all other actions for compensatory

[578] Accident Compensation Act 2001, Section 20(2)(a)

damages, by common law or statute, are barred within NZ. In case the interpretation applied, in respect of SN2 and SN3 is incorrect however, this chapter will engage in an overview of two other relevant courses of action: consumer protections and common law negligence. Negligence will principally be discussed in respect of the still available exemplary damages. Alongside an overview of their potential application, the potential issues that could arise for their application due to the involvement of an AI will be discussed.

## 8.3 Consumer guarantees and protections

As this thesis is concerned with the implementation and utilisation of new technologies, and how this may affect patients, a potential course of action relevant is consumer protections. Consumer protections in NZ arise from a number of different sources, which depend on who is seeking remedies (i.e. a patient against the manufacturer, or a hospital against the manufacturer), and the type of good or service involved (whether a device is for private or commercial use for example). Different legislation that could be relevant include: the Consumer Guarantees Act 1993 (CGA), and the Fair-Trading Act 1984 (FTA), and Contract and Commercial Law Act 2017 (which replaces the repealed the Sale of Goods Act 1908 (SOGA)).

With this thesis' focus on actions brought by, or protections afforded to, patients, the most relevant of the three scenarios to discuss would be SN3; an individual utilises an AI mobile application for their treatment. However, because consumer protections are an expansive area of law, and quite separate from the protections and processes involved in the Code and healthcare more broadly, this will not be discussed. An application and analysis of consumer protections to situations involving AI, both within healthcare or more generally, warrants being conducted in its own work. It is noted however that if the interpretation and application of the ACA above is correct, that consumer protections are also barred from being used to seek compensation for harm.

## 8.4 Negligence in NZ healthcare

Broadly speaking, negligence is the failure to exercise appropriate care, skill, or consideration of one's conduct. Within healthcare, this conduct can be criminally negligent, if

there is a statutory regime overseeing specific standards of conduct, or, conduct can be classified as tortious negligence, if the failure causes some form of harm, either economic or physical, to another person.

If a patient believes their care provided was negligent, there are two main courses of actions: either they complain that their right under the Code to services of a reasonable standard was breached, or they bring an action in the common law for the tort of negligence. As discussed in [8.2], when the ACA applies, utilising either to seek compensatory damages is barred. The relevant right within the Code, is Right 4, which provides[579]:

> (1) Every consumer has the right to have services provided with **reasonable care and skill**.
>
> (2) Every consumer has the right to have services provided that **comply with legal, professional,** ethical, and other relevant **standards.**
>
> (3) Every consumer has the right to have **services provided in a manner consistent with his or her needs.**
>
> (4) Every consumer has the right to have services provided in a manner that **minimises the potential harm to**, and optimises the quality of life of, that consumer.
>
> (5) Every consumer has the right to co-operation among providers to ensure quality and continuity of services.

The language of this right (see emphasis added) is many ways parallels that language used within the tort of negligence. Especially the requirement of "reasonable care and skill", as well as the provision of services consistent with "his or her needs". Because of this, to avoid repetition, only the tort of negligence will be discussed in depth within this chapter. Doing so provides the broadest possible coverage for the issues involved and allows the analysis to be applicable to a wider variety of contexts than just healthcare. Additionally, due to the overlap in language, the discussion of issues identified is largely applicable in both contexts. It is noted however, that the Code applies to a wider variety of situations, and provides a wider right as a result (particularly in respect of the right to services provided on ethical standards, and the right of co-operation). This inquiry will focus only on the components most relevant to the tort of negligence.

This section will first provide an overview of the test of negligence's three constituent parts: the duty of care, breach of the duty, and causation. Following this, there will be an

[579] The Code of Patient's Rights, see Appendix II, Right 4

overview of how "exemplary damages" are determined (the notable cause of action if compensatory actions are barred by the ACA), and the role vicarious liability (VL) plays in healthcare situations.

To conclude this section, a discussion of the issues that the use of AI pose to the application of these tests and requirements will be given. The following section will then engage in a detailed discussion of how these issues may impact the application of the test.

### 8.4.1 The test of negligence

Within NZ, negligence currently operates on a three-step inquiry: (1) the defendant owed a duty of care towards the plaintiff, (2) the defendant breached that duty; and (3) the breach caused harm to the plaintiff.[580] The second step, the breach of the duty owed, contains an additional step: that the breach suffered caused some form of loss (which the victim is seeking to compensate). This step will not be discussed within this chapter, as it is largely a factual issue; it will be presumed that within medical scenarios, that real loss occurs. This final step involves a three-step discussion of causation; whether the breached was a cause in fact; and whether the breach was a cause in law, and then whether this cause is sufficiently proximate to the harm incurred. Each of these steps will be discussed in turn.

#### 8.4.1.1 The duty of care

The first step is to establish that the party that caused the harm (commonly called the tortfeasor) owed a duty of care to the person harmed (such as the patient). The duty of care creates a legal obligation to take reasonable care to achieve the safety or well-being of others in regard to the conduct the duty encompasses. In *Donoghue v Stevenson*[581] Lord Atkin phrased this as:

---

[580] J.K.C Kingston, "Artificial Intelligence and Legal Liability," (2018) at 2.3.1; oftentimes (3) is split into two separate stages of whether harm has occurred, and whether this harm is sufficiently causal or too remote to be considered reasonable. The specific number of stages in the negligence inquiry has sometimes been represented as more or less than three stages, however the principal components are consistent. It is noted that this test has been recently expressed as 6 steps in a number of cases, such as by the majority *Khan v Meadows* [2021] UKSC 21, [2021] 3 WLR 147, which is discussed next. The choice to rely on the more summarised "3 step test" is for both ease of discussion, and to focus in on the key areas of concern for AI, as opposed to a fully technical discussion of negligence. However, it is noted that the reality of the test is much more complex than expressed here.

[581] (1883) AC 562

> "You must take reasonable care to avoid acts or omissions which you can reasonably foresee would be likely to injure your neighbour."[582]

A "neighbour" is someone who is both reasonable to foresee as being harmed through the conduct, and someone who is sufficiently proximate to the conduct.[583] Subsequent decisions included an additional requirement of whether it is "fair, just and reasonable" to impose a duty on the party in question, which may bring into account any broader policy reasons as to why imposing the duty would be desirable or undesirable.[584] All components of this duty formulae are unproblematic in regular medical situations; it is well-established that a medical professional owes a duty of care to their patients.[585]

For novel or borderline situations, to identify who might owe a duty of care outside the principal parties, the duty inquiry must be filtered through the requirements of what is "fair, just and reasonable."[586] This is not a limitless inquiry, although it can often seem to consider any number of considerations but is instead based on some form of underlying judicial principles. Todd summarises these guiding principles down to four main considerations:

1. A duty to take care should not interfere inappropriately with the autonomy of the defendant in deciding whether to act
2. The existence or extent of any duty that is imposed on the defendant should represent a proportionate burden of liability in respect of the wrongdoing in question
3. It should be appropriate for the courts to recognise a duty to protect a person in the position of the plaintiff
4. The proposed duty should operate coherently in the working legal system as a whole.[587]

The legal test for novel cases, which AI is most likely to be considered, has two parts: (1) whether the defendant should reasonably have foreseen the injury to their "neighbour," in the sense of a person who is closely and proximately affected by the defendant's conduct; and (2) whether there are any broader implications for the community in recognising or denying the

---

[582] At 580
[583] At 580
[584] *Caparo Industries plc v Dickman* [1990] 2 AC 605 (HL)
[585] *Kent v Griffith,* [2000] 2 All ER 474
[586] *Body Corporate No 207624 v North Shore City Council* [2012] NZSC 83, [2013] 2 NZLR 297
[587] Todd and others, above n549, at [5.40]

duty.[588] Both of these steps are important to discuss in relation to AI. Firstly, the range of people involved with AI's development and implementation, and their different understanding of its capabilities, results in a wide array of people with differing degrees of what constitutes "reasonable" foreseeability. Secondly, the technology itself comes with a number of societal considerations to weigh, in particular the value of the use of the technologies, as well as the continued stability and efficacy the healthcare industry.

In 2021, two decisions of the UK Supreme Court delivered on the same day, sought to clarify the scope of the duty and the nature of the principle. These cases, *Khan v Meadows*[589] and *Manchester Building Society v Grand Thornton UK LLP*[590] both sought to clarify the nature of the scope of the duty. A summary of both cases is provided below:

> **In *Khan*,** the claimant (M) approached her doctor to determine if she carried the haemophilia gene. Her doctor arranged blood tests for her, and the results were normal. However, K's tests were not sufficient for an accurate determination of the haemophilia gene. K told M that her results were normal, and her potential child would not have haemophilia. M later gave birth to a son who suffered from both haemophilia and autism. M sought damages in respect of both disabilities, which both parties agreed were reasonably foreseeable as a result of K's breach of duty.
>
> At first instance, K was held liable for costs associated with both disabilities. However, the Court of Appeal, and later Supreme Court, both held that K was only liable for the damages associated with the haemophilia.
>
> **In *Manchester*,** the defendant firm negligently advised their client, the claimant, that their accounts could be prepared using "hedge accounting" and that such a method gave a true and fair view of their financials. In reliance of this advice, the claimant carried out a long-term financial strategy, which unaware to them, served to hide their financial volatility. After seven years, the defendant realised the error of their advice and the claimant was forced to restate its accounts which showed significant losses. To extricate

[588] At 5.3.04(3)
[589] [2021] UKSC 21, [2021] 3 WLR 147
[590] [2021] UKSC 20, [2021] 3 WLR 81

itself, the claimant closed out their swapped positions at significant cost. They sought to recover the loss from the defendant firm.

After both trial judge and Court of Appeal held that they could not recover this loss, the Supreme Court allowed the appeal (although qualifying the amount available, due to contributory negligence from the claimant).

The majority judgments in both cases, delivered by Lords Hodge and Sales, both followed the same six-step test. The step of note here is the "scope of the duty" (or "question 2" as it was referred to in the judgment): What are the risks of harm to the claimant against which the law imposes on the defendant a duty to take care? Or phrased more simply, what, if any, risks of harm did the D owe a duty of care to protect the claimant against? Within healthcare, this requires a consideration of the nature of the service which is being provided (i.e. genetic testing for haemophilia), to determine which risks the law imposes a duty on the practitioner to exercise reasonable care to avoid.[591] In *Manchester* they said the court should determine the scope of the duty by looking to see what the risk the duty was supposed to protect against is.[592] The defendant (the accounting firm) had a duty to ensure that the claimant could assess the risk of its financial trades, and by providing their advice, the system adopted by the claimant obscured this risk. In *Khan,* this meant K only had a duty to exercise reasonable care to avoid the risk of haemophilia, not autism. This meant that K should have referred M for more specialised genetic testing, in order to establish whether she had the haemophilia gene.

In both cases, the minority judgments largely agreed with the outcomes, but varied in their specific approaches. Importantly, Lord Leggatt in *Manchester* phrased the scope of duty question in the language of causation. His version of the question was then: was there a sufficient causal relationship between what made the defendants advice wrong and the "basic loss" (the factual loss incurred)? Lord Burrows in *Manchester* was criticised for his focus on policy, with the others on the bench remarking that the scope of the duty was not dependent on issues of policy such as judgments of fair and reasonable.[593] This appears to suggest that the

[591] For example, see *Parkinson v St James and Seacroft University Hospital NHS Trust* [201] EWCA Civ 530, [2002] QB 366
[592] At [13]
[593] It is important to remember this case was decided in the UK, which has diverged somewhat from NZ's application of negligence. The "fair, just and reasonable" requirement is found within *Caparo,* above n574, and is still required within NZ.

court views the scope of the duty as an independent and established principle in the eyes of the UK court.

As it is unclear how these two cases will impact the formulation of negligence within NZ is, this chapter will presume that the NZ test in its formulation still applies, but will defer to the formulation of the scope of the duty employed by the majority judgment. This is because Lord Leggatt's gave his own view of the matter, which did not form part of the majorities reasoning. Additionally, Lord Leggatt's test already aptly fits within the terms of legal causation discussed later, and posing the question of duty as "what risk or risks are posed by the defendant's activity?" allows for the simplest application.

#### 8.4.1.2 Breach of the duty

When an individual or other legal person owes a duty of care, the person may breach this duty of care, by not acting with "reasonable care" and causing harm to the defendant. This breach can be caused by either positive actions, or omissions, but requires proof that the standard of care expected within the specific context has not been met. The standard of care is the standard of "reasonable" care expected of the theoretical "reasonable person" in the specific situation.[594] This reasonable person is a person with the skills and expertise of the person, irrespective of their personal idiosyncrasies, within the given position. As examples; doctor is held to the standard of the reasonable doctor, a surgeon to the standard of the reasonable surgeon.[595] This means that the reasonable person within medicine is one held out as competent for the particular procedure or process they are undertaking. Other factors then affect this standard, such as the social benefit of the conduct,[596] the known probability of risks,[597] and the gravity of such risks.[598] Professional standards can be identified by reference

[594] Originating in *Blyth v Birmingham Waterworks* (1856) 11 Ex 781 at 784
[595] See *Bolitho v City and Hackney Health Authority* [1998] AC 232 (HL) for an example of doctors.
[596] See *Tomlinson v Congleton Borough Council* [2004] 1 AC 46
[597] See *Bolton v Stone* [1951] AC 850
[598] See *Read v J Lyons & Co Ltd* [1947] AC 156 at 173

to professional manuals, training standards, or boards of certification and these are often used as evidence.[599][600]

The reasonable person must take reasonable care, which is care to prevent harm that is foreseeably likely to occur.[601] This harm can be physical, mental, or economic. Within a medical situation, this could be physical injury sustained during a procedure, emotional suffering caused by discriminatory care, or financial harm caused by mismanaged care that results in continued expenses. But importantly, the harm must be shown to have been caused by the breach itself, and not simply a conveniently timed occurrence. Nor can it be too remote from the breach to be considered reasonable to hold one liable for it. This last step is the "causation" inquiry.

#### 8.4.1.3 Causation

The causation inquiry is separated into three steps. These are: (1) whether the harm was a cause in fact,[602] (2) whether this was then a cause in law, and (3) the proximity between the cause and the harm incurred.

The starting point to determine cause in fact, is the "but for" test; would the plaintiff have suffered their loss but-for the defendant's wrongdoing? If the harm would have arisen without the defendant's wrongdoing, there would be no legal liability.[603] This first step is an objective factual inquiry, whereas the following two steps act as inquiry into the scope of the risk involved, and whether that risk was reasonably foreseeable for the parties who caused the harm.[604] In some circumstances, the "but for" test is replaced by the *Fairchild* principle.[605] Here, the question is whether the defendant's conduct in some way "materially increases the risk" of the kind of harm that befell the victim. This deviates from the balance of probabilities

---

[599] *Sulco Ltd. ES Redit & Co. Ltd* [1959] NZLR 45 at 88
[600] Importantly the court is not bound by such evidence, see *McLaren Maycroft & Co. v Fletcher Development Co. Ltd* [1973] 2 NZLR 100 at 107-108
[601] *Hamilton v Papakura District Council* [2000] 1 NZLR 265 (CA) at 279-281
[602] J Stapleton (2013) 129 LQR 39
[603] *Kuwait Airways Corporation v. Iraqi Airways Co. (Nos 4 and 5)* [2002] 2 AC 833 at [72]; the exception to this is when the tortious acts of two or more persons combine to produce indivisible damage.
[604] *Snell v Farrell* [1990] 2 SCR 311, at 326-328
[605] Which originates from *Fairchild v Glenhaven Funeral Services Ltd* [2002] UKHL 22, building on the earlier judgment of *McGhee v National Coal Board* [1972] UKHL 7, 1 W.L.R. 1

standard under the but for test, and is concerned with the creation of a risk of harm, as opposed to the causing of an outcome. This was because it was impossible to determine exactly where and when the fibres had been inhaled, and which of the plaintiff's five bosses were therefore liable. Lord Hoffman said in some cases to require that "liability depends upon proof that the conduct of the defendant was a necessary condition of the injury, it cannot effectively exist."[606] In this case, it was held that the material increase in a risk was sufficient to satisfy the law's requirement for a causal connection for liability.

For situations as complex as medicine, harm in which multiple parties contribute may be divisible or indivisible when discussing cause in fact. 'Divisible' means the harm can be individually attributed in parts to each party, and each party is then liable only for their specific contribution. Where the specific contributions are intermingled and it can be difficult to identify the apportionment, the test is whether the amount of each contribution can be reasonably ascertained.[607] Indivisible harm is when this apportionment cannot reasonably be achieved because the acts of two or more parties combine to produce the same damage. In this instance, both (or all) parties may be fully liable for the harm incurred.[608] The person harmed can also be responsible (although not liable) for having been a contributory cause of the harm.[609] In instances where two independent acts combine or coincide to cause harm, the question becomes whether each individual act materially contributed to the harm.[610] If a defendant's wrongdoing is found to the cause in fact of the harm, it may still not be considered the cause in law. Generally, this occurs because some intervening act or event occurs. The question then becomes whether the plaintiff's harm is within the scope of the risk created by the defendant's conduct.[611] This intervening conduct can either occur through a third-party,[612] natural or outside forces[613], or the plaintiff's own conduct.[614] The plaintiff's

[606] At [62]
[607] See *Wright v Cambridge Medical Group* [2011] EWCA Civ 699 at [80]-[84], in which divisible physical injuries caused by medical negligence were apportioned.
[608] Todd and others, above n 549, at [20.2.02]
[609] *Williams v The Bermuda Hospital Board* [2016] UKPC 4
[610] See, for example, *Kuwait Airways Corporation v Iraqi Airways Co. (Nos 4 and 5)* [2002] 2 AC 883
[611] Todd and others, above n549, at [7.5.02]
[612] *Al Kandari v J R Brown & Co.* [1988] QB 665
[613] *BNZ v NZ Guardian Trust* [1999] 1 NZLR 664
[614] *Corr v IBC Vehicles Ltd* [2008] 1 AC 844

intervention can either be apportioned as contributory, or even considered the "true" source of the harm, overriding the previous conduct.[615]

Lastly, a negligent act can result in a wide range of unexpected, distant, and far-reaching consequences and at what point these consequences are no longer considered legally liable is a question of remoteness. This means the relationship between harm and cause must be sufficiently proximate, or closely related, to hold the tortfeasor liable. Largely the rule is that tortious actions can only be sustained when the harm is a "foreseeable consequence" of the wrongdoing.[616] The harm has to be reasonably foreseeable as likely to happen, unless the risk is so minute that "a reasonable man would in the whole circumstances feel justified to neglect it."[617] If a plaintiff is especially susceptible to harm, and thus the harm suffered is greater, this is irrelevant; it is only a question of whether the harm of this kind was foreseeable.[618] Not all components of the resultant harm need to be foreseeable, it is largely determined as whether the kind of harm in question is reasonable,[619] and its extent does not need to be fully realised.[620]

### 8.4.2 Exemplary damages

Exemplary damages may be awarded when the tortfeasor's conduct is "advertent" or "reckless"[621] and warrants some form of punishment.[622] This is often discussed as a matter of breach of the duty, by recognising that there are degrees of breaches which can occur, some

---

[615] See *David Kimber TV & Sound Ltd v Kaiapoi BC* [1988] 1 NZLR 376

[616] *Hamilton v Papakura DC* [2000] 1 NZLR 265 at 282-284, affirming the judgment of *Cambridge Water Co. v Eastern Counties Leather plc* [1994] 2 AC 264

[617] Stephen Todd, Tort Law in New Zealand (3rd Edition, Kluwer Law International, Netherlands), at [544]

[618] Commonly known as the "thin skull" principle from *Stephenson v Waite Tileman Ltd* [1973] 1 NZLR 152. A plaintiff being of an abnormal character is a matter of circumstance and does not invalidate the connection between cause and harm. Affirmed in *Taupo Borough Council v Birnie* [1978] 2 NZLR 397 in respect of financial matters.

[619] *Corr v IBC Vehicles*, above n 604 (where an individual who suffered injuries became depressed and committed suicide. The suicide itself did not need to be foreseeable, but the depression was).

[620] *Taupo Borough Council v Birnie*, above n608

[621] *Couch v Attorney-General (No 2)* (on appeal from *Hobson v Attorney-Genera*l) [2010] NZSC 27, [2010] 3 NZLR 148

[622] *Taylor v Beere* [1882] 1 NZLR 81 (CA)

more severe than others. The intention is for these damages to punish the acts of the tortfeasor,[623] and may also serve to deter such conduct from occurring.[624]

In *Donselaar*[625] it was generally accepted that exemplary damages could only be awarded for intentional torts, where some degree of contempt, malice, or active disregard was involved.[626] On the same day, *Taylor v Beere*[627] determined that the test for such damages should be where the defendant was guilty of "outrageous misconduct." In *Bottrill v A*[628] the Court of Appeal held that ED's should only be awarded in cases of "advertent wrongdoing", where one was "consciously aware that the conduct is wrong."[629] The Privy Council accepted that to meet this standard, what was necessary was "gross negligence."[630] The minority rejected this, saying it was contrary to the "well-established principle that punishment should be inflicted on a defendant, no matter how gross his negligence, unless he had a guilty mind."[631]

*Couch*,[632]which now serves as the leading authority in NZ, clarified that this criterion of condemnation was too subjective and uncertain. Rejecting the Privy Council's majority judgment, and instead favouring the minority argument, Tipping J summarised the majorities position, and therefore new test, as[633]:

> "Exemplary damages should be confined to torts which are committed intentionally or with **subjective recklessness**, which is the close moral equivalent of intention. … Applying that principle to the case of negligently caused personal injury (that is,

---

[623] *Bottrill v A* [2001] 3 NZLR 622 (CA)
[624] An example of a medical situation that resulted in the award of exemplary damages was in *Green v Matheson* [1979] 3 NZLR 564 (CA). This case involved alleged medical experimentation on a plaintiff without their consent. "Green" in this case is Herbert Green, discussed in Chapter Four, responsible for the cervical smear controversy at Christchurch Women's Hospital.
[625] *v Donselaar* [1982] 1 NZLR 81 (CA)
[626] Todd and others, above n 549, at p1342
[627] [1982] 1 NZLR 81
[628] [2001] 3 NZLR 622 (CA), affirming the position taken in *McLaren Transport Ltd v Somerville* [1996] 3 NZLR 424 (HC)
[629] At [42]
[630] *Bottrill v A* [2003] 2 NZLR 721 (PC); interestingly, the PC also expressed disagreement with *Bottrill's* earlier determination that ED's are for punishment. Instead, the PC felt that their primary function was to illustrate the courts disapproval of "outrageous conduct", as a matter of condemnation, at [20].
[631] At [76]
[632] *Couch v Attorney-General (No 2) (on appeal from Hobson v Attorney-General)* [2010] NZSC 27, [2010] 3 NZLR 149
[633] At [178]-[179]

> injury caused through breach of a duty of care), **exemplary damages may be awarded if, but only if, the defendant deliberately and outrageously ran a consciously appreciated risk of causing personal injury to the plaintiff.** Whether running such a risk should be regarded as outrageous will depend on the degree of risk that was appreciated and the seriousness of the personal injury that was foreseen as likely to ensue if the risk materialised."

This means that largely accidental harms that arise from a doctor perhaps being careless will not amount to the necessary standard. Also, the award of exemplary damage is not determined by the degree of harm suffered, but by the degree of wrongdoing that *caused* the harm. If the doctor meets the necessary mental requirements set out by Tipping J, then their wrongdoing can support an action for exemplary damages.

### 8.4.3 Vicarious liability

Vicarious liability (VL) is a form of liability imposed on a third party, who did not personally commit the tort. VL is a form of strict liability imposed on those who are, in some way, expected to exercise control or authority over the tortfeasor. VL is determined based on the relationship that exists between tortfeasor and the third party, and if there is a connection between the tort committed and this relationship.[634] The most common relationship, and the one most relevant for healthcare, is that of employers and employees; a hospital (as employer) may be vicariously liable for conduct of the tortfeasor (a RHP). While the law has begun to recognise informal relationships akin to employment as giving rise to VL,[635] these are highly unlikely to be relevant in a strictly regulated industry like healthcare. The intention of VL is summarised by Todd as having three main considerations:

> (1) as an employer benefits from their employees conduct, they in fairness should be liable for their actions too;
>
> (2) an employer is more likely to be able to compensate a victim of harm; and

[634] Todd and others, above n 549, at [22.1]
[635] See P Morgan (2013) 129 LQR 139 for a discussion of *Various Claimants v Catholic Child Welfare Society* [2012] UKSC 56, [2013] 2 AC 1 [Christian Brother's case] which is the leading case in this area.

(3) VL distributes liability and therefore the cost of a tort which can be insured against.[636]

These considerations also incur an added deterrent effect for both tortfeasor and employer; the tortfeasor is motivated to act with greater care to avoid potential professional discipline, and the employer is motivated to impose greater oversight or mechanisms of control to avoid wrongs occurring.[637]

The employer is responsible for wrongs committed by their employee in the "course of employment".[638] The wrong committed by the employee is imputed to the employer as the authority under which they act. There needs to be a sufficient connection between the employment relationship and the negligent act, and not simply an act that occurred due to an opportunity provided.[639] The important thing to keep in mind about VL is that there needs to be an identifiable and "actionable" claim against the wrongdoer, before VL can be imposed. In instances where an action against a RHP may be difficult for whatever reason, VL against the hospital is not a "patchwork" solution. But VL may provide a mechanism to mitigate the burden on RHP's, particularly where the application of negligence may be considered unfair in light of the emergent technologies.

VL does not absolve the tortfeasor of liability, but on a practical level does typically substitute the other party in their place in respect of consequences. Often the employer will be responsible for the payment of damages but the wrongdoer is still jointly liable in principle, often explicitly through their employment obligations.[640] The same as regular liability, VL claims for compensation are barred by the ACA in the event of personal injury. In *S v Attorney-General*[641]it was held that exemplary damages could not be imposed on a vicarious basis, as the intention of these awards is to punish the wrongdoer for their conduct. The intended "moral condemnation" imposed would be on the wrong party, and therefore ineffectual.[642]

---

[636] Todd and others, above n 549, at [22.2.01]
[637] At 1217
[638] *Lister v Hesley Hall Ltd* [2001] UKHL 22, [2002] 1 AC 215 at [40]
[639] At [59]
[640] See *Lister v Romford Ice and Cold Storage Co Ltd* [1957] AC 555 (HL).
[641] [2003] 3 NZLR 450 (CA)
[642] Todd and others, above n549, at [22.2.02]

Due to the likely coverage provided by the ACA, and the bar on compensatory damages through negligence, the only likely available action to patients is for exemplary damages. As an employer cannot be vicariously liable for exemplary damages, VL is unlikely to be of great relevance to situations involving AI systems. If the discussion in [8.2.5] is correct, the only potential actions taken would be exemplary damages, and therefore the sole responsibility of the tortfeasor.

## 8.5 Issues posed by the involvement of an AI

With the requirements for negligence in mind, this section will now discuss the issues posed for applying these requirements to situations involving AI systems.

The issue that arises in relation to the duty of care in medicine is how far a duty of care can extend, and who owes a duty of care beyond the regular physician-patient relationship. While duty usually is not a live issue within standard medical negligence situations,[643] policy and proximity issues arise when discussing how far this duty extends, and how the chain of involvement is affected by the inclusion of an AI system. In situations where the care is provided "autonomously" it needs to be determined who owes a duty of care when a doctor or other medical professional is involved, or when there is no conventional first point of contact, like a doctor. And at what proximity, or distance, does it become unduly harsh to impose a duty. Issues arising when attempting to draw the line between those liable, and those not, are primarily due to the way AI situations can involve a number of parties conventionally not involved in medical negligence claims.

The issue that arises with the requirement of breach, when an AI is involved, is two-fold: firstly, what is the standard of care expected of different parties when an AI is being utilised, and secondly, whether fault for this breach can be determined, when an AI is capable of acting independently, or entirely unexpectedly.[644] Both of these issues are centred around the concept of fairness in negligence; the standard of care expected cannot be super-human, or the assignment of liability imposes a strict liability on doctors who are attempting to provide cutting-edge care.

---

[643] *Kent v Griffith,* above n575; Chief Medical Office, 2003 at 51 refers to the duty as "seldom challenged" regardless of circumstance within medicine; Todd and others, above n549, at [5.2.04(2)]
[644] This would be expressed as an issue of remoteness, wherein an AI acts in a way entirely unpredictable to someone interacting with it.

Whether the breach has caused the harm, the causation inquiry, is an issue of complexity; the decisions of AI systems are difficult to decipher in respect of "why" or "when" they occurred, and harm can be caused by different components provided by different sources. While the law has some mechanisms in place to deal with these issues in other areas, these need to be weighed against the benefits or needs of the technology in question. It is possible that too restrictive an approach could hamper the benefits of, and continued uptake of AI systems in the future. When determining whether something is the cause in law, it is important to factor in whether some harms of this kind are an acceptable concession, considering the needs of the technology to operate, and the benefit they provide. Discussions of indivisible harm and assumption of a risk need to also consider the fairness. While the harm may not be able to be adequately apportioned, this may result in too harsh a burden on different people involved in the situation. It is possible that the judgement of *Fairchild* provides some guidance here, both in the determination of liability for multiple parties, as well as the use of the "materially increasing the risk" test.[645] If this test were applied instead, the question becomes whether the use of the AI, which caused the harm, can be said to have materially increased the risk of that harm occurring. Some of the causal issues associated with intervening events, and the unforeseeable, still arise here, however it is perhaps more applicable than the standard but for test.

The discussion of causation and remoteness will focus on one main concern for each component (in fact and in law): firstly, the technical complexity and difficulties in determining causal issues in this context, and secondly, the potential cooling effect of applying negligence principles in novel circumstances. While the law has mechanisms to deal with, or mitigate, the issues of evidentiary complexity and identification, the discussion of causation serves to emphasis caution on the difficulties of applying causation principles to AI.

With the requirements and boundaries of negligence outlined, and the potential issues posed by the inclusion or involvement of AI have been noted, this section will discuss how these requirements may be applied or interpreted in situations involving AI systems. The three stages of the duty inquiry – duty of care, breach, and causation – will each be discussed in turn with the intention of illustrating the difficulties posed by AI, and where potential

[645] Mentioned above in [8.4.1.3]

shortcomings of the current formulation of negligence may arise. In light of this discussion, areas which may warrant reform will be discussed in the following chapter, Chapter Nine.

### 8.5.1 The duty of care

This section will identify those who potentially owe a duty of care to a patient when a human doctor is making use of an AI, or an AI is performing its task independently but cannot itself be held accountable. In other words, on whom is a duty of care imposed when an AI is responsible for treatment or diagnosis? Identifying where a line could be drawn, and where best it should be drawn, is an important step in formulating potential reform or regulation for AI's implementation, as regulation can cover the gaps in liability that cannot easily be encompassed by negligence.

This section will separate the potentially liable parties into two separate categories. Firstly, those within the hospital environment. The term "hospital" will be used to refer broadly to any conventional medical practice, including private general practices and community clinics. Secondly, those beyond the hospital setting. This will discuss those involved with an AI's development and management, within the technology, development, or administrative world.

#### 8.5.1.1 Within the hospital

For a number of parties, their duty of care is not a live issue. A doctor or nurse who treats a patient owes a duty of care irrespective of the technology or mechanisms they utilise. A hospital owes a duty of care to patients within it for the care provided by their employed health professionals.[646] The discussion will therefore only discuss parties beyond these conventional duties, highlighting those that arise specifically from an AI's involvement.

In 2018 the UK Supreme Court, in *Darnley v Croyden Health Services NHS Trust*,[647] held that as soon as a patient is acknowledged and "booked in" by a receptionist at a hospital, a duty of care has been established. The reasoning given was that it was reasonable to assign a duty of care if the situation falls within an "established category in which the law imposes a

[646] Note that *Woodland v Swimming Teaching Association* [2013] UKSC 66 suggested that this duty was a non-delegable duty as opposed to vicarious liability, although this has been met with criticism. See Christine Beuermann "Do hospitals owe a so-called "non-delegable" duty of care to their patients?" (2017) Vol.26:1 Medical Law Review, 1-26

[647] *Darnley v Croyden Health Services NHS Trust* [2018] UKSC 50

duty of care" already, as opposed to a case-by-case inquiry into what was fair, just and reasonable.[648] A summary of the facts is provided below.

> **In *Darnley.*** D was struck in the head and sought care at the Croydon Accident & Emergency (A&E) department. He informed the A&E receptionist he had a head injury and felt very unwell. The receptionist told him he would have to wait up to four to five hours before being seen. D replied, he felt he was about to collapse and could not wait.
>
> Normal practice (according to two receptionists, interviewed for D's claim), when a person with a head injury asked about wait times was either: (a) to say they could expect to be seen by a triage nurse within 30 minutes of arrival (according to one receptionist); or (b) to say that the triage nurse would be informed, and that patients would be seen as soon as possible.
>
> D left the A&E after 19 minutes, because he felt too unwell to remain. Shortly after, an ambulance was needed to take him back to hospital, and a CT scan found a significant extradural haematoma. D required emergency surgery, unfortunately, he suffered permanent brain damage in the form of a severe and greatly disabling left hemiplegia.

In this case, the established category is likely those who "hold themselves out" as having a particular skill or ability, which the victim relies upon. This can also be phrased as a form of induced reliance; by creating the impression of expertise that is relied upon by a patient, a duty is created.

In *Darnley,* the situation was not novel and fell within a recognised category of duty of care, but this decision did consider the scope of those owing this particular duty. That a duty of care is owed by those who provide and run an Accident & Emergency (A&E) department to those individuals presenting or complaining of an illness or injury, is now clear. The SC here decided that a duty should be owed by those who provide any information to waiting patients, so that the patients can make decisions based on correct information. In this case, where misleading information was provided about the time at which medical attention would be available, it is not considered appropriate to distinguish between medical and non-medical staff. Here the receptionist owed a duty not to provide misleading information in respect of

[648] At [16]

treatment where the information could foreseeably cause harm.[649] The decision appears to suggest that the role of the "holding out" party is not critical to the duty, once they become involved in the overall mechanisms of healthcare through provision of advice, they assume a duty of care. An individual could also hold themselves out as an authority on when best to (or how to) utilise an AI system.

At what point a defendant is holding out is a factual inquiry and would depend on specific circumstances involved in each use of an AI (or independent action of an AI). Logically this duty would exist for anyone involved in the process of a patient engaging with an AI system, if they provided information or a referral for it.[650] This means in instances like SN2, where there is no medical operator present, a potential action could be targeted towards someone owing a duty of care for referring the patient to that AI system.[651] Similarly, in SN3, if the patient utilised the mobile application through a referral or advice of someone within their GP practice, that person could be reasonably said to have induced reliance in the patient on the AI. In the case of SN1, an individual who refers a patient to a doctor who is known to use AI may owe a unique duty in how they make this referral or represent the role of the specialist and their use of the AI.

#### 8.5.1.2 Beyond the hospital

AI also presents the possibility for duties of care to be owed by people beyond those working within the confines of a medical setting. Most notably, this would be those responsible for the manufacture, development, maintenance, and implementation of any AI systems in place. This differs from standard manufacturer liability due to their creation of systems that make decisions; instead of designing and manufacturing an inanimate object to be manipulated by a professional, they instead create the professional. While traditional medical technological advancements augment the skills and expertise of human professionals, these new technologies may instead supplant them entirely. If a human doctor

[649] At [16]
[650] Whoever is the administering specialist, or "overseer" of the particular AI system is likely to be the sole individual with complete knowledge of it. However, those referring to this person would likely still be subject for the quality of information they express.
[651] Their duty would only be to provide factual and appropriate information in this referral, and to evaluate the risks of the patient interacting with the system.

performing a specific task has a duty to exercise care, then surely this duty of care also exists for those responsible for the design of the device that will replace that human doctor.

With uncertainty surrounding how exactly AI systems will develop, and the kinds of environments in which they will be employed, this line of inquiry is difficult to pursue. Issues that may arise from this line of thinking, particularly in respect to causation, are certainly worth investigating currently. However, the question of whether a duty is owed is perhaps too premature, without waiting to see how the technology in this area develops. Manufacturer liability would likely still be encapsulated by existing consumer guarantees and trade obligations at the current level of technology available on the market, however this would likely shift as the technologies became more "cognitive" in nature.[652]

#### 8.5.1.3 Conclusion

It appears likely that those within a hospital setting, regardless of their specific role, will owe a duty of care to patients with which they engage when an AI system is involved. Exactly what conduct they must take care of will vary (say, between advice given, or actions performed) but the involvement of an AI does not appear to raise issues in this case. For those beyond the hospital setting, it is less clear. While it may be likely that those responsible for the AI systems maintenance owe a duty of care, this is dependent on several factors still unknown. These are: how AI systems continue to develop, how they are regulated, and the interaction with other areas of law such as consumer protections.

### 8.5.2 Breach of the duty

Once a duty of care is established, the next step for liability is that the party must, in some way, breach this duty. Acting without "reasonable care and skill" or the skill expected of the "reasonable person", here the reasonable AI user or operator, will breach the duty of care. This section will discuss the issue of breach of duty in respect of three parties: a doctor who utilises an AI system, a hospital in which an AI system is utilised (in the event that the system acts autonomously), and the party responsible for the management and maintenance of the system. There are only a few issues that arise in respect of the first two parties, so only minor

[652] This is a position argued by John Buyers in "Artificial Intelligence: The Practical Legal Issues" (Law Brief Publishing: London, 2018). Due to the current realities of the technology in question, this position is still largely speculative and therefore outside the framing of this thesis.

comments will be made. This discussion focusses on breach of duty by human parties only. The idea of a "reasonable AI" will not be discussed here but is included in the discussions of negligence and legal personhood of AI in Chapter Nine.

#### 8.5.2.1 Reasonable care and skill

The standard of care for a doctor within any medical scenario is that of a reasonable doctor who has the appropriate skill and training, for that specific scenario.[653] For example, a surgeon performing open heart surgery (which suggests a certain, specific, level of training), is held to the standard of the reasonably qualified open-heart surgeon. It would logically follow then that a doctor utilising an AI system, who has had the appropriate training to do so, is held to the standard of care expected of a reasonable doctor trained in using that same AI system. In this instance, the issue seems to be relatively simple; when AI is utilised by a human doctor, the standard of care expected is simply the same as if it were another (albeit specialist) tool. For those parties other than the primary healthcare provider, their standard of care is directly related to the form of holding out they have exercised. For example, a receptionist would not be expected to give technical medical advice but can be expected not to provide misleading information in respect of the care.[654] In the event that a doctor calls upon the help of another in utilising the AI, such as a designated operator within the hospital, they will be considered to have acted with reasonable care. The standard rule is that when a person calls a reputable expert or source, and defers to their expertise appropriately, they have taken reasonable care. Failure to call upon expertise or help may result in liability, as a lack of skill or experience for oneself is not a defence.[655] This would be important when discussing who is permitted to utilise AI systems, and in the event of unforeseen outcomes or difficult circumstances. It is reasonable to assume that like most highly technical machinery, a sub-discipline within healthcare would develop for the use of AI systems (akin to a radiologist), at least until such systems become ubiquitous within healthcare generally.

Common practice, or ordinary practice, is the claim that one's actions did meet the reasonable standard of care because it accorded with what is conventionally expected or

---

[653] *Bolitho v City and Hackney Health Authority* [1998] AC 232 (HL)
[654] *Darnley*, above n637, at 24-27
[655] *Lyons v Nicholls* [1958] NZLR 409 (CA)

done. This evidence is important but not decisive[656] and will commonly arise during the discussion of whether reasonable care and skill were exercised.[657]

In *Bolam v Friern Hospital Management Committee*[658]it was held that a doctor was not guilty of negligence if they had acted according to properly accepted practice, dictated by a responsible authoritative body of the profession. In this case, even contrary opinion within the field was not enough to discredit the standard set by the profession.[659] This position was accepted for quite some time within the UK, however in *Bolitho v City and Hackney Health Authority,*[660] the test was clarified to require a "logical basis", "responsible" actions," and to be "reasonable" as opposed to simply being a black-and-white inquiry as to whether practice was or was not regular or common. The current position in the UK was set out in the case of *Montgomery v Lanarkshire Health Board.*[661] This case chose to move on from the infallible doctor who knows best, and instead regarded patients as persons holding their own independent rights which should be respected. Now, whether conduct is considered common practice is a case-by-case inquiry, considering the specific circumstances and needs of the patient involved.

There is a potential time of transition that raises some questions. Early adoption of AI can be said to have a higher risk associated with it, as the necessary adjustments and decisions are yet to occur to remediate the prospective problems. The question might then be asked: are those hospitals, or practitioners, who opt to utilise an early AI system acting with "reasonable care"? As discussed in the previous chapter, consent can be provided by a patient to undergo experimental treatment, which by its very nature is not the norm. However, failure to adhere to the norm (or at least some reasonable variation of it) is evidence of a failure to act with reasonable care. The use of experimental treatment needs to be weighed up against its value and necessity in each instance. Simply put, utilising an experimental treatment option for a patient who does not require it, when a normal treatment method is available, is a departure from the norm, and would likely constitute a failure to take reasonable care. When the use of new diagnostic or treatment methods shifts from being "experimental" to "common practice"

---

[656] *Sulco Ltd v ES Redit & Co Ltd* [1959] NZLR 45 (CA) at 87-88
[657] As a result, this is not a necessarily a "defence" employed after the fact, and is only an intermediary justification within the negligence inquiry.
[658] [1957] 1 WLR 582
[659] At 88
[660] *v City and Hackney Health Authority* [1988] AC 232 (HL)
[661] [2015] USKC 11, [2015] AC 1430

is unclear. Conversely, at what point innovative treatment becomes expected treatment is similarly imprecise. It could be said that once statistically better performing AI are commonplace, the choice to not utilise the system might become inherently negligent. It is almost certainly the case that cutting-edge medical technologies will be approved and implemented institutionally, and doctors will not be the ones making these determinations of use independently. But it is worthwhile to note that how early adoption of these technologies will be treated is somewhat unclear.

There are further discussions that could arise from this line of inquiry as to common practice. In the event of conflicting diagnostic conclusions, whether the doctor or AI should have the superior authority in treatment decisions, and whether a doctor who opts to not rely on an AI's decision has acted outside the accepted standard are both questions which will need to be answered in the future. However, much like the previous point, this is a grey area that will only begin to be understood and debated once these technologies become more commonplace, or at least more readily accepted.

Gerstner discusses the standard of care expected of an individual who creates a system. She asks whether the manufacturer of an AI is designing a system that performs a task that would otherwise be performed by a surgeon, are they to be held to the standard of the reasonable surgeon, or is the standard that of a reasonable manufacturer of a robotic surgeon? Gerstner recognises that there is no issue that a vendor of software owes a duty of care to the customer, in this instance the hospital, through contract, but whether this duty extends to the patient that is treated, and how the standard of care is applied in that case is problematic.[662] Gerstner argues that in the creation of an expert system, the creator must exercise care expected of that particular expert.[663] The logic being that they are creating a system that will supplant the expected and certified expertise of that person, the party responsible for its creation should at least exercise the supplanted human's care. Furthermore, the system itself must be capable of performing its tasks with the same care as the reasonable human equivalent.[664] How this standard would operate in practice is difficult to ascertain. Of course, an individual involved in designing how the AI functions, or "thinks" through its surgical choices, would not have the expertise or knowledge of a professional surgeon. Whether this

[662] M.E Gerstner, "Comment: Liability Issues with Artificial Intelligence Software," (1993) 33 Santa Clara Law Review
[663] At 12
[664] At 14

means that failure to consult with, or involve, appropriate medical experts in the development of an AI amounts to failure to take reasonable care is unclear. Attempting to compare how the standard of reasonable care applies across distinct disciplines is likely to be a difficult task. Following the earlier discussion of deferring to superior expertise, it could be argued the former position is more credible to apply. A developer of an AI system failing to adequately involve, and at least consider, the perspectives of the medical professional they are attempting to emulate, would likely constitute a failure to take adequate precautions.

A practical problem with the "equivalent care" position is that it shifts the standard of care into one that cannot be emulated, or does not practically exist. As an AI system is capable of performing tasks to a much higher degree of skill and care than their human counterparts, this means that there is no comparable human equivalent. Instead, AI systems would be only comparable to other AI systems, and those who develop them would somehow be required to exercise the care of a "reasonable superhuman surgeon." This is not dissimilar from the "blind faith in a prophet" problem discussed earlier in Chapter Two, in respect of verifiable outcomes, and raises the problem that it creates an expected standard of a reasonable person that does not exist. The practicalities, and even necessity, of these considerations is unclear.

Chapter Nine includes a discussion on speculative changes to the standard of care applied or expected that may arise from the emergence of AI. These changes, largely argued by John Buyers, present two main scenarios: firstly, the expectation of collaborative and pre-use determinations of care, and secondly, the prospective changes to the role of, and therefore expectations of, the human doctor.

### 8.5.3 Causation and remoteness

Once a breach has occurred, it must be shown that this breach was the cause of the harm, and that this harm was not too remote from the breach to be considered reasonable. This causation requirement is problematic for most situations involving AI, due to the complexity of the technology itself[665] and the difficulty of identifying causes or reasoning for actions of said systems.[666] Schiff provides the list of coders and designers, medical device companies, physicians, and other healthcare professionals, hospitals and healthcare systems, regulators, insurance companies, pharmaceutical companies, medical schools, and others, who have

[665] Buyers, above n 642, at 21-22
[666] See Chapter 2.4 for a discussion of the Black Box Problem (BBP)

important responsibilities in developing, testing, training, applying, and evaluating sophisticated technologies in healthcare.[667] This results in the "many hands" problem that the fact that there are different professional or ethical standards, different timings of involvement, and (as discussed in Chapter Six) different jurisdictional boundaries mean it can be difficult to fairly apportion fault.[668]

#### 8.5.3.1 Cause in Fact

Cause in fact (CIF) can be a difficult determination in medicine for two main reasons: (1) harms that remain unseen for an extended period, or change over time, and (2) the wide range of parties, and the complexity of the technologies in question involved, create a difficult web of complexity for pinpointing specific causes. Buyers discusses the issues of causation in AI inquiries in a number of different areas of law, and concludes that the complexity of AI is an immense roadblock for this requirement in practice.[669]

Current state-of-the-art AI often do not provide for these systems to self-report, or record the cause of outcomes, which is further complicated by the nature of AI as a black box. Buyers uses the example of driverless cars for his discussion, primarily focusing on issues of product liability within the United States of America (USA) but the reasoning can be applied to negligence more generally. Driverless cars rely on the aggregation of several technologies, all designed, patented and manufactured by different parties to create an overall system. It is not as simple as saying "the brakes failed" like in a conventional vehicle, and then blaming the brake manufacturer. Within a driverless car, the smart parts all function as components of a singular system, as opposed to numerous distinct parts working in unison. This mechanical complexity creates an almost endless sea of liability targets and is compounded by the difficulty of identifying where the specific defect that caused the harm originated.[670] These machines, depending on the complexity of each component, are also not acting on a prescriptive instruction set, but a system of rules that may or may not have anticipated a particular circumstance. As a result, the way they interact or respond also increases the

---

[667] Daniel Schiff and Jason Borenstein, "How should clinicians communicate with patients about the roles of artificially intelligent team members? (2019) Vol.21(2) AMA Journal of Ethics, E138-145 at 141

[668] This is a problem Tokio Matsuzaki briefly discusses in "Ethical Issues of Artificial Intelligence in Medicine" (2018) Vol.55 (1) California Western Law Review

[669] Buyers, above n 642, at 25

[670] At 28

complexity of outcomes, especially when there are systems from different providers involved.[671]

Treating an AI system as a complete, whole system and therefore the harm as being indivisible would be a possible solution to this issue. This is a question of how AI is being defined in use; how it is regulated and permitted in use, and who then is responsible for its use. There are three potential options that could be taken:

1) treat the harm as indivisible, and all those responsible for the creation and use of the AI are equally liable for its faults.

The resistance to this would come from companies feeling unduly burdened to ensure the contributions by each different entity meet their own standards, or the standards prescribed by the law.[672] If harm is indivisible, all parties who contribute to the harm are fully liable. This rule, known as "liability in solidum." This would serve however to ensure more responsible integration and openness between developers and those utilising an AI. It has been noted that indivisible harms would lead to higher industry standards, or at least a centralisation or internal mechanism of policing, to ensure that the risk is appropriately accounted for.[673] Conversely there has been criticism of the idea suggesting it would lead to more secrecy and monopolies by larger entities attempting to protect themselves by providing all components themselves.[674]

2) To require more in-depth standards of self-reporting within AI, so that individual apportionment is more practical.

This is largely an issue of policy and not legal application.

3) Once an AI system has been approved for use and integrated into a healthcare system, the operating institution (or overarching system) assume responsibility for the risk.

---

[671] At 29
[672] However, in some instances this may be considered a worthwhile concession, this would likely result in issues for competition and anti-trust law. Larger companies may attempt to acquire smaller companies to prevent the potential unfair liability applied by treating components as indivisible.
[673] Sarah Green, "The risk pricing principle: a pragmatic approach to causation and apportionment of damages" (2005) Vol. 4, Law, Probability and Risk, 159-175
[674] Scherer, above n 567, at 390

This would mean that those responsible for the systems availability, assume responsibility for any harms it can be said to have caused. Of course, this would rely on exceptions being included for blatantly egregious conduct by those utilising the system, that which amounts to exemplary damages, but this is largely a matter of drafting.

#### 8.5.3.2 Cause in Law and remoteness

Intervening events (IE), in which a secondary event occurs in between the breach and the harm, can further complicate causation inquiries. These is defined by Scherer as "an intervening force or act that is deemed sufficient to prevent liability for an actor whose tortious conduct was otherwise a factual cause of harm."[675] They are often treated as raising causal issues but can also be considered evidence that the harm was too remote, due to the IE occurring.[676] IEs are likely to occur within medicine, especially when there are a number of doctors involved in a patient's care, and when a number of different technologies, medicines, and practices are employed. IEs have been discussed as a likely outcome in AI cases in general, especially within medicine.[677] The issue in respect of IEs is, whether the IE itself falls within the risk of the conduct that is in question. Answering this is a highly contextual inquiry, and will be reliant on what the identified risks for which the party which owes a duty of care is responsible for.

It has been argued that due to the unpredictable nature of an AI, its advice or actions could always be considered an IE to harm that might be caused by a doctor.[678] This will likely be the case until understanding of AI has advanced considerably, especially in the realm of accountable AI systems that self-report. These situations are mostly dependent on determining that it is no longer reasonable to hold one individual liable for harm caused, when such an unforeseeable event occurs. However, such argumentation would likely rely on the use of the AI being the only reasonable course to pursue; if other options were available, being reliant on accepting the unpredictable nature of an AI would be less convincing.

---

[675] At 365
[676] *McKew v Holland & Hannen & Cubitts* [1969] 3 All ER 1621 (HL) at 1623
[677] For example, Scherer, as above n 567.
[678] At 82

It is possible (especially in the early days of implementation) that the actions, decisions, or "thought-process" of AI will be too obscure and varied to be considered foreseeable. An issue with foreseeability of harm in respect of AI is the fundamental difference in decision-making. As Nate Silver explains, when discussing the example of a chess-playing computer:

> 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 circumstances. For a human player, this would have required the creativity and confidence to see beyond the conventional thinking.[679]

Foreseeability in causation is decided on the basis of human-decision making, and how their decisions are understood. However, the logic of AI is different, and how the system reaches its conclusions is not bound to the same cognitive limitations, biological processes, or personal experiences that guide a human's judgment. As a result, trying to apply similar foreseeability metrics to the outcomes that may arise from an AI system is difficult, not only due to their disparity with people but also to all other conventional technology in play.

Buyers uses the example of the UK test in *Scott and Bennett*[680], which applied a three-stage inquiry for causation not too dissimilar to the NZ test. The point he isolates is that identifying an AI as being "under the control" of its user is an issue which will undoubtedly result in great debate. Holding a doctor liable for the actions of a thinking, learning machine, by asserting that they were in control of it, will no doubt be the main hurdle of negligence actions involving AI.[681] Buyers also remarks that the test in Scott does not provide a mechanism for dealing with the inexplicable isolated incident, which due to the learned environmental nature of AI is likely to the primary example of harms.[682]

As discussed by Chinen, the more autonomous AI systems become, the "more tenuous becomes the strategy of attributing and distributing legal responsibility for their behaviour to

---

[679] Nate Silver, "The Signal and the Noise: Why So Many Predictions Fail – But Some Don't" (Penguin Group, New York, 2012)
[680] *Scott and Bennett v Chemical Construction (GB) Ltd* [1971] 3 All ER 822
[681] Buyers, above n 642, at 32
[682] At 33

human beings.[683] While the decision to utilise the system firmly rests with the human agent, it is debated that there will come a point where their prospective outcomes are both unforeseeable, and sufficiently autonomous to be considered a distinct action. Causation within AI negligence decisions will likely be a practical barrier for some years to come internationally, due to the relatively juvenile state of the AI field. Specific determinations of the limits of causation, when harm is too remote, or exactly what qualifies as a cause of law is a series of cases and decisions that will need to be made over time. However, for now, it's important only to identify where these potential problems may lie.

## 8.7 Conclusion

The issues of how legal liability for negligence will be apportioned and determined in situations involving AI has become a hot topic in recent years internationally. It is apparent that there are a number of issues now (and prospective issues in the future) as the technologies in question continue to develop. It appears however, that due to the Accident Compensation Act 2001, NZ has a sufficient (and readily available) mechanism to avoid many of these issues. AC, by providing coverage for harms caused, and barring the ensuing negligence actions, provides NZ with the time to prepare for the coming technological developments more adequately, by alleviating the most immediate concerning situations. AC appears to cover the three situations discussed throughout this thesis, albeit with a somewhat stretched application, eliminating the practical bulk of potential problematic claims that international academics fear.

It is likely the case that the problems discussed in respect of negligence largely amount to speculative concerns, where the discussed technology reaches levels of cognition and independence currently not practical. In these cases, it is difficult to adequately discuss the legal challenges that arise, because the evolution of law in the interim may make such comments a nullity. For now, the most generic of situations likely to occur are likely covered by the ACA. This buffer affords NZ the time necessary to continue these inquiries more minutely, and observe the developments occurring in similar overseas jurisdictions.

---

[683] MA Chinen, "The co-evolution of autonomous machines and legal responsibility" (2016) Virginia Journal of Law & Technology

In light of the conclusion of this chapter, Chapter Nine's discussions in respect of negligence will largely engage with proposals and speculative problems that may arise. Because there is little immediate concern for this area in NZ's jurisdictions, attention will be paid to potential future changes that may arise internationally and be relevant in NZ, and to some of the broader proposals which may amount in systemic changes arising.

# Part C

# Reform

Part B was dedicated to identifying and analysing the potential legal issues that arise when applying conventional legal rights and tests to situations involving AI systems. Part C will consolidate the conclusions reached and provide a series of recommendations for how best to approach the near-future implementation of AI in New Zealand healthcare.

The purpose of this part is three-fold:

(1) to summarise the thematic ideals present within healthcare and the conclusions reached about the application and interactions of legal mechanisms to AI systems;

(2) to provide discussion on potential issues associated with reform, and then recommendations on how to proceed; and

(3) to provide guidance on future research that seems warranted considering this thesis' scope and conclusions.

This purpose will be fulfilled in Chapter Nine, "Reform and Recommendations," which will serve as the final substantive chapter of the thesis.

Chapter Ten, "Conclusions", will summarise the overall coverage of this thesis and its contributions.

# Chapter 9: Reform and Recommendations

## 9.1 Introduction

In this thesis so far, the issues that arise when attempting to utilise current legal formulations of patient rights in situations involving AI have been discussed. This chapter will now serve to summarise these preceding discussions as well as provide a series of recommendations for how to best approach the near-future implementation of AI in NZ healthcare.

This chapter will begin with a summary of the different conclusions reached throughout Part B of the thesis. This summary will be divided into "General issues" identified with AI itself, and chapter-specific summaries for each of the areas discussed in Chapters Five to Eight. Both the areas in which the law is currently acceptable or readily able to manage the introduction of AI, as well as the areas in which conflicts or difficulties with applying the law have been identified, will be outlined to provide a comprehensive overview of the conclusions of this thesis.

Following this, there will be a discussion of some different conceptual reforms for some of the issues discussed and their practicality, for example, the idea of "embedding fairness" into AI systems, to avoid the issues identified in Chapter Five "Non-Discrimination and Bias." This section will not serve to provide an exhaustive look at ideas that have been published relating to healthcare AI, but instead will focus on a select few areas which further emphasise the unique issues involved in AI technologies and their regulation. Perhaps the most notable reform which will be discussed is regarding the legal personhood of AI. This will be engaged with briefly but will not engage with the wider debate of AI cognition and status, which warrants a more comprehensive interdisciplinary analysis.

Following these summaries, there will be a series of recommendations on the way in which NZ should prepare for the implementation of, and continued use of, AI systems in healthcare. These recommendations will be a mix of legal and policy-based recommendations intended to facilitate a smoother transition from the current state-of-affairs to the prospective smart healthcare of the future.  is for these recommendations to be of minimal severity, so they are easy to both apply and enforce, while future research is done into other technical or specific

areas required. The intention is not to provide a definitive set of changes which must be undertaken, but a range of changes which best serve the themes outlined in [9.2]. If these changes are implemented, the immediate future of NZ's smart healthcare may not be without fault but should be equipped with the ideas necessary to manage the transition.

The chapter will conclude with some brief thoughts on reform and AI more generally, before the thesis' overall conclusion is provided in Chapter Ten.

## 9.2 Summary of thesis position

This section will provide a detailed overview of the different issues identified throughout the preceding chapters in respect of managing the impact of AI on healthcare. First, general issues associated with AI, largely discussed in Chapter Two, will be outlined. Secondly, the specific issues associated with each right discussed in Chapters Five to Eight will be provided. A table is provided afterwards to summarise the conclusions drawn in the application of the three scenarios, and to supplement the explanations provided for the specific rights.

### 9.2.1 General issues associated with AI

The most fundamental issue of AI from both legal and policy perspectives, is how to define the specific technologies at issue. Discussed in Chapter Two, AI comprises a broad spectrum of technologies with different purposes, functionalities, and guiding principles or processes. Treating these technologies as homogenous is potentially deceptive, and even potentially dangerous due to the breadth of outcomes involved. To ensure proper regulation or application of the law it is important to ensure that the definition of the technologies has been clearly described and agreed upon. This is important both in the context of common law discussions, where too narrow of a definition may result in other AI situations which are not captured by the decision, or too broad definitions which result in the decision having unintended consequences. Clarity is needed within the legislative regulatory sphere to ensure that the purpose and intention of the regulations can be fulfilled. While this thesis did provide a working definition of an AI within the context of this discussion, it is important that such a broad-spectrum approach is not taken in every context. The differences in each technology's

use, development and capabilities will warrant a careful analysis and definition for each circumstance, and sector, of implementation.

Although less at issue within the context of this thesis, the often international nature of Big Data is something that will affect all facets of AI regulation and use in the future. Beyond the scope of this thesis, but mentioned in Chapter Six, it is important to keep in mind the limitations involved in jurisdiction-specific regulations and the problems that may arise in practice. Because AI is often international in nature, and its development is currently at the behest of major corporate conglomerates, it is important that any decisions made in respect of jurisdictional issues are not too restrictive as to "shut out" NZ from the technology's future. However, it is crucial to always ensure that the domestic protections afforded to patients in NZ are respected and effective, to prevent exploitation of patient's data.

Closely related to the issue of internationality, but also pertinent within the domestic context, is the issue of scale involved in AI. An AI has the capacity to breach an individual's rights in a wide variety of ways in a very short period, and by extension can do this to a large number of people. A practical barrier to legal protections and their enforcement is the ability to manage the volume of prospective claims and disputes that may arise. The protections afforded by the Code are personal; an individual files a complaint, and this individual complaint is investigated by the Commissioner. Similarly, common law actions (as demonstrated in the discussion of negligence) are typically pursued on an individual basis, for harms incurred by an individual. However, AI represents a future in which the breach of a right to privacy is violated simultaneously between thousands of patients. Maintaining a system in which individual courses of action is the method of recourse is unlikely to be effective in terms of both redress and efficiency.

Finally, when discussing AI systems that are adaptive, there is often a correlation with systems that operate as a "black box". Although this is not the case, as discussed in respect of Kroll and others, there is significant work being done in verifiable or accountable AI systems. In instances in which AI systems do operate as a black-box, this will pose substantial problems. Being able to identify the origins of decisions or the processes involved is crucial to most of the actions discussed throughout this thesis and the inability to properly identify these things within most AI will cause considerable evidentiary issues. While "explainable" AI are possible, and there is substantial research into different models of doing so, how the law chooses to approach this in terms of regulation will need to be considered.

All these issues, while discussed in this thesis in the context of healthcare, are also relevant in other contexts as well. Situations in which an AI violates a patient's rights in healthcare may also engage with a wide variety of areas of the law, and the interaction of (and prioritisation of) different interests in these situations will need to be considered.

### 9.2.2 Issues associated with specific rights in Chapters Five to Eight

#### A) Discrimination

The capabilities of adaptive AI are such that it is extremely likely that connections are being made, and conclusions are being drawn, that distinguish based on characteristics that are protected under non-discrimination laws. Attempts to mitigate this will likely dampen the benefit of the AI's inclusion in general, and the understanding of AI necessary to prevent this occurrence is likely insufficient.

The tests for discrimination, under NZBORA, can be applied without issue due to NZ's focus on output, as opposed to a specific mental element. Because of this, patients maintain an effective means of redress in situations in which they have been discriminated; an AI's involvement does not complicate their available action. It will likely be difficult to identify if discrimination has occurred though due to the requirements of evidence and causation. The opaque nature of some AI, as well as the difficulty in understanding how they reach their conclusions generally, means identifying when discrimination has occurred will be difficult.

The current formulation of discrimination can lead to harsh liability being applied onto parties. Those held responsible for the discrimination occurring (hospitals, doctors) are likely to have been unaware of the potential discrimination by the AI, and unlikely to possess the ability to have mitigated or avoided it (short of refusing to utilise the AI at all). While the application of the test in its current form is desirable for protecting patients, the potential cooling effects on the use of, and development of, AI need to be considered.

#### B) Privacy

AI is fundamentally at odds with privacy from the outset due to its desire to aggregate and manipulate the largest amounts of data possible. The current protections and limitations imposed on data sharing and use are ineffectual at mitigating the risks to patients, such as being identified or having their data shared between different parties. Additionally, the way

in which AI utilises data can potentially breach a patient's privacy as well, although this is something that could be mitigated during the informed consent process.

Privacy is one of the areas in which the jurisdictional difficulties of AI will be most relevant. Due to how data is obtained, as well as the nature of the companies most involved within the AI industry, the way in which privacy can be breached internationally, as well as the limitations of methods of redress, will be an area of concern in the coming years. Additionally, while not explicitly an issue, careful consideration should be given to the role of, and treatment of, Māori data within AI's development and use. NZ must ensure that any protections maintained or altered continue to fulfil the obligations set out in Te Tiriti o Waitangi (The Treaty of Waitangi).

The test associated with a breach of the right to privacy is able to be applied, and patients have an effective means of redress available in respect of compensation (not considering the issue discussed in [9.2.1] relating to volume). However, the ability to enforce breaches of privacy that occur at second or third instance, as a result of data sharing or manipulation, is highly unlikely. Doing so could lead to a cooling effect on the development and integration of AI systems. The standards expected in respect of patient privacy need to be carefully considered to balance their effectiveness in protecting the patient, but not being too restrictive on those parties associated with the AI.

C) Informed consent

Like the previous two rights, the way in which an AI functions makes the likelihood of it performing actions (or making decisions) being at odds with the consent provided by patients. Dynamic consent, the ability for one's consent to adapt and accommodate changes, will be an issue in situations involving AI. It is likely that a patient's consent will need to be formulated as broadly and purposively as possible to ensure that breaches do not occur when attempting to utilise AI. Additionally, like privacy, while there is effective redress for violation of a patients' consent at first instance, the use and manipulation of data at later stages is likely. These situations are unlikely to be able to be properly addressed for patients.

Many of the issues associated with informed consent can be mitigated through careful disclosure and communication of the risks and uncertainties involved. While medical professionals should exercise caution in the information they communicate, it is unlikely that informed consent will cause significant issues in practice. In situations involving autonomous

care, the process of disclosure and obtaining informed consent will need to be especially thorough, but again, will likely pose no significant issues. The test for informed consent can be effectively applied, and patients have an effective means of redress available through the Code.

D) Harm and compensation: Negligence and the Accident Compensation Act 2001

Finally, there are a number of issues associated with the application of negligence principles to situations involving AI. Significant judicial analysis will likely need to occur to determine questions about the scope of the duty, and who owes a duty of care in situations involving an AI system. This will be of particular importance for situations involving autonomous care, in which no human agent is readily involved. Similarly, what the standard of care is for the parties involved in situations involving AI, and how "fault" can be attributed in autonomous situations, requires further analysis. Like discrimination and consent, there are clear issues of causation in applying negligence principles, exacerbated by the opaque nature of many AI and its potentially "unforeseeable" outcomes.

Whether the issues associated with negligence are of serious import is unclear. This relies on a clearer picture of whether or not the ACA will cover the kinds of scenarios discussed within this thesis. While the conclusion drawn in Chapter Eight is that AC will be available to the scenarios tested, it was based on a somewhat stretched interpretation of the Act. Most notably, the requirement that treatment is provided "by or at the direction of" an RHP is problematic. It is possible a court would disagree with the interpretation taken in this thesis, and therefore the coverage provided by ACA is unavailable. Although this is an issue that is remediable through minor reform, and is not of substantial concern. If the ACA does not cover the situation involved, the difficulties associated with the requirements of breach and duty of care will likely mean a situation warranting exemplary damages will be especially difficult to establish, effectively leaving a patient without remedy.

Below, a summary of the conclusions reached in applying the three scenarios has been provided. To re-iterate, the three scenarios are:

1. An individual goes to a doctor, who utilises an AI tool in their treatment ("SN1")
2. An individual is treated only by a machine ("SN2")
3. An individual uses an application or online AI tool as a substitute for visiting a doctor in person. ("SN3")

**Table I: A summary of the conclusions reached in each chapter on the application of the three scenarios:**

| Chapter | Right | Does the nature or involvement of an AI likely lead to a breach of the right afforded? | | | Are the obligations imposed on healthcare sufficient to reduce the likely harm? | Is the test or rule associated with the rule able to be applied effectively? | | | Does this application lead to an outcome which is undesirable? |
|---|---|---|---|---|---|---|---|---|---|
| | | **SN1** | **SN2** | **SN3** | | **SN1** | **SN2** | **SN3** | |
| 5 | Non-discrimination | Situational, but likely | Situational, but likely | Situational, but likely | Unlikely | Yes | Yes | Yes | Yes |
| 6 | Privacy | Likely | Likely | Likely | No | Yes | Yes | Yes | Yes |
| 7 | Informed Consent | Likely | Likely | Likely | Yes | Yes | Yes | Yes | No |
| 8 | Accident Compensation | Situational | Situational | Situational | N/A | Likely | Likely | Likely | No |
| | Negligence | Situational | Situational | Situational | Unclear | Unlikely | Unlikely | Unlikely | No |
| | Exemplary Damages | Situational | Situational | Situational | Unclear | Unlikely | Unlikely | Unlikely | No |

## 9.3 Select issues with reform and proposals

This section will discuss some of the issues associated with prospective reforms, and some areas in which questions arise when considering common avenues of discussion of AI systems and their legal impact. Most notably, this section will discuss the idea of making AI independently liable for the harms they cause, a common occurrence in discussions of AI in both academic and mainstream contexts. These selected issues are not intended to represent the most supported reforms, or to provide a comprehensive overview of issues involved. Instead, these are intended to highlight issues with a select few reforms, to highlight the difficulty that AI presents. It will also provide a select few suggestions made by other academics for some of the issues discussed within this thesis.

### 9.3.1 Embedded fairness

When considering the issues associated with how an AI "acts", particularly in respect of discrimination, it has become commonplace, particularly within the EU, to discuss embedding fairness. This involves designing AI in such a way that they are "fair" from the outset, and therefore not capable of being discriminatory or unjust.

Wachter and others in *Why Fairness Cannot be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI*[684] provide a robust overview of the arguments for embedding fairness, as well as a systematic critique of its impracticalities. Their first conclusion is that non-discrimination laws (within the European context) are insufficient for the coming technological developments of AI.[685] In fact, the authors go so far as to suggest that not only is it not possible to create "fair AI" in relation to non-discrimination, but other more tailored guidelines will be necessary to regulate both the technologies themselves, and their development. Their argument, built around the General Data Protection Regulation (GDPR) largely focuses on the fact that concepts of non-discrimination and other "social" wrongs are inherently human; their formulation, interpretation, and boundaries are based on

[684] Wachter and others, above n 267
[685] At 6

human characteristics and understandings.[686] The limitation here is that AI do not emulate these processes in turn, even when intended, or designed, to do this. The inherent "person" within the situation is lost, and thus the associated protections logically fails. Creating an AI system that can appropriately apply the data it relies on, to each contextual situation, to ensure that it is fair is an almost impossible technical task.

It is the view of Wachter and others that automating fairness is not possible, simply due to the differences in how discrimination occurs. In the context of the EU's GDPR, it was said:

> Due to the disparate nature of algorithmic and human discrimination, the EU's current requirements are too contextual, reliant on intuition, and open to judicial interpretation to be automated. Second, we show how the legal protection offered by non-discrimination law is challenged when AI, not humans, discriminate. Humans discriminate due to negative attitudes (e.g. stereotypes, prejudice) and unintentional biases (e.g. organisational practices or internalised stereotypes) which can act as a signal to victims that discrimination has occurred.[687]

While this study was focused on the EU's current interpretation of discrimination, it is relevant within the NZ context, particularly with its understanding that discrimination is a largely contextual, case-by-case analysis. It is difficult to automate a process that can emulate this heavily independent, and ironically, biased, interpretative cause. The study's position is that the highly contextual and personal nature of fairness and discrimination will be evaporated by the advent of AI. As they say, oftentimes discrimination claims are brought by those who *feel* they have been discriminated against intuitively; however, this intuition will no longer work when AI discriminates in new ways, on massive scales, and in ways people had never considered. While the GDPR has been argued in other contexts as being a benchmark for future privacy legislation internationally, within the AI context it appears at least to be lacking.

More broadly, calling back to John Buyer's argument discussed in Chapter Eight,[688] it is possible that conventional standards of ethics and concepts of fairness are insufficient for discussions of AI. The fundamental functionalities of an AI's "mind" and an organic mind

---

[686] At 6-8
[687] At 8
[688] Mentioned in [8.5]

are simply too disparate for proper analogous application or reasoning. Further complications arise in that difference does not always equate to inequality, as has been discussed already. Again, biological differences between races or genders can affect the efficacy of different pharmacological compounds, requiring a difference in caution and application. Given that there is then an epistemic uncertainty between the relationship of protected characteristics and health outcomes, particularly in new frontiers of technology, the use of traditional ethical fairness models (particularly in relation to algorithms) is problematic. It creates both empirical challenges, but also issues of application and even understanding, for those enforcing these mechanisms later.

A final concern is the disconnect between a patient's care, and the predictions associated with whether that care is fair. For example, a patient of a particular race is given differential treatment to a patient of a different race, and an AI model predicts – corrected for this fairness change – how well that the first patient will respond to treatment. What happens when that patient does not follow that predicted response, particularly when the prediction was based on a new, previously unused avenue of care? In this case, the model's idealised attempts at fairness are incompatible with the non-ideal, real world in which it is operating. It also may have made an ultimately ineffective recommendation, based on this perceived ideal of fairness, which in turn further harms the patient (or at least causes a disruption in care otherwise). The result of this is that the perceived new ethical "fair model" of AI is not actual benefitting the patient, but instead is camouflaging existent health inequalities under the veil of new fairness. In this case, the empirical evidence model of fairness falters, so too does traditional ethical models of balanced fairness.

Closely related to this idea of embedded fairness are the issues associated with corrections. What is sometimes referred to as a feedback loop can occur when parties attempt to rectify wrongful outcomes that have been instilled in an AI system. The idea being that through their attempts to "correct" the AI, the developers instead cause additional issues or cause greater breaches of a patient's rights. As an example:

> An AI system is designed by within a corporate environment that is primarily made up of Caucasian, educated male programmers. Amongst this group are a number of individuals, in positions of authority, known to have racist ideations towards African Americans. Consequently, the developed AI has these biases inherently built in,

through implicit manipulation of the datasets used, and methods of analysis. In a later attempt to remove this bias, the health authority engages another corporate body to adjust the dataset, drawing from more data, from an increasingly varied pool of sources, in order to feed enough competing data, from diverse enough sources, that the system corrects its own bias.

An issue becomes apparent; when is the end point of this process identified? Endlessly funnelling data into a system runs the risk of producing unknown, unintended, or potentially even more harmful, outcomes in the short-term as the data is improperly curated, and the system attempts to analyse it. Most notably, privacy protections may be circumvented or relaxed during an attempt to correct racial, or other forms of discriminatory biases within an AI system. This illustrates a conflict within the management of AI, and the facilitation of existing rights. The tension that arises is no longer between AI and conventional legal doctrine, but pressure between the rights protected themselves; a form of hierarchy becomes implied when concessions made to one right, in favour of another, begin to become necessary. This fear then encourages continued sourcing and implementation of data which can hinder processes designed to ensure privacy, because individual privacy and anonymisation could hinder the ability to gather enough useable data to overcome the bias. As such, the very real potential for bias can encourage design that does not give privacy sufficient due because privacy could hinder the ability to gather enough data to reduce bias, creating a self-defeating cycle of AI implementation.

### 9.3.2 The liability of AI

In much of the literature written in this area, there is an inevitable shift towards discussing the liability of AI themselves; not the liability of those interacting with or utilising the AI, but instead how to hold the AI liable independently. I would argue that this is both an undesirable, and a wholly impractical, avenue of inquiry. However, for completeness, the clearest issue with the idea – it's practicality – as well as one example where NZ has perhaps implemented a similar idea, will be discussed briefly.

The primary benefit to holding an AI independently liable is to simplify the inquiries into liability that have been discussed in the previous chapters, notably in Chapter Eight.

Allowing an AI to be held personally liable, it would narrow the potentially liable parties and largely mitigate the issues discussed in respect of the duty of care. In doing so, potential litigants would have greater clarity on who/what they could seek damages from. The parties most likely to no longer owe a duty of care, if the AI itself was liable, are those who were responsible for its functionality; the designers, developers, or manufacturers. It would follow logically that if the AI itself is liable it then assumes responsibility for its own risk and actions. This would, of course, be qualified by what status of legal personhood is bestowed on the AI; children for example are not independently liable in full, and some obligations still rest on their guardians.

If an AI itself is held liable, there is an obvious practical burden to overcome; how can an AI right the harm it has caused? Conventionally, a human or organisation found to have acted negligently can be ordered to pay financial compensation to restore the plaintiff to the position they would have been in, however as it stands, an AI is not legally capable of possessing property.[689] Without property, an AI lacks the ability to offer proper compensation toward a harmed patient. It could be argued that in situations where an AI is liable, its "host" organisation would also be vicariously liable. In this scenario the compensation would come from this organisation, i.e. the hospital, and the AI's liability would be purely for doctrinal "tidiness" by avoiding some of the issues outlined in Chapter Eight. This simply leads to more questions, such as how the "relationship akin to employment" test would apply to a non-sentient machine, as well as whether this is a better solution than simply finding the hospital principally liable in some way instead. At its outset, the idea of holding an AI liable seems largely a matter of speculation, with little practical value.

Beyond the impracticality of an AI remedying the wrongs caused to a patient, the way in which an AI is "controlled" or disciplined provides similar problems. If an AI were to be liable, it holds to reason that it should also be "punished" within the professional system it operates. Again, this runs affront to not only practical but also logical difficulties; for a punishment of this kind to have proper effect, the party punished needs to understand it is being punished. The value of exemplary damages, or some form of punishment (or "righting

[689] Although this is another area of debate occurring right now, particularly within the area of intellectual property law. See Rafael Dean Brown, "Property ownership and the legal personhood of artificial intelligence" (2020) Vol.30 Information & Communications Technology Law, for a discussion on this debate.

the wrong") arises because the defendant is aware of their conduct, the punishment, and in theory learns from it.[690] However, this assumes a great deal of awareness and cognitive ability on the part of an AI. In the event of non-human legal persons, such as corporations, there are at least human agents within the body corporate that understand the consequences and gravity of the outcome. While some forms of AI in the future may be able to adequately perceive and understand consequences, the breadth of what constitutes "AI" means that there would need to be an intricate, form-specific, arrangement for both legal status and appropriate remedies for different forms of AI. Non-monetary remedies, such as an injunction or declaratory order would in theory be possible, although it would require the actions of another party to enforce it in practice. And once again, the AI itself is not going to understand or "learn" from this in any way.

On a more logical basis, John Buyers argues that when holding a party accountable for their actions that caused harm, society is conceptually deeming this action to be "wrong."[691] In some way, their behaviour is being condemned, or at least reprehended, and correction is required. He argues that even though an AI might produce an undesirable result, or even harm a person, that does not mean the decision made was wrong, but instead an example of what is known as an "edge case."[692] Because the information from which an AI operates is information provided; they are developed and learn in a controlled environment with information only attainable by provision. As a result, any decisions determined to be wrong are based only on information provided, and therefore cannot be considered "wrong" for an AI. If a doctor was trained on, and only presented half of the information necessary, but was still properly registered and allowed to perform surgeries, the fault could hardly lie with the doctor. They were operating only with what was available, expected, and approved of them. In this case, it does not necessarily follow that it is appropriate to hold an AI accountable at all, simply because it is only acting in the way that it was designed to do so.

For the purposes of comparison, there is some precedent in NZ for a non-human (or corporate) entity being given legal status. In a similar way to how children are not fully liable, an AI could be given legal status and thus be liable for its actions, but have its burdens born by another party. A likely candidate would be the hospital itself as discussed, which

---

[690] Buyers, above n 642**,** at 25
[691] At 23
[692] At 25

would distribute the burden onto shareholders or the Ministry of Health. A similar mechanism has been employed in recognition of Tikanga Māori concepts in New Zealand through the Te Awa Tupua (Whananui River Claims Settlement) Act 2017 which grants Te Awa Tupua, a river, its own legal status. In this situation, the river itself cannot be expected to represent itself or fulfil remedies but a panel is established under the Act which acts as its agents. A comparable mechanism may be desirable for either larger AI systems that provide a wide-array of services, or all systems under a umbrella of processes in which the corporation responsible for them act as their agents. In the case of the river, the intention of this change was to protect the river; granting its legal personhood was so the river could "defend" itself against environmental harms, and other actions, that conflicted with Tikanga Māori. Following this logic, a panel of agents could be appointed to defend an AI if its actions lead to harm. While those who have suffered because of this loss may understandably wish to bring actions against the hospital (or in theory the AI) to prevent its use, there may be misunderstandings or wider interests at play that need to be represented. Conversely, if an AI is legitimately at fault, this panel or chosen party can be responsible for any necessary compensation, fulfilment of duties such as mandatory auditing or repairs, or other remedies that result from an action.

#### 9.3.2.1 "The reasonable machine"?

If an AI were to be held liable, what standard would it be held to? The current formulation for breach of duty is the "reasonable person"; which person is an AI being compared to? Gerstner, when discussing whether a software vendor owes a duty of care to a customer (as a matter of product liability), suggested there might be some difficult in choosing what standard is owed and proposed two options.[693] Traditionally, an action against a human would require that this vendor acted as the "reasonable software vendor", however Gerstner argues that for AI "expert systems"[694] the standard of care should be that of an expert, or at least professional, in that AI skillset.[695] Gerstner extended this to suggest that an individual operating alongside or in conjunction with an AI might be held to the same standard as an

[693] Gerstner, above n 652, at 13
[694] At 14-15
[695] At 18

expert who would normally perform the task themselves.[696] Logic then follows that an AI system capable of performing a task would be held to the same standard as its human equivalent. However, problematically, one of the main appeals of AI in medicine is their capability of performing tasks that their human counterparts are incapable of doing – for example precision surgeries and far more extensive diagnostics. This would then mean that the AI is being held to the standard of a reasonable precision surgeon, who does not exist.

Alternatively, the AI could be held to the standard of the "reasonable AI" within its position. The immediate problem here however is that AI of a particular role or quality would vary immensely between manufacturer and usage. Whilst doctors are trained at a limited number of accredited institutions, on set curriculum to ensure their knowledge and skills are up to par, the variety and economic competition associated with large entities means similar homogeneity does not exist within AI technology. An AI system that performs a task made by one business might function completely differently, with a different base dataset and different protections, than a machine performing a similar task for a similar purpose from another business. Working around this would require technologies used to work with specific curated datasets, or only be acquired by a singular tendering business to ensure consistency. However, restricting the implementation of AI in such a way serves primarily to disrupt the development, and effective improvement, of AI systems ultimately hindering its benefit in practice.

### 9.3.3 New standards of care

If NZ does not seek to make AI independently liable (and as a result in so far as human parties would still be potentially liable) this section will discuss potential changes to the standard of care (for human parties) generally. The standard of the "reasonable person" has been subject to some criticism over the years,[697] so the potential of a new standard is worth mentioning. One of the principal criticisms is that the apparent objective standard is a mask

[696] At 19

[697] John Gardner, "The many faces of the reasonable person" (2015) 131 Law Quarterly Review, contains a discussion of some critiques, particularly from a feminist perspective. Also, Mayo Moran, "Rethinking the Reasonable Person: An Egalitarian Reconstruction of the Objective Standard" (Oxford Scholarship Press: London, 2003) explores a robust analysis of why this test has come under fire.

for bias of different groups, through whichever judge is presiding. While the standard is claimed to be "objective" it is still determined on a case-by-case basis, by whichever judge is presiding.[698] This section will discuss an alternative that may arise because of the use of AI.

Fiser has argued the possibility of a shift from the reasonable care standard to a "what is best available" standard instead.[699] This would mean that if AI is available ,that is more effective than human diagnosis, the choice to not use AI would inherently be negligent.[700] "Best available technology" (BAT) standards are common in other areas of law already, such as pollution and environmental control regulations.[701] This position has seen growing strength within the US, with proponents like Andrew D. Selbst arguing in favour of modifying negligence laws to new emergent technologies to require that provenly superior technologies become the default, when implemented and regulated.[702] A problem with this standard in practice however is in determining when a technology or process becomes the BAT; on implementation an AI is likely to require to undergo "training", and those utilising or interacting with it will likely be undertaking extensive training and upskilling. Determining at what point this technology shifts from innovative or experimental, to the BAT and thus required to be utilised, will not only be highly circumstantial but highly burdensome to expect doctors to identify.

Fiser also presented the possibility that if a reasonable person standard were to remain, it could result in a lowering of the standard over time, if doctors deskill due to the implementation of AI. This deskilling would happen as a result of different tasks being shifted onto AI, and thus the daily responsibilities of doctors are lessened.[703] Another way to think of this, without perhaps the negative connotation associated with the word "deskilling", would be that the profession of the doctor changes. While radiologists for example are required to understand how to interpret scans and patient information themselves, radiologists of tomorrow may be required to understand how to interpret AI decision-making. This would not require them to verify the decision itself, but instead to verify that the decision is reliable.

---

[698] Moran, above n 687, at 5-6
[699] Harvey Fiser, "Automating Medicine" (Neuroscience and Society Conference, North Sydney Harborview Hotel, Sydney, August 2018) under heading "What are the legal consequences?"
[700] Fiser, above n 689
[701] These are common within European Union directives, such as Directive 96/61/EC.
[702] Andrew D. Selbst, "Negligence and AI's Human Users" No. 20-01 (2018) Public Law & Legal Theory Research Paper (unpublished)
[703] Fiser, above n 689

While this is not necessarily a reduction in skill, it is a realigning of it. If the doctors themselves have fewer responsibilities, they are likely to experience a lowering of what a "reasonable doctor" can be expected to do. This is not a problem in and of itself but was discussed by Fiser as largely an interpretative possibility where hospitals of different technology development resulted in different standards being applied within the same profession. Both possibilities are dependent on seeing how AI continues to develop and be utilised.

#### 9.3.3.1 A policy consideration: The reduction of doctor liability

In light of the discussion of new standards, it is worth noting also the idea that doctor liability (as in the degree to which they are held responsible) being reduced is beneficial. Conventional thinking is that medical malpractice rules, such as medical negligence, registration, and regulations like the Code protect patients – or at least enhance their care. However, if this were not necessarily true, then it could be arguable that in some instances, loosening restrictions that doctors are subject to may also be a viable option. To phrase it another way; instead of trying to find ways to fairly hold doctors accountable, it could be decided instead that the law is not concerned with holding doctor's liable in as many situations as they are, or could be. Instead, the law finds new mechanisms – for example aimed at holding hospitals themselves responsible for all acts, and allowing doctor's greater freedom to practice. On a practical level, removing tortious negligence actions for doctors would likely be replaced with higher standards of certification, as well as harsher internal discipline. However, by allowing things to occur behind closed doors, doctors may provide more effective care without potential worries about patient's alleging their care was negligent.

There is little evidence currently that stronger medical malpractice regulation actually improves patient care.[704] In fact there is evidence to suggest that these regimes instead lead doctor's to practice "defensive medicine" – medicine that is tailored to covering as many bases as possible, to avoid possible liability after the fact.[705] Defensive medicine is often

[704] Christina A. Minami and others, "Association between state medical malpractice environment and postoperative outcomes in the United States," (2016) 224:3 Journal of the American College of Surgeons,
[705] At 15-16

written about within the context of the US healthcare system, due to its high expense and funding differences. However, there is also evidence that this practice both exists and has a notable effect on doctor conduct within the UK as well,[706] so it is not unreasonable to assume it also occurs within NZ. Northwestern University researchers found that post-operative care patients, in states with stronger malpractice protections, were 22% more likely to become septic, 9% more likely to develop pneumonia, and 15% more likely to suffer acute kidney failure.[707] The increase in malpractice claims by area was also directly correlated with significantly worse medical outcomes for that area, suggesting that these areas with more avenues for actions were not actually facilitating better care, simply more aggressive responses.

Alongside potentially helping doctors to act more freely and effectively, there is a cost benefit as well. Medical malpractice is expensive, both in the course of damages, personal losses, and the defensive medicine being practiced by doctors. Those doctors who considered themselves more risk adverse in the UK were found to order far more diagnostic tests and treatments than the patients illnesses would generally warrant.[708] In a Gallup 2010 poll of private-hospital doctors in the USA, 73% admitted that they practiced defensive medicine; they confirmed that they actively attempted to mitigate their liability, even in situations in which their conduct was unnecessary.[709] This tactic adds an estimated $210 billion in expense for the US healthcare system annually.[710] While a cost comparison between the US and NZ is not a fair comparison due to their immense disparities in regulation and healthcare coverage, it is still important to highlight that doctors conduct does come with a cost. AI can not only mitigate this cost in terms of efficiency and efficacy in practice, but also potentially by giving doctors an "out" in liability, to lessen their self-perceived need to protect oneself. This thesis is not necessarily advocating for this position, but only highlights it so that during a

[706] Osman Ortashi and others, "The practice of defensive medicine among hospital doctors in the United Kingdom" (2013)14 BMC Medical Ethics
[707] Minami, above n 694, at 20
[708] At 25
[709] M Sonal Sekhar and N Vyas, "Defensive Medicine: A Bane to Healthcare," (2013) Vol3:2 Ann Med Health Sci Res, at 6
[710] Institute of Medicine (US) Roundtable on Evidence-Based Medicine, ed. Pierre L Yong, "The Healthcare Imperative: Lowering Costs and Improving Outcomes" (Washington DC: National Academies Press, 2010) at 55

discussion of potential reforms, it is established that reducing liability broadly may also be beneficial.

## 9.4 Recommendations for NZ healthcare

This section will now provide a series of recommendations on how NZ should approach the implementation of AI in healthcare in the near future for AI. These recommendations are intended to be relatively "simple" and therefore easy to implement. This recognises the reality that there are still multiple unanswered or unresolved issues associated with both AI and its place within healthcare. Greater investigation is needed before there can be more robust or comprehensive reforms suggested.

### 9.4.1 Guidelines

Guidelines can take on two main forms: guidelines for those developing or implementing the technologies in question, or guidelines for those interacting with or utilising those technologies. Both are recommended because the issues associated with AI are unlikely to be managed by a unilateral approach; attempting to mitigate the issues associated with AI purely through developing them differently, or through requiring doctors to take extra precautions, will likely result in an ineffectual response. In recent years, NZ has developed a robust range of guidelines and principles for the use, development, and integration of AI in a wide variety of contexts. Examples include:

1) Principles for the Safe and Effective Use of Data and Analytics (Privacy Commissioner and Government Chief Data Steward, 2018);
2) Government Use of Artificial Intelligence in New Zealand (New Zealand Law Foundation and Otago University, 2019);
3) Trustworthy AI in Aotearoa – AI Principles (AI Forum New Zealand, 2020);
4) Open Government Partnership, an international agreement to increase transparency Data Protection and Use Policy (Social Wellbeing Agency, 2020);
5) Privacy, Human Rights and Ethics Framework (Ministry of Social Development); and
6) The Algorithm Charter for Aotearoa New Zealand (Stats NZ, 2020).

These provide a comprehensive and effective basis on which to base future developments and reforms, and thus this thesis will not seek to provide another set of guidelines. Instead, it is noted that the principles most important for healthcare, commonly occurring within these different sources, are:

1. Transparency;
2. Dignity of people;
3. Privacy and human rights; and
4. Ensuring human oversight.[711]

In respect of practical guidelines, the MoH has already begun releasing its series of advice and information on emerging health technologies.[712] An area in which guidelines are perhaps most important is in the process of obtaining informed consent. Effective communication of both the realistic benefits and capabilities of, as well as the risks and uncertainties around, an AI is crucial to ensure that informed consent is properly obtained and respected. While it was concluded in Chapter Six that the application of the right to informed consent is not especially interfered with by the introduction of an AI system, it is still desirable to mitigate potential courses of action by exercising caution. While not necessarily legally necessary, it is recommended that health professionals are provided with an AI-specific set of information which likely needs to be communicated, as well as providing some introductory guidance on the likely common questions which doctors are themselves unlikely to know the answers to.

Health professionals in situations involving an AI system should be advised to communicate:

1. That an AI is involved in the process and to what extent;
2. The value or necessity of the AI's involvement, as well as its purported benefits;
3. What mechanisms of oversight or human control exist over the AI system involved;

---

[711] This is most important for ensuring the continued trust of the system as it develops. Irrespective of the abilities and successes associated with AI technology, maintaining the assurance that any approved or utilised technologies are under close observation will be crucial in their continued acceptance.

[712] For example, the Ministry of Health's "Emerging Health Technology Advice" series and "Introductory Guidance to Emerging Health Technology" series are both available at https://www.health.govt.nz/our-work/digital-health/vision-health-technology/emerging-health-technology-advice-and-guidance

4. The unique risks posed by an AI system, which otherwise may not arise from the treatment being delivered conventionally (such as dynamic use of data); and
5. The ways in which the patient's data may be utilised after the fact.

While these points are largely required under the test of informed consent, they are stressed as important within Chapter Six, as well as the results of the conducted survey, and in some instances may not be perceived as crucial otherwise.

### 9.4.2 Legislative changes

In this section, a number of regulatory requirements or legislative changes will be outlined. These fall into two categories: (1) changes made to the protections and processes involved in healthcare itself, to allow for the effective use of the mechanisms available, and (2) suggestions for regulation of how AI is developed and implemented, to pre-emptively mitigate the risks posed before the situations arise.

#### A) Liability

It is recommended that a statutory determination of liability is established for situations involving the use of an AI. As discussed in several chapters, traditional methods of attributing liability could be seen as both unfair, unduly harsh, or simply impractical in situations in which the harm was brought about by an AI itself, or because of an AI's use. This statutory attribution should only apply in situations without clear responsibility of the human agent involved. Available situations might be limited to instances of:

1. Indirect discrimination, where discrimination occurred because of an AI's functionality, as opposed to the intention of a discriminatory human agent;
2. Breaches of a patient's privacy or informed consent, because of an AI's functionality, which might occur through:
    - Predictive medicine, in which an AI utilises data in a way beyond the consent of the patient, or determines and utilises something which the patient did not wish to be known; or
    - Re-identification, where the AI lead to a situation in which a patient who was previously anonymous was able to be identified, without deliberate intention by a human agent.

3. Situations in which a patients' rights were breached where no human agent was involved (i.e. during autonomous care).

In these situations, attempting to attribute liability by the conventional means is undesirable, and to clarify any prospective decisions should be determined from the outset. One way this could be phrased in statute is:

> In situations where liability for breach of any Act must be determined in which an AI was utilised by another party, or acted independently of another party, liability shall be attributed to the overseeing medical facility, except where –
>
> a) The actions of a legal person involved can be causally attributed to the wrong incurred; and
> b) A legal person is shown to intentionally, or recklessly, act in such a way as the wrong would occur

In this situation, the overseeing medical facility (most likely a hospital, or GP practice) would be responsible for implementing and utilising the AI system. This would also allow for situations which warrant exemplary damages, for example where a doctor deliberately utilised an AI known to have a racial bias, to attract liability under the conventional rules.

B) Accident compensation changes

Similar to the recommendation made under (A), it is recommended that the requirements for "treatment injury" under the ACA[713] are amended to accommodate the developing realities of the technology. This change should only be implemented in the event that autonomous treatment, without an available or responsible human agent, is permitted within NZ.[714]

Requiring that "treatment" be given "by or at the direction of" a registered health professional[715] results in situations where truly autonomous care may not be covered by the ACA. As discussed in Chapter Eight, it is likely that a broad, purposive interpretation of the

[713] Accident Compensation Act 2001, Section 32
[714] While it is argued that this is the natural trajectory of the technology in question, it is also possible for NZ to require that all situations involving AI maintain Human oversight. In doing so, NZ would perhaps mitigate some of the efficiency benefits of AI, but would ensure that the ACA could always apply to treatment situations as well as maintain public trust in the treatment processes provided.
[715] Section 32(1)(a)

provision would result in these situations being covered, especially in the short term where the use of independent AIs is being recommended or guided by human professionals. While this amendment may therefore be pre-emptive, it is still worth considering.

Healthcare provided within NZ is regulated by bodies such as MedSafe and the Ministry of Health, and therefore the forms of treatment provided are sanctioned. The requirement of treatment being administered by an RHP, a human, is arguably only necessary as a matter of current practical realities, to ensure a sufficiently narrow and precise definition within the ACA. Amending provision 32(1) to all sanctioned care would allow for autonomous AI situations to be covered as treatment, and would not expand the coverage to undesired "non-official" healthcare. A drafting of this section would look like:

> **32 Treatment injury**
>
> (1) Treatment injury means personal injury that is—
>
> (a) suffered by a person—
>
> (i) seeking treatment through any service or provider approved and accredited by the Ministry of Health;
>
> or
>
> (ii) receiving treatment through any service or provider approved and accredited by the Ministry of Health

This would eliminate the, albeit narrow, restriction that currently exists for autonomous care, but still only allow care that the state considers professional or official to be covered by the ACA. It makes little sense for treatment that is in theory available within a hospital to not be covered by the ACA in the potential smart-health future. This amendment is a relatively simple adjustment to make, with no readily identifiable consequences beyond the intended scope.

C) Māori involvement in development and management

As ideologies and biases have changed over time, one of the great positives has been the modern push towards greater diversity in voices, and inclusion of viewpoints in the legal and medical process. This diversity is also crucial in attempting to combat discrimination in AI systems. Including different voices in the development of AI, and the management of their datasets, is an important step in ensuring that explicit biases are excluded, and implicit biases

are at least mitigated in their intensity.[716] Gender, ethnic and socio-economic diversity are important for a variety of reasons. Careers in AI are both well-paid and an area of rapid growth. Dominance by a group (i.e. white males) will continue to occur if not rapidly confronted in the coming years. Biases embedded in data may be exacerbated rapidly, and with the AI revolution underway it is possible this accelerates simply too fast to correct. As CognitionX stated to the HoL:

> "…one of the reliable ways we know we can mitigate [the problem of bias and discrimination] is to have more diverse development teams in terms of specialisms, identities and experience."[717]

In addition to its benefit of reducing potential discrimination, inclusion can help ensure that NZ's obligations under the Treaty are fulfilled. Rules should be established around the development and training of AI systems to ensure that the obligations of the Treaty towards Māori are met from the outset. Under the newly reformed healthcare system, the use of an AI system without healthcare should be subject to the approval of the Māori Health Authority, secondary to any requisite approval by other regulatory agencies (such as by MedSafe).

Some things to consider requiring include:

1. That AI systems for use within NZ healthcare are developed in consultation with Māori health experts, to ensure they properly respect and represent Māori patients;
2. That such systems are appropriately trained on data representative of the health of the NZ population;
3. That any data used for training is appropriately curated to minimise biases against Māori that may result from insufficient data, inadequate research or understanding of Māori health issues, and any other details which may disproportionately affect Māori; and
4. That the beliefs and opinions of different iwi are appropriately respected in the functionality and utilisation of the AI in question.

[716] House of Lords Select Committee on Artificial Intelligence, above n 56, at [155]
[717] at [173]

Determining when these components are achieved, or who qualifies as a "Māori health expert" is something that needs to be determined carefully. Such controls need to be designed to not be too restrictive of the development of AI but to be restrictive enough to ensure that the benefits of the technology are not lost for Māori in NZ. How such controls or processes are designed warrants a more focused inquiry later and is included in the recommended future research in [10.5].

## 9.5 Concluding thoughts on reform and AI

The known difficulties associated with AI, as well as the uncertainties surrounding its capabilities and continued development, mean that any discussion of reforms is likely to be inadequate or incomplete. This chapter set out only to consolidate the positions concluded throughout this thesis in respect of issues with applying the current rules to situations involving AI, and then highlight a select few issues that arise when considering how to mitigate or remedy these. While perhaps appearing defeatist in approach, I do acknowledge that it is unlikely any reform will "get it right the first time", and that identifying these seemingly endless issues is vital to the continued conversation around the technology. While healthcare and its associated protections cannot be said to be perfect, NZ has a long history of incremental reforms aimed at maintaining the inclusive, responsive and effective system that currently operates.

The recommendations made in [9.4] are intended to provide the most immediate and readily available changes that will facilitate a smoother transition to the AI-infused healthcare of tomorrow. The changes recommended under [9.4.2] (A) in respect of liability are arguably the most immediate in terms of legal consequence, as the early adoption of AI is likely to incur situations in which the issues identified in Chapters 5 to 8 are readily apparent. It is also stressed that while this thesis was largely concerned with the operation of responsive rules of law, that the changes recommended in under [9.4.1] and in (B) in respect of development guidelines and obligations, are perhaps even more important to consider. The potential scale of breaches of patient rights by AI is an area of special concern, where even if the legal tests involved are more readily applicable once amended, attempting to mitigate the wrongs occurrence from the outset is far more desirable.

# Chapter 10: Conclusions

## 10.1 Introduction

The aim of this thesis was to identify the impact of the introduction of AI technologies on NZ healthcare, with the primary focus on the rights patients are afforded during care. The driving concern is that the existing interpretations and applications of these rights would not be able to be applied to situations involving AI technologies. In this event, patients who suffer wrongs such as discriminatory or negligent treatment, would be unable to seek effective redress. The broader consequence would be that public trust in the system at large would be hindered due to a lack of certainty around the protections afforded to patients. This trust is reliant on the assurance that patients who engage with these new avenues of treatment are not in some way worse off than if engaging with conventional care. In turn, this decay of trust would affect both public health, as well as negate the purported benefits of the technology's introduction.

The presumption throughout this thesis is that NZ will desire, seek and attempt to develop into a "smart" healthcare system. This is based on NZ's already rapidly growing technological industry as well as the country's history of modern, innovative care. With this presumption in mind, the aim was to outline the areas in which the law would be required to adjust so that two things could be ensured: (1) that the protections afforded to patients are not lost in the wake of innovation, unless the benefit for doing so was a worthwhile concession, and (2) that the attempts to maintain established protections is not so restrictive as to hinder the development of the technologies.

The primary conclusion of this thesis is simple: the issues associated with the introduction of AI to healthcare that NZ must grapple with are not unique to its system. The incompatibilities or difficulties associated with privacy and consent for example are such that they will affect any system, and NZ is neither ill prepared nor incapable of addressing these. However, NZ is in the unique position thanks to the Accident Compensation scheme to mitigate some of the potential dangers of early adoption (such as harm from misuse or

inadequate understanding of the technology) that other jurisdictions will be forced to resolve through conventional negligence mechanisms. NZ's purposive and rights-based approach to AI is appropriately adaptable to the challenge at hand, and the country has sufficient time before widespread adoption of AI technology to ensure effective management of the issues identified within this thesis.

This thesis illustrated that NZ's current legal framework is sufficient to manage the introduction of AI systems into healthcare. The flexibility and adaptability of the system is appropriately designed to manage a broad range of scenarios, even those without human involvement. For example, AC's no-fault compensation scheme is not concerned generally with the cause as much as the scenario in which the harm arose. This allows for great flexibility, with perhaps one minor change for clarity, to any smart health scenarios that may arise. Similarly, discrimination lacking any requirement of intent or *mens rea* means that it can be applied to situations without direct human oversight, and can in theory even be applied to autonomous non-human systems. There is still significant difficulty surrounding causation and matters of intent or recklessness when a human agent is still involved. However much of these are reliant on further contextual information, such as whether the AI involved is a BP or not, to properly determine. As stated, importantly few of these problems are unique to AI, and thus a rich array of jurisprudence is available in many cases to provide guidance on possible solutions.

A repeated conclusion throughout this thesis is that both scale and jurisdictional issues will arise in respect of AI systems. How the legal system, largely designed around resolving issues on smaller inter-personal, or at least inter-corporate, scales will adjust to situations involving potentially thousands or hundreds of thousands of people is unclear. The example of AC may be one in which future legal developments are based in an attempt to mitigate the impact of this. Similarly, how the international regulation, cooperation or coordination of AI systems continues to develop will need to be monitored carefully. While solutions such as Taiuru's localised data-banking, discussed in Chapter 6, are feasible, they result in further difficulties of cost and technical expertise. The recommendations made in the previous chapter were largely concerned with the short-term introduction of AI systems, and mitigating some of the minor issues identified in the body chapters. The future research listed later is where these larger issues of scale and jurisdictional boundaries are discussed, as they will likely require far greater reaching change and control.

The following sections provide greater detail on the specific chapter-based conclusions reached throughout this thesis. Following this, a brief reminder of the recommendations made in Chapter Nine, in light of these conclusions, is provided. Finally, there will be a set of recommendations for future research that should be conducted, to either supplement or build on the issues raised within this thesis.

## 10.2 General conclusions about AI and healthcare: Chapters 2 to 4

Part A provided the necessary background and framing context for the thesis. This part served two purposes: (1) to define the technology in question and some of the broader issues associated with its interaction with the law, and (2) to illustrate the conceptual framework which underpins the healthcare system in NZ; namely it's flexibility and historical inclusiveness, to provide a basis on which to approach the analysis in later chapters.

**Chapter Two "Artificial Intelligence"** detailed the technology in question and the difficulty in specificity in both discussions, and regulation, of AI systems. What constitutes "AI" is both hard to define and not universally agreed upon, leading to discussions and analyses being highly contextual. Prospective regulation of introduced technologies, and the interpretations taken by tribunals or courts, must be carefully crafted to ensure that the consequences are limited to those technologies intended. Improper definition could lead to situations in which too many technologies are captured by a regulation, leading to unduly harsh applications on some technologies, and perhaps lax applications on others. Conversely, too narrow a definition runs the risk of creating gaps in the legal protections afforded to patients, and situations in which the law is slow to respond to the developing field of AI.

**Chapter Three "The Health System of New Zealand"** outlined the system in current operation within NZ, and its genesis. By highlighting the systems' history, a picture was given of a system that is both flexible to change and has strived to be inclusion of different people's and their needs. While this is of little legal consequence, it is important to show that the system itself does not aim to be restrictive or narrow in its applicable, and that broad purposive approaches which maintain this inclusive intention are both valuable and appropriate. This chapter also highlighted the sources of law which are relevant for the

succeeding discussions, most importantly the "Code of Patient's Rights" which served as the representative source of the rights and protections discussed later in the thesis. Also, the Accident Compensation Act 2001 and its role within healthcare was introduced, which serves as one of the principal advantages NZ has in this discussion of healthcare in AI.

**Chapter Four "The Code of Patient's Rights"** served three purposes: (1) to outline the underlying ideals and themes of the healthcare system, as inferred from the Code, (2) to provide the framework for the following chapters' focus and application, and (3) to provide guidance on the concerns and considerations of the public. This Chapter acted as the bridge between Part A's conceptual background and introductions, and the technical analysis conducted in Part B.

The dominant themes identified within the Code, which were applied within this thesis, were trust and respect. As mentioned already, trust is integral to ensure the effective operation of the healthcare system and its associated protections. Closely related, respect is an important component of ensuring public trust in the system; a system that respects and accommodates patients' individualities and dignity is more likely to be trusted and engaged with. It also provides a guidance on the ways in which concepts discussed in Part B, such as discrimination, privacy and informed consent, should be approached to maintain this respect of the patient.

While each chapter did not necessarily engage in depth with the right as it exists in the Code, using the Code as the framing device was important to show what is considered integral to the healthcare system. The rights afforded to patients within the document and the underlying principles which can be inferred from this, show both the priorities of the system and the areas in which the most careful analysis is necessary. The ten rights codified were distilled down to four main applicable concepts: the right not to be subject to discrimination; the right to have their privacy respected; the right to provide informed consent; and the right to not receive negligent treatment. These four rights were discussed in their respective chapters as a mixture of their Code representation as well as their representation within the common law. The reason for this was to provide a robust and comprehensive analysis that could be applied within a variety of contexts, both within and outside healthcare, as well as to illustrate the deeply embedded issues associated with AI and the concepts, tests and language utilised by these rights.

The survey conducted, outlined and analysed in Chapter 4, helped enforce the guiding direction established by showcasing what the public – those who are potentially are impacted by the introduction of AI in healthcare – thought. Whilst not definitive in itself, the responses and conclusions drawn from the survey helped to indicate a number of important ideas. These are:

1) Public understanding of AI is poor, and thus misunderstanding of their prospective issues or dangers is common.
2) There is public concern about the impact of AI on healthcare, which is currently unaddressed.
3) Prospective patients consider the "human connection" of healthcare to be a top priority, as opposed to the lauded efficiency and benefits AI may afford them; and
4) The use of patient data both in, and outside, of healthcare is a primary concern.

The survey also served to provide some guiding evidence and considerations in later chapters, namely in Chapters 6 and 7 when discussing the rights to privacy and informed consent. As these two rights largely turn on subjective considerations or understandings, the insight provided by the survey was effective for showing the potential react to the implementation of AI in healthcare.

## 10.3 The impact of AI on patient rights and mechanisms of redress: Chapters 5 to 8

As summarised in Chapter Nine, the impact of AI on healthcare is not evenly distributed, and some rights and concepts fair better than others. What is clear is that irrespective of whether the basic functionality of the rights is intact, the technological capabilities, and functional uncertainties of AI leads to situations in which they are both impractical to execute, as well as situations in which it perhaps undesirable to apply the rules as conventional. This section will now summarise the discussions in Chapters Five to Eight, and their conclusions. These summaries will reflect on the research questions posed in Chapter One at [1.3.1.2] and their answers.

**Chapter Five "Discrimination**": The formulation of discrimination, and its associated tests, under both the NZBORA and the Code are able to be applied to circumstances involving AI systems. Due to NZ's focus on outputs, as opposed to intention or mental state, discriminatory care can occur in situations with little or even no human involvement. There are potential issues with how liability is determined, and the harshness associated with this. Situations in which discrimination can be determined may result in individuals or parties being held liable for that discrimination when they were unable to properly prevent or account for its occurrence. This can result in a cooling effect on the technology's use, due to the output focused discrimination rules being perhaps too readily applicable. In this case, a patient treated by an AI is not at any disadvantage in terms of their protection and is afforded the same protections with little problems in the way of application.

**Chapter Six "Privacy":** this is an area in which considerable attention needs to be given in the coming years. Big Data's reliance on largescale data sets, and the sharing and manipulation of this data, means that the conventional principles and application of privacy are largely incompatible in practice. The technology itself also presents issues for some of the conventional mechanisms of protection, such as anonymisation of data, which means that different considerations or standards for how to achieve patient privacy will need to be considered. Privacy is one of the areas in which the interterritorial nature of AI and data is at its most problematic, something which was beyond the scope of this thesis for an in-depth analysis.

**Chapter Seven "Informed Consent":** this is an area in which first instance scenarios (patient interacting with their doctor or a hospital) prevent little issue. It was shown that caution needs to be exercised when doctors are determining the information they communicate or disclose to patients, and how they represent the technology itself when obtaining consent. This is of little consequence to the application of the right however, and will only impact the liability and expectations placed on the healthcare professionals themselves. However, the difficulties of applying the requirements of informed consent, and enforcing them, become apparent when considering that AI is a transformative and ongoing process. In these second-instance situations, the concerns of jurisdiction present in the privacy discussion are also of importance. For patients within NZ engaging with healthcare directly, there is little problem. The immediate situations in which a patient's consent is

necessary to obtain and respect will be unaffected by the introduction of AI, so long as care is taken when determining what needs to be disclosed and communicated.

**Chapter Eight "Harm and Compensation":** Accident compensation prevents much of the issues associated with negligence (summarised next) from being a concern in NZ. In the three scenarios discussed within this thesis, it is likely that the ACA would cover harms caused, at least in incidents likely to occur in the near future. The issues for AC will arise when AI begins to be utilised in a more autonomous, and independent, capacity where the requirement of treatment being "by" or "at the direction of" a registered health professional creates a complication. In these situations, it was concluded in Chapter 8 that coverage likely falls short of the situation, leaving patients reliant on common law negligence which was determined to be full of potential issues. The ability for the ACA to cover the situations, with only minor amendments necessary, means that NZ has a readily available mechanism putting it as a significant advantage to other jurisdictions.

Negligence, or care of an acceptable standard, is an area with a number of notable issues in application. How it is determined who owes a duty of care in situations involving AI is a difficult question; while not impossible to apply, it will likely result in early common law decisions needing substantial judicial analysis to determine. Similarly, there are several issues with both the breach of this duty and how it is determined, and the requirement of causation in negligence decisions (namely in respect of evidentiary and conceptual limitations). With the flexibility of the common law in respect of negligence, it is not concluded that negligence is "incompatible" with AI in healthcare, but there are some questions that remain. Chapter Eight highlighted however that this is almost entirely a set of issues which will rarely be of concern or relevance. Due to the application of Accident Compensation, most negligence actions will be barred by section 317 of the Act. The only situations in which negligence will continue to be relevant is in exemplary damages actions, but as shown, the issues associated with breach and the duty of care will make these difficult to apply as well.

A commonality across these four chapters was the issue of scale. Healthcare wrongs are typically a one-on-one occurrence, or in extreme instances like Green's medical experimentation, some hundreds of people. AI presents the very real likelihood that when individuals rights are found to have been violated, a considerable volume of people and

parties may be entitled to redress of some kind. While the tests and concepts have varying likelihood of being applicable, the more practical concern is how the system is able to adequately address situations of this scale. Similarly, while discussed in the context of privacy and patient data, the international nature of AI systems (both in terms of datasets, responsible parties, and connectivity) raises issues for the enforcement of any available legal protections.

## 10.4 Recommendations made: Chapter 9

In Part C, Chapter Nine first discussed some select issues associated with prospective approaches to reform. These approaches had been referenced in earlier chapters and represented different ways in which some of the issues identified could be managed or circumvented. As shown within Chapter Nine however, these approaches were often unlikely to be effective themselves or led to situations requiring considerable reform and reworkings of the law to implement. Considering this, it is my recommendation to adopt a "minimal interference" approach to reform. Changes made should be the minimum required to mitigate some of the most immediate or obviously remediable issues identified. In doing so, this allows more time for analysis and debate to occur to determine more effectual methods of reform for the issues which are either of a less immediate concern or require more substantial reforms.

The broadest applicable recommendation made was for the development of further guidelines and advice for medical professionals to be produced. Guidance in respect of the principled and ethical development and use of AI is important, and it was noted that NZ has already begun to do so comprehensively. It was recommended that specific AI guidelines are developed for the process of informed consent, detailing the things that medical professionals should disclose and communicate to patients to mitigate the potential issues identified within Chapter 6.

A statutory attribution of liability was proposed, in which situations involving AI that lacked clear human fault would result in liability being imposed on the overseeing institution or responsible body. This would replace the conventional tests for attributing liability under

each right, which were argued to present potentially cooling effects, undue harshness, or simply immense difficulty in applying.

It was recommended that the approval and sanctioning of the use of AI within healthcare should be subject to the approval of the Māori Health Authority. To obtain approval, an AI developer would require to illustrate that the system fulfilled a number of requirements to ensure the respect of, and effective care of, Māori patients within NZ. This recommendation was made in part to mitigate some of the issues discussed, such as in Chapter 5 in respect of discrimination, but also to ensure that NZ effectively fulfilled its obligations under Te Tiriti o Waitangi (The Treaty of Waitangi).

Finally, it was recommended that the requirement of a treatment injury being suffered when treatment is "by or at the direction of" a registered health professional, under section 32 of the Accident Compensation Act 2001, be amended. This proposed amendment expands the definition of a treatment injury to include any treatment that is received in a sanctioned or approved capacity, irrespective of who (or what) delivers the treatment. While this amendment is pre-emptive, as the situations which warrant it are unlikely to occur in the near future, it has little risk associated with it and is therefore recommended now.

## 10.5 Recommendations for future research

This thesis was, by necessity, selective in its coverage of issues relating to AI and healthcare. Some areas of the system and the potential issues associated were omitted, and others were only mentioned in passing to try maintain a focused scope. Additionally, some issues discussed were noted as being too extensive for sufficient analysis within the context they appeared and warranted more focused inquiries. To close the thesis, this section will make list areas in which future research is recommended to supplement this thesis, as well as to address some of the broader conceptual issues identified in greater depth.

In Chapter Two, the issues associated with defining and therefore appropriately regulating AI were discussed. Considerable research needs to be conducted into how AI is defined as well as how NZ should choose to define it within healthcare, and other contexts. The definition provided in this thesis was utilised purely to ensure that the resulting discussion

could be applied contextually, and should not be considered an authoritative definition. Different types of AI, employed in different contexts, will all result in different applications of the rights discussed and other legal protections. It is important that more work is done into this variety, and that any legal definition presented of AI is carefully qualified.

A common issue discussed throughout this thesis is the impact that the black-box problem (BBP) would have on legal tests, particularly in respect of causation. While "explainable" AI are possible, and were discussed in Chapter Two, there needs to be research into both the effectiveness of such systems, as well as any issues that may arise from requiring systems be developed in such a way. For example, as mentioned in Chapter Five when discussing COMPAS, the proprietary nature of software is something that needs to be considered in situations involving AI. Requiring AI to explain their processes may result in issues for competitive developments and trade secrets.

In Chapter Seven, it was briefly mentioned that AI will almost certainly result in jurisdictional conflicts and enforcement limitations, both within healthcare, and in a variety of other contexts such as international finance, diplomacy and business. How NZ's domestic law and international obligations respond to the involvement of AI presents room for significant research. One potential area of inquiry is into the benefits of, or potential creation of, an international governing agency of AI, akin to the WTO or WIPO in function. Doing so would assist in facilitating consistent interpretation and applications of the technologies, as well as ensuring that human rights, national jurisdictions, and international collaboration is maintained and championed.

This thesis made brief mention of the relevance of consumer protections under the Consumer Guarantees Act 1993 (CGA), Fair-Trading Act 1984 (FTA) and Sale of Goods Act 1908 (SOGA). There is room for future research into the application of, and limitations of, these consumer protections in respect of AI in contexts beyond healthcare. Additionally, the trade protections available to business or institutional bodies like hospitals under other aspects of trade regulation, such as the Fair-Trading Act 1984 (FTA) is another area of potential research.

Chapter Nine made recommendations on the involvement of Māori in the development and regulation of AI. There are potential avenues of research for other considerations of development standards and regulation, both within a legal context but also within a practical

context (such as, does MedSafe or another regulatory body have the technical expertise or capability to properly approve an AI system?).

Finally, it was noted in Chapter Six that AI presents significant difficulties for the concept of "anonymisation" of data. It was also noted that NZ has received some criticism for its comparative lack of clear anonymisation standards, or at least inadequacy compared to other jurisdictions. Future research into what standard of anonymisation, or alternative mechanisms of data privacy, are expected in NZ is recommended.

## 10.6 Closing comments

AI presents an exciting frontier for both technological and legal innovation in healthcare which NZ should take the opportunity to exploit. And while many of the issues discussed within this thesis may appear either overly cautious or pre-emptive, it is my view that the advances of the technology will almost certainly outpace those of the legal system and policy. By adopting a proactive and forward-thinking approach, NZ can ensure that the law is prepared to accommodate the changes to healthcare that are likely, and to ensure that the people and their dignity is not lost in the torrent of technological benefits, innovations and new possibilities. This is of particular importance in healthcare, where the consequences of being ill-prepared are often for people at their most vulnerable. Ensuring that the law can response promptly, and effectively, is crucial for maintaining trust in the healthcare system, and in turn, its effectiveness.

This thesis is intended to complement the discussion happening in other sectors and contexts within NZ to prepare the country for the technological future. The hope is that through analyses such as these NZ can embrace these new possibilities and innovate on the global stage. NZ's inclusive and adaptive healthcare system, and some of its specific mechanisms like accident compensation provide examples to emulate within other jurisdictions, and these will facilitate a transition in NZ to smart healthcare.

# Bibliography

## I. Cases

### New Zealand


*Abel v Brownlee* [2002] DCR 407

*ACC v Ambros* [2008] 1 NZLR 340

*Accident Compensation Corporation v Calver* [2021] NZCA 211; [2021] NZLR 721

*Accident Compensation Corporation v Shand* [2020] NZHC 2743, [2020] 3 NZLR 507

*Adlam v Accident Compensation Corporation* [2017] NZCA 457, [2018] 2 NZLR 102

*Allenby v H* [2012] NZSC 33, [2021] 3 NZLR 425

*Auckland District Health Board v W & W* [2012] NZHC 1563

*B v DGSW* [1995] 2 NZLR 134

*BNZ v NZ Guardian Trust* [1999] 1 NZLR 664

*Body Corporate No 207624 v North Shore City Council* [2012] NZSC 83, [2013] 2 NZLR 297

*Bottrill v A* [2001] 3 NZLR 622 (CA)

*Bottrill v A* [2003] 2 NZLR 721 (PC)

*C v Holland* [2012] NZHC 2155

*Caparo Industries plc v Dickman* [1990] 2 AC 605 (HL)

*Child Poverty Action Group Inc v AG* (2011) NZCA 397

*Coburn v Human Rights Commission* [1994] 3 NZLR 323

*Couch v Attorney-General (No 2) (on appeal from Hobson v Attorney-General)* [2010] NZSC 27, [2010] 3 NZLR 148

*Cumberland v Accident Compensation Corporation* [2014] 2 NZLR 373

*Daniels v Thompson* [1888] 3 NZLR 22 (CA)

*David Kimber TV & Sound Ltd v Kaiapoi BC* [1988] 1 NZLR 376

*Donselaar v Donselaar* [1977] Wellington, A 454/76

*Donselaar v Donselaar* [1982] 1 NZLR 97 (CA)

*EM v Accident Compensation Corporation* [2016] NZHC 2535

*Green v Matheson* [1989] 3 NZLR 564 (CA)

*Hamilton v Papakura District Council* [2000] 1 NZLR 265 (CA)

*Hawthorne v Cox* (2008) 1 NZLR 409

*Hosking v Runting* [2005] 2 NZLR 1

*Human Rights Commission v Eric Sides Motor Co* [1981] 2 NZAR 447

*J v ACC* [2017] NZCA 441, [2017] 3 NZLR 804

*L v Robinson* [2000] 3 NZLR 499

*Lyons v Nicholls* [1958] NZLR 409 (CA)

*McAlister v Air New Zealand* [2009] NZSC 78, [2010] 1 NZLR 153, (2009) 8 HRNZ 801

*McAlister v Air New Zealand Ltd* [2009] NZSC 78

*McLaren Maycroft & Co. v Fletcher Development Co. Ltd* [1973] 2 NZLR 100

*McLaren Transport Ltd v Somerville* [1996] 3 NZLR 424 (HC)

*P v D* [2000] 2 NZLR 591

*Queenstown Lakes District Council v Palmer* [1888] 1 NZLR 548 (CA)

*Ransfield v Radio Network Limited* [2005] 1 NZLR 233, (2004) 8 HRNZ 185 (HC)

*Re Chase* [1888] 1 NZLR 325 (CA)

*Shortland v Northland Health Ltd* [1998] 1 NZLR 433 (CA)

*Smith, Valerie v Air New Zealand Ltd* [2011] 2 NZLR 171

*Stephenson v Waite Tileman Ltd* [1973] 1 NZLR 152

*Sulco Ltd v ES Redit & Co Ltd* [1959] NZLR 45 (CA)

*T v Attorney-General* (1988) 4 FRNZ 582 (HC)

*Taupo Borough Council v Birnie* [1978] 2 NZLR 397

*Taylor v Beere* [1982] 1 NZLR 81 (CA)

*Te Pou Matakana Ltd and Whanau Tahi Ltd v Attorney-General* [2021] NZHC 3319

*Television New Zealand Ltd v Rogers* [2007] NZSC 91, [2008] 2 NZLR 277

## Australia

*Purvis v New South Wales* (2003) 217 CLR 92 (HCA)

*Rogers v Whitaker* (1992) 175 CLR 479

*Rosenberg v Percival* (2001) 205 CLR 434; [2001] HCA 18

*Secretary, Department of Health and Community Services v JWB and SMB [Marions Case]* (1992) 175 CLR 218 (HCA) 250

## The United Kingdom

*Airedale NHS Trust v Bland* [1993] 1 All ER 821, 860

*Al Kandari v J R Brown & Co.* [1988] QB 665

*Al Kandari v J R Brown & Co.* [1988] QB 665

*Blyth v Birmingham Waterworks* (1856) 11 Ex 781

*Bolam v Friern Hospital Management Committee* [1957] 1 WLR 582

*Bolitho v City and Hackney Health Authority* [1998] AC 232 (HL)

*Bolton v Stone* [1951] AC 850

*Bonnington Castings Ltd v Wardlaw* [1956] AC 613

*Cambridge Water Co. v Eastern Counties Leather plc* [1994] 1 All ER 53

*Cambridge Water Co. v Eastern Counties Leather plc* [1994] 2 AC 264

*Campbell v MGN Ltd* [2004] UKHL 22, [2004] 2 AC 457

*Canterbury v Spence* 464 F 2d 772, 780 (1972)

*Corr v IBC Vehicles Ltd* [2008] 1 AC 844

### Canada



### The United States of America



## II. Legislation

### New Zealand

Care of Children Act 2004

Consumer Guarantees Act 1993

Crimes Act 1962

Crown Entities Act 2004

Fair Trading Act 1984

Health Act 1956

Health and Disability Commissioner Act 1994

Health and Disability Services Act 2001

Human Rights Act 1993

Information Act 1982

New Zealand Bill of Rights Act 1990

New Zealand Public Health and Disability Act 2000

Official Information Act 1982

Privacy Act 1993

Public Records Act 2005

Sale of Goods Act 1908

Sentencing Act 2002

Social Security Act 1964

The Health Practitioners Competence Assurance Act 2003

The Treaty of Waitangi (Te Tiriti o Waitangi) 1948

Workers' Compensation Act 1956

Workers Compensation for Accidents Act 1900

## Other jurisdictions

The Health Insurance Portability and Accountability Act of 1996 (HIPAA) (USA)

The General Data Protection Regulation 2016 (GRPR) (EU)

# III. Regulation or Schedules

## New Zealand

Health and Disability Commissioner (Code of Health and Disability Services Consumers' Rights) Regulations 1996

Health Information Privacy Code 2004

# IV. Treaties or Covenants

International Covenant on Civil and Political Rights 1966

Universal Declaration of Human Rights 1948

# V. Decisions of Tribunals or Commissioners

## New Zealand

Health and Disability Commissioner, Decision 08HDC20258 (11 November 2009)

# VI. Books and Chapters

Barry John, The Great Influenza: The Story of the Deadliest Pandemic in History, (Penguin Books, London, 2005)

Brownsword Roger and Goodwin Morag, Law and Technologies of the Twenty-first Century (Cambridge University Press, Cambridge, 2012)

Bryder Linda, A History of "The Unfortunate Experiment" at National Women's Hospital, (Auckland, Auckland University Press, 2009)

Butler Andrew and Butler Petra, New Zealand Bill of Rights Act: A Commentary, 2nd ed (Wellington: Thomson Reuters Press, 2015)

Buyers John, “Artificial Intelligence: The Practical Legal Issues” (Law Brief Publishing: London, 2018).

Coney Sandra, “The Unfortunate Experiment: The Full Story Behind the Inquiry into Cervical Cancer Treatment” (Wellington: Penguin Books, 1988)

Cook Karen S., “Trust in Society” Editor Karen S. Cook, (Russell Sage Foundation, London, 2001)

Crompton Laura, "A Critical Analysis of the Trust Human Agents Have in Computational and Embodied AI" in” in J. Seibt M. Nørskov, O.S Quick (ed) Culturally Sustainable Social Robotics (IOS Press, 2021)

Flood Collen M. and Régis Catherine, *AI and Health Law*, in Florian Martin-Bariteau & Teresa Scassa, eds., “Artificial Intelligence and the Law in Canada” (Toronto: LexisNexis Canada, 2021)
Fried Charles, Contract as Promise: A Theory of Contractual Obligation (Harvard University Press, Cambridge, 1981)

Institute of Medicine (US) Roundtable on Evidence-Based Medicine, ed. Pierre L Yong, “The Healthcare Imperative: Lowering Costs and Improving Outcomes” (Washington DC: National Academies Press, 2010)

Madigan Michael T. and Martinko John, Brock Biology of Microorganisms (11th edition, Prentice Hall, London, 2006)

Michael T. Madigan and others, Brock Biology of Microorganisms (12th ed) (Washington, Pearson Education Heg USA, 2018

Miller Vincent C. and Bostrom Nick, Fundamental Issue of Artificial Intelligence (Synthese Library, Berlin: Springer, 2014)

Moran Mayo, “Rethinking the Reasonable Person: An Egalitarian Reconstruction of the Objective Standard” (Oxford Scholarship Press: London, 2003)

Nie and Anderson, “Bioethics in New Zealand: A Historical and Sociological Review,” Annals of Bioethics, ed. Peppin and Cherry (Lisse, Swetz and Zeitlinger, 2003)

Paul Charlotte, “Medicine in Context”, in Joanna Manning ed, The Cartwright Papers: Essays on the Cervical Cancer Inquiry 1987-1988 (Wellington, Bridget Williams Books, 2009)

Penk Steven, and others, Privacy Law in New Zealand (2nd Edition, Thomas Reuters, Wellington, 2016)

Porritt Arthur BT, "History of Medicine in New Zealand" (Cambridge University Press, Cambridge, 1967)

Russell Stuart and Norvig Peter, Artificial Intelligence: A Modern Approach (3rd Ed) (Boston, Pearson Education Heg USA, 2016)

Silver Nate, "The Signal and the Noise: Why So Many Predictions Fail – But Some Don't" (Penguin Group, New York, 2012)

Simester AP, Brookbanks Warren, and Boister Neil, Principles of Criminal Law (5th ed) (Wellington, Thomson Reuters, 2007)

Skegg David, "Foreword" in Joanna Manning ed, The Cartwright Papers: Essays on the Cervical Cancer Inquiry 1987-1988 (Wellington, Bridget Williams Books, 2009),

Southwick Karen, Silicon Gold Rush: The Next Generation of High tech Stars Rewrites the Rules of Business (California Press, Stanford, 1999)

Todd Stephen, and others, Todd on Torts (8th ed, Thomson Reuters New Zealand, Wellington, 2019)

Todd Stephen, Tort Law in New Zealand (3rd Edition, Kluwer Law International, Netherlands)

van Rysewyk Simon Peter, "Machine Medical Ethics" (Matthijs Pontier (ed), Springer, Intelligent, Systems, Control and Automation: Science and Engineering, (2015

# VII. Journal Articles

Allain Jessica S. "From Jeopardy! To Jaundice: The Medical Liability Implications of Dr. Watson and Other Artificial Intelligence Systems" (2013) Vol.73(4) L.L Rev

Allen Beth, "Information as an Economic Commodity" (1990) Vol 80(2) The American Economic Review

Barocas Solon and Selbst Andrew D. "Big Data's Disparate Impact" (2014) 571:104 California Law Review

Bathaee Yavar, "The Artificial Intelligence Black Box and the Failure of Intent and Causation" (2018) Vol. 32:2 Harvard Journal of Law & Technology

Beck Andrew H and others “Systematic Analysis of Breast Cancer Morphology Uncovers Stromal Features Associated with Survival” (2011) Vol.3(108) Sci Transl. Med,

Beuermann Christine, “Do hospitals owe a so-called “non-delegable” duty of care to their patients?” (2017) Vol.26:1, Medical Law Review 1-26

Brown Rafael Dean, “Property ownership and the legal personhood of artificial intelligence” (2020) Vol.30 Information & Communications Technology Law

Bryan Ava Ferguson, and others, “Unknown Unknowns: Surgical Consent During the COVID-19 Pandemic,” (2020) Vol.272(2) Surgical Perspective

Carroll Stephanie and others, “Data as a Strategic Resource: Self-Determination, Governance, and the Data Challenge for Indigenous Nations in the United States” (2016) Vol 8(2) The International Indigenous Policy Journal

Chalmers Iain, “Defendants of the Cartwright Inquiry are Unable to Provide a Description of ‘Adequate Care’ for Cervical Carsinoma in Situ” (2010) New Zealand Medical Journal No.1322

Chinen MA, “The co-evolution of autonomous machines and legal responsibility” (2016) Virginia Journal of Law & Technology

Coglianese David and Cary Lehr, “Regulating by Robot: Administrative Decision Making in the Machine-Learning Era” (2017) 1147(105) Geo. Law Journal

Daikos G.K, “History of Medicine: Our Hippocratic Heritage,” (2007) International Journal of Antimicrobial Agents 29

Else Anne “The “Unfortunate Experiment” and the Cartwright Inquiry, twenty years on: Why getting it right matters, (2010) Vol.24(2) Women’s Studies Journal

Else Anne, “The “Unfortunate Experiment” and the Cartwright Inquiry, twenty years on: Why getting it right matters”, (2010) Women’s Studies Journal Vol.24(2)

Emanuel A “To whom will ye liken me, and make me equal? Reformulating the Role of the Comparator in the Identification of Discrimination” (2014) 45 VUWLR 1

Fjelland Ragnar, “Why general artificial intelligence will not be realized” 2020 Humanities & Social Sciences Communications

Gardner John, “The many faces of the reasonable person” (2015) 131 Law Quarterly Review

Gaskins Richard, Tort Reform in the Welfare State: The New Zealand Accident Compensation Commission, (1980) 18:2 Osgoode Hall Law Journal

Gerke Sara, Minssen Timo and Cohen Glenn, "Ethical and legal challenges of artificial intelligence-driven healthcare" (2020) Artificial Intelligence in Healthcare, 295-336

Gerstner M.E, "Comment: Liability Issues with Artificial Intelligence Software," (1993) 33 Santa Clara Law Review

Gillis Talia, "False Dreams of Algorithmic Fairness: The Case for Credit Pricing" (2019)

Godfrey Brett, "Law Must Keep Pace with Increasingly Sophisticated Technology to Protect Public Health," WL 7247054 (2014)

Gooding Piers and Kariotis Timothy, Ethics and Law in Research on Algorithmic and Data-driven Technology in Mental Health Care: Scoping Review, (2021) JMIR Mental Health

Green Herbert, "Cervical Carcinoma in Situ," (1970) Australian and New Zealand Journal of Obstertrics and Gynaecology Vol10(1)

Green Sarah, "The risk pricing principle: a pragmatic approach to causation and apportionment of damages" (2005) Vol. 4, Law, Probability and Risk, 159-175

Heslop Barbara "All about research – looking back at the 1987 Cervical Cancer Inquiry" (2004) New Zealand Medical Journal, Vol.117

Hoffman Sharona and Podgurski Andy, "Artificial Intelligence and Discrimination in Health Care" (2020) Vol 19(3) Yale Journal of Health Policy, Law, and Ethics

Holme Anne Lise, Berland Astrid Karin and Severinsson Elisabeth, "Older Patients' Involvement in Shared Decision-Making—A Systematic Review," (2016) Open Journal of Nursing, Vol:6(3)

Howard Virginia J. "Reasons underlying racial differences in stroke incidence and mortality" (2013) Stroke Vol44:5

Hulkower Raphael, "The History of the Hippocratic Oath: Outdated, Inauthentic, and Yet Still Relevant," (2010) Einstein Journal of Biology and Medicine 25

Javid Christine, "Predictive Medicine: Genes indicate diseases before symptoms do" (2004)

Johnson Neil, and others, "Abrupt Rise of New Machine Ecology Beyond Human Response Time," (2013) Vol.3 Scientific Reports

Kalra D, Gertz R, Singleton P, Inskip "Confidentiality of personal health information used for research" (2006) HM BMJ. Vol333(7560):196-8

Kersting Kristian, "Machine Learning and Artificial Intelligence: Two Fellow Travelers on the Quest for Intelligent Behaviour Machines" (2018) Vol.1(6) Frontiers in Big Data

King Thomas C. and others "Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions" (2019) Vol. 25 Science and Engineering Ethics,

Kingston J.K.C, "Artificial Intelligence and Legal Liability" (2018)

Kokkoris Michael D and Kamleitner Bernadette, "Would you sacrifice your privacy to protect public health? Prosocial responsibility in a Pandemic paves the way for Digital Surveillance" (2020) Vol 18 Front. Psychology

Kovalenkto Ekaterina and others, "Distinguishing between Parkinson's Disease and essential tremor through video analytics using machine learning: A pilot study" (2020) Vol 21(10) IEEE Sensors Journal,

Kroll Joshua A. and others "Accountable Algorithms" (2015) 165(633) University of Pennsylvania Law Review

LaRosa Emily and Banks David, "Impacts on Trust of Healthcare AI," Carnegie Mellon University, 2018.

Lewicki R.J, Tomlinson E.C, and Gillespie N. "Models of Interpersonal Trust Development: Theoretical Approaches, Empircal Evidence, and Future Directions" (2006) Vol.32 Journal of Management,

Lum Kristian and Isaac William, "To predict and to serve?" (2016) 13(5) Royal Statistical Society: Significance

Matsuzaki Tokio, "Ethical Issues of Artificial Intelligence in Medicine" (2018) Vol.55 (1) California Western Law Review,

McCall Becky, "15 ways Silicon valley is harnessing Big Data for Health" (201) 26 7-10 Nature Medicine

McCarthy John, and others, "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence" (31 August 1955) : http://raysolomonoff. com/dartmouth/boxa/dart564props.pdf [accessed 5 February 2018

McCredie and others “Consequences in Women of Participating in a Study of the Natural History of Cervical Intraepithelial Neoplasia” (2010) Aust NZ J Obstet Gynaecol, Vol. 50(4)

Minami Christina A. and others, “Association between state medical malpractice environment and postoperative outcomes in the United States,” (2016) 224:3 Journal of the American College of Surgeons

Motague E.N.H and others, “Trust in Medical Technology by Patients and Health Care Providers in Obstetric Work Systems,” (2011) Vol.29:5 Behaviour & Information Technology, , 541-554

Nicholson Price II W. “Describing Black-Box Medicine,” (2015) 21 B.U J. Sci & Tech L. 347

Nicholson Price II W., and Cohen Glenn, “Privacy in the Age of Medical Big Data” (2019) 25(1) Nat Med

Obermeyer Zaid and others, Dissecting racial bias in an algorithm used to manage the health of populations, (2019) 366(6464) Science, p447-453

Ohm Paul, “Broken Promises of Privacy: Responding to the Surprising Failure of Anonymisation” (2010) UCLA Law Review

Ortashi Osman, and others, “The practice of defensive medicine among hospital doctors in the United Kingdom” (2013) 14 BMC Medical Ethics

Paul Charlotte and Brookes Barbara, “The Rationalization of Unethical Research: Revisionist Accounts of the Tuskegee Syphilis Study and the New Zealand “Unfortunate Experiment”, (2015) American Journal of Public Health Vol. 105(10)

Paul Charlotte, “Internal and External Morality in Medicine: Lessons from New Zealand” (2000) Biomedical Journal Vol.320(7233)

Phillips Michael, and others, “Detection of Malignant Melanoma Using Artificial Intelligence: An Observational Study of Diagnostic Accuracy” (2020) 10(1) Dermatol Pract Concept

Porter Christine, “Constitutional and Regulatory: De-Identified Data and Third Party Data Mining: The Risk of Re-Identification of Personal Information” (2008) Vol(5) University of Washington Shidler, Journal of Law, Commerce & Technology,

Ramamoorthy A and others “Racial/ethnic differences in drug disposition and response: a review of recently approved drugs” (2015) 97(3) Clinical Pharmacological Therapy

Richards Bernadette and Hutchison Katrina, “Consent to Innovative Treatment: No Need for a New Legal Test” (2016) Vol.23:938 Journal of Law and Medicine,

Richter Gesine and others, “Patient views on research use of clinical data without consent: Legal, but also acceptable?” (2019) Vol.27 European Journal of Human Genetics, 841-847

Rocher Luc, Hendrickx Julien M. and de Montyoe Yves-Alexandre, “Estimating the success of re-identifications in incomplete datasets using generative models” (2019) Vol.10 Nature Communications,

Rodrigues Michelle A and Clancy Kathryn B.H, “A comparative examination of research on why women are more underrepresented in some STEMM disciplines compared to others, with a particular focus on computer science, engineering, physics, mathematics, medicine, chemistry, and biology” (2014) University of Illinois Women and Gender Global Perspective, Vol. 2

Rotar-Pavlic Danica, Svab Igor, and Wetzels Raymond, “How do older patients and their GPs evaluate shared decision-making in healthcare?” (2008) BMC Geriatrics Vol.8

Scherer Matthew, “Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies” (2016) Vol.29 (2) Harvard Journal of Law & Technology,

Schiff Daniel and Borenstein Jason, “How should clinicians communicate with patients about the roles of artificially intelligent team members? (2019) Vol.21(2) AMA Journal of Ethics, E138-145

Selbst Andrew D., “Negligence and AI’s Human Users” No. 20-01 (2018) Public Law & Legal Theory Research Paper (unpublished)

Simon Gregory E. and others, “Predicting Suicide Attempts and Suicide Deaths Following Outpatient Visits Using Electronic Health Records” (2018) The American Journal of Psychiatry

Simon Herbert, “Rational Choice and the Structure of the Environment” (1956) 63(2) Psychological Review

Singleton Peter and Wadsworth Michael, “Consent for the use of personal medical data for research” (2006) 333(7561), BMJ 255-258

Sonal M and Vyas N, “Defensive Medicine: A Bane to Healthcare,” (2013) Vol3:2 Ann Med Health Sci Res

Stepanov Ivan, “Introducing a property right over data in the EU: the data producer’s right – an evaluation” (2019) Vol 34(1) International Review of Law, Computers & Technology,

Taber Jennifer M., Leyva Bryan, and Peroksie Alexander, “Why Do People Avoid Medical Care? A Qualitative Study Using National Data,” (2005) Vol30:3 J Gen Intern Medical,

W. Nicholson Price II, “Artificial Intelligence in HealthCare: Applications and Legal Implications” (2017) 14 The SciTech Lawyer

Wachter Sandra, Russell Chris and Mittelstadt Brent, “Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI” 41 (2021) Computer Law & Security Review

Young Kate, Fisher Jane, and Kirkman Maggie, “”Do mad people get endo or does endo make you mad?”: Clinicians’ discursive constructions of Medicine and women with endometriosis” (2018) 29(3) Feminism & Psychology

Zhang Xingyu and others, “Racial and Ethnic Disparities in Emergency Department Care and Health Outcomes Among Children in the United States," (2019) Vol.10 Frontier Paediatrics

# VIII. Reports and Government Releases

## Government

Ad Hoc Expert Group (AHEG) for the Preparation of a Draft text of a Recommendation the Ethics of Artificial Intelligence “First draft of the Recommendation on the Ethics of Artificial Intelligence” (SHS/BIO/AHEG-AI/2020/4 REV.2)

Australian Trade and Investment Commission, “Insight – Artificial Intelligence to drive Australia’s economic growth”, (2018) available at < https://www.austrade.gov.au/news/insights/insight-artificial-intelligence-to-drive-australia-s-economic-growth>

Cartwright Silvia Rose, The Report of the Cervical Cancer Inquiry: The Report of the Committee Inquiry into Allegations Concerning the Treatment of Cervical Cancer at National Women’s Hospital into Other Related Matters, (Wellington, Government Printer, 1988)

Department for Business, Energy and Industrial Strategy, Industrial Strategy, Building a Britain fit for the future (November 2017)

Department of the Prime Minister and Cabinet, “The new health system” (2021), available at < https://dpmc.govt.nz/our-business-units/transition-unit/response-health-and-disability-system-review/information>

Digital Government, New Zealand’s Digital Transformation, Online (https://www.digital.govt.nz/digital-government/nz-digital-transformation/) accessed on 15 June 2018

European Commission Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions: Artificial Intelligence for Europe (2018)

Executive Office of the Presidency of Barack Obama, "Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights" (May 2016), available here < https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf>

Health Quality & Safety Commission New Zealand, "First annual update on increasing transparency in New Zealand healthcare" by Health Quality Intelligence (2017) available at <ttps://www.hqsc.govt.nz/our-programmes/health-quality-evaluation/publications-and-resources/publication/2962/>

House of Lords Select Committee on Artificial Intelligence, *AI in the UK: ready, willing and able?* Report of Session 2017-19 (Published 16 April 2018)

Labour (Labour Communications Office), "Building a stronger health and disability system that delivers for all New Zealanders" (April 2021), available at https://dpmc.govt.nz/sites/default/files/2021-04/heallth-reform-white-paper-summary-apr21.pdf

LabourVoices (Labour Communications Office), "A health system that works for everyone" (21 April 2021), available at < https://www.labour.org.nz/news-health-reforms-2021-system_that_works_for_everyone>

Minister for the Digital Economy and Communications Office, "Towards a Digital Strategy for Aotearoa: Discussion Paper" (Government Press, Wellington, 2019)

Ministry of Business, Innovation and Enterprise, "Digital economy research" (Accessed 15 Feb 2021), at https://www.mbie.govt.nz/business-and-employment/economic-development/digital-economy/digital-economy-research/

Ministry of Health "Emerging Health Technology Advice" (2018-) available at https://www.health.govt.nz/our-work/digital-health/vision-health-technology/emerging-health-technology-advice-and-guidance

Ministry of Health "Introductory Guidance to Emerging Health Technology" (2018-) available at https://www.health.govt.nz/our-work/digital-health/vision-health-technology/emerging-health-technology-advice-and-guidance

Ministry of Health, About Primary Care Organisations, Online (https://www.health.govt.nz/our-work/primary-health-care/about-primary-health-organisations) accessed on 26 March 2018

Ministry of Health, Crown Entities and Agencies, Online (https://www.health.govt.nz/new-zealand-health-system/key-health-sector-organisations-and-people/crown-entities-and-agencies) accessed on 17 April 2018

Ministry of Health, Professional and Regulatory Bodies, Online (https://www.health.govt.nz/new-zealand-health-system/key-health-sector-organisations-and-people/professional-and-regulatory-bodies) accessed on 15 April 2018

Ministry of Justice "Restorative Justice: Best Practice in New Zealand" (2011) Wellington, MOJ

New Zealand Health and Disability System Review, "Final Report Pūrongo Whakamutunga" (2020) Wellington, HDSR

NZ Health Partnership, DHB Procurement Strategy, published May 2016, Online http://www.nzhealthpartnerships.co.nz/wp-content/uploads/2016/05/DHB-Procurement-Strategy-Original-Publishing-May-2016-Colour-Cover.pdf

Office of the Ombudsman. "Request for Complications data by named cardiothoracic surgeon and neurosurgeon. Case numbers 402136/402138/402140/402142/402144." (2016) Available at <http://www.ombudsman.parliament.nz/system/paperclip/document_files/document_files/1635/original/402136_etc_-_request_for_surgical_complications_data.pdf?1467187036> (accessed on 05 June, 2020)

Report of the Commission of Inquiry, "Compensation for Personal Injury in New Zealand" ("The Woodhouse Report") (Government Press, Wellington, New Zealand, 1967)

Report of the Royal Commission on Social Policy (Government Press, Wellington, New Zealand 1988)

The Prime Minister's Business Advisory Council The Future of Work report (2019) available at < https://bac-staging.beingbui.lt/the-future-of-work-report>

Waitangi Tribunal WAI 2275 (2015)

United Nations, Records of the General Conference, 40th session, Paris, 12 November-27 November 2019, volume 1: Resolutions (40 C/RESOLUTIONS VOL.1 + CORR)

US National Science and Technology Council, "Preparing for the future of artificial intelligence" (2016) available at <https://publicintelligence.net/whitehouse-preparing-artificial-intelligence>

CMS Law AI in Life Sciences: Legal perspectives on the opportunities and challenges of AI for life sciences companies (2019)

Holder Child, Khurana Vikram and Watts Mark, "Artificial Intelligence: Public Perception, Attitude and Trust" (Bristows Law Reform, 2018)

Larson Jeff, and others, "How we analyzed the COMPAS Recidivism Algorithm" (ProPublica, May 23, 2016). Available here < https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>

## Public (NGOs, charities, or research entities)

AI Forum and the Ministry of Business, Innovation and Enterprise (MBIE) *Artificial Intelligence: Shaping a Future for New Zealand* (2018), available here <https://www.mbie.govt.nz/dmsdocument/5754-artificial-intelligence-shaping-a-future-new-zealand-pdf>

AI Forum, AI Forum and Precision Driven Health, Artificial Intelligence for Health in New Zealand, (2020), available at https://aiforum.org.nz/wp-content/uploads/2019/10/AI-For-Health-in-New-Zealand.pdf

AI Forum, Towards Our Intelligence Future: An AI Roadmap for New Zealand (2020), available at https://aiforum.org.nz/wp-content/uploads/2019/09/Towards-our-Intelligent-Future_v1.01.pdf, at 109

Gavaghan Colin, and others, "Government Use of Artificial Intelligence in New Zealand" (New Zealand Law Foundation, Otago University, 2019)

Media Release from eHealthNews.NZ, New Zealand Doctor (2018) Accessed at: <https://www.nzdoctor.co.nz/article/undoctored/ai-health-raises-privacy-concerns>

Medical Council of New Zealand, "Good Medical Practice" (2021) Available here < https://www.mcnz.org.nz/assets/standards/b3ad8bfba4/Good-Medical-Practice.pdf>

State of Open Data, "Indigenous Data Sovereignty" Available here < https://www.stateofopendata.od4d.net/chapters/issues/indigenous-data.html>

Te Mana Raraunga, The Māori Data Sovereignty Network, "Principles of Māori Data Sovereignty: Brief #1" (October 2018) Available here <https://cdn.auckland.ac.nz/assets/psych/about/our-

research/documents/TMR%2BM%C4%81ori%2BData%2BSovereignty%2BPrinciples%2BOct%2B2018.pdf?

The Privacy Foundation, "The Ignorance of Anonymisation to Protect Privacy" (2015) at https://www.privacyfoundation.nz/wp-content/uploads/2020/09/The-Ignorance-of-Anonymisation-to-Protect-Privacy.pdf

University College London, "Anonymisation". Available here: https://www.ucl.ac.uk/data-protection/guidance-staff-students-and-researchers/practical-data-protection-guidance-notices/anonymisation-and

## IX. Speeches or Conference Presentations

Boden Margaret A. "Perspective: AI – Utopia or Dystopia?" (World Economic Forum) 2019

Fiser Harvey, "Automating Medicine" (Neuroscience and Society Conference, North Sydney Harborview Hotel, Sydney, August 2018) under heading "What are the legal consequences?"

Fjelland Ragnar, "Why general artificial intelligence will not be realized" 2020 Humanities & Social Sciences Communications

Hack Aoteraroa, NZ's 1st Artificial Intelligence in Healthcare Conference (Auckland: January 2019)

McCarthy John, and others, "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence" (31 August 1955), at 1: http://raysolomonoff. com/dartmouth/boxa/dart564props.pdf

O'Neill O. "Some limits on informed consent" (2003) Symposium on consent and confidentiality (Newnham College, Cambridge, UK)

## X. Media

BBC, Singapore Personal Data Hack hits 1.5m Health Authorities Says, (20 July 2018) online (https://www.bbc.com/news/world-asia-44900507)

Browne Ryan, “Elon Musk warns A.I. could create an ‘immortal dictator from which we can never escape’” (2018) CNBC, available at https://www.cnbc.com/2018/04/06/elon-musk-warns-ai-could-create-immortal-dictator-in-documentary.html>
Confessore Nicholas, “Cambridge Analytica and Facebook: The Scandal and the Fallout so far” (8 April 2018) New York Times
Dastin Jeffrey, “Amazon scraps secret AI recruiting tool that showed bias against women” (Reuters, 11 Oct 2018), Reuters International, available here < https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>

Hocquet Philippe, Trust and Don’t Verify: The AI Black Box Problem, avail. on medium, accessed on 15 August 2018 at < https://medium.com/@PhilippeHocquet/trust-and-dont-verify-the-ai-black-blox-problem-442c2b15e79e>

Jackson Gabrielle, “The female problem: how male bias in medical trials ruined women's health” *The Guardian* (International Edition, 17 Nov 2019)

Kerry Cameron F., and Morris John B. Jr. “Why data ownership is the wrong approach to protecting privacy” (June 26 2019) Brookings, online https://www.brookings.edu/blog/techtank/2019/06/26/why-data-ownership-is-the-wrong-approach-to-protecting-privacy/

Lo Clifford, “After Singapore medical data hack, Hong Kong’s Department of Health becomes latest cyberattack victim”, (02 August 2018) South China Morning Post, online (https://www.scmp.com/news/hong-kong/hong-kong-law-and-crime/article/2158023/after-singapore-medical-data-hack-hong-kongs) accessed on 18 October 2018

Mahoney Luke, in “The Rise of Companions Robots in Japan,” (2019) Japan Today, available here < https://japantoday.com/category/tech/the-rise-of-companion-robots-in-japan>

Nicolas Sartor, “Data Anonymisation Software – Differences Between Static and Interactive Anonymisation” (2019), available at https://www.datasciencecentral.com/profiles/blogs/data-anonymisation-software-differences-between-static-and

Noone Greg, “How New Zealand’s Māori people are fighting for their data sovereignty” (2021) available at < https://techmonitor.ai/policy/privacy-and-data-protection/maori-data-sovereignty-new-zealand-indigenous>

Pool Ian, 'Death rates and life expectancy - Effects of colonisation on Māori', Te Ara - the Encyclopedia of New Zealand, (http://www.TeAra.govt.nz/en/death-rates-and-life-expectancy/page-4) accessed 15 August 2018

Rohner Thomas, “Nunavut woman claims mistreatment in Ottawa hospital” (2020) CBC Canada (https://www.cbc.ca/news/canada/north/nunavut-woman-ottawa-hospital-mistreatment-1.5801844)

Sachdeva Sam, “NZ to push for autonomous weapons treaty” (30 Nov 2021), Newsroom, available at <https://www.newsroom.co.nz/nz-to-push-for-autonomous-weapons-treaty>

Taiuru Kairitiana, Māori Data Sovereignty (2019), Available at <https://www.taiuru.maori.nz/maori-data-sovereignty-and-digital-colonisation/#Treaty_of_Waitangi_overview>

Te Ara Encyclopedia, Dental Profession and Services, Te Ara Encyclopedia on New Zealand, edited by A.H. McLintock (http://www.TeAra.govt.nz/en/1966/medical-services/page-6%20on%2028) accessed on 03 March 2018

Sam Corbett-Davies and others, “A computer program used for bail and sentencing decisions was labelled biased against blacks. It’s actually not that clear.” (2015) Washington Post, available at (https://www.washingtonpost.com/news/monkey-cage/wp/2015/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/).

Wiggers Kyle, “’Anonymized’ X-ray datasets can reveal patient identities” (2021), available at https://venturebeat.com/2021/03/16/anonymized-x-ray-datasets-can-reveal-patient-identities/.

# XI. Other

Collins English Dictionary, available here <https://www.collinsdictionary.com/dictionary/english/treatment> Accessed on 02/08/2021.

Lexico UK English Dictionary, available here <https://www.lexico.com/definition/treatment> Accessed on 02/08/2021

# Appendices

## I. The Code of Patient's Rights

Full name: Health and Disability Commissioner (Code of Health and Disability Services Consumers' Rights) Regulations 1996

1. Consumers have rights and providers have duties:

(1) Every consumer has the rights in this Code.

(2) Every provider is subject to the duties in this Code.

(3) Every provider must take action to—

(a) Inform consumers of their rights; and

(b) Enable consumers to exercise their rights.

2. Rights of consumers and duties of providers:

The rights of consumers and the duties of providers under this Code are as follows:

**Right 1**

*Right to be treated with respect*

(1) Every consumer has the right to be treated with respect.

(2) Every consumer has the right to have his or her privacy respected.

(3) Every consumer has the right to be provided with services that take into account the needs, values, and beliefs of different cultural, religious, social, and ethnic groups, including the needs, values, and beliefs of Māori.

**Right 2**

*Right to freedom from discrimination, coercion, harassment, and exploitation*

Every consumer has the right to be free from discrimination, coercion, harassment, and sexual, financial or other exploitation.

**Right 3**

*Right to dignity and independence*

Every consumer has the right to have services provided in a manner that respects the dignity and independence of the individual

**Right 4**

*Right to services of an appropriate standard*

(1) Every consumer has the right to have services provided with reasonable care and skill.

(2) Every consumer has the right to have services provided that comply with legal, professional, ethical, and other relevant standards.

(3) Every consumer has the right to have services provided in a manner consistent with his or her needs.

(4) Every consumer has the right to have services provided in a manner that minimises the potential harm to, and optimises the quality of life of, that consumer.

(5) Every consumer has the right to co-operation among providers to ensure quality and continuity of services.

**Right 5**

*Right to effective communication*

(1) Every consumer has the right to effective communication in a form, language, and manner that enables the consumer to understand the information provided. Where necessary and reasonably practicable, this includes the right to a competent interpreter.

(2) Every consumer has the right to an environment that enables both consumer and provider to communicate openly, honestly, and effectively.

**Right 6**

*Right to be fully informed*

(1) Every consumer has the right to the information that a reasonable consumer, in that consumer's circumstances, would expect to receive, including—

(a) an explanation of his or her condition; and

(b) an explanation of the options available, including an assessment of the expected risks, side effects, benefits, and costs of each option; and

(c) advice of the estimated time within which the services will be provided; and

(d) notification of any proposed participation in teaching or research, including whether the research requires and has received ethical approval; and

(e) any other information required by legal, professional, ethical, and other relevant standards; and

(f) the results of tests; and

(g) the results of procedures.

(2) Before making a choice or giving consent, every consumer has the right to the information that a reasonable consumer, in that consumer's circumstances, needs to make an informed choice or give informed consent.

(3) Every consumer has the right to honest and accurate answers to questions relating to services, including questions about—

(a) the identity and qualifications of the provider; and

(b) the recommendation of the provider; and

(c) how to obtain an opinion from another provider; and

(d) the results of research.

(4) Every consumer has the right to receive, on request, a written summary of information provided.

**Right 7**

*Right to make an informed choice and give informed consent*

(1) Services may be provided to a consumer only if that consumer makes an informed choice and gives informed consent, except where any enactment, or the common law, or any other provision of this Code provides otherwise.

(2) Every consumer must be presumed competent to make an informed choice and give informed consent, unless there are reasonable grounds for believing that the consumer is not competent.

(3) Where a consumer has diminished competence, that consumer retains the right to make informed choices and give informed consent, to the extent appropriate to his or her level of competence.

(4) Where a consumer is not competent to make an informed choice and give informed consent, and no person entitled to consent on behalf of the consumer is available, the provider may provide services where—

(a) it is in the best interests of the consumer; and

(b) reasonable steps have been taken to ascertain the views of the consumer; and

(c) either,—

(i) if the consumer's views have been ascertained, and having regard to those views, the provider believes, on reasonable grounds, that the provision of the services is consistent with the informed choice the consumer would make if he or she were competent; or

(ii) if the consumer's views have not been ascertained, the provider takes into account the views of other suitable persons who are interested in the welfare of the consumer and available to advise the provider.

(5) Every consumer may use an advance directive in accordance with the common law.

(6) Where informed consent to a health care procedure is required, it must be in writing if—

(a) the consumer is to participate in any research; or

(b) the procedure is experimental; or

(c) the consumer will be under general anaesthetic; or

(d) there is a significant risk of adverse effects on the consumer.

(7) Every consumer has the right to refuse services and to withdraw consent to services.

(8) Every consumer has the right to express a preference as to who will provide services and have that preference met where practicable.

(9) Every consumer has the right to make a decision about the return or disposal of any body parts or bodily substances removed or obtained in the course of a health care procedure.

(10) No body part or bodily substance removed or obtained in the course of a health care procedure may be stored, preserved, or used otherwise than

(a) with the informed consent of the consumer; or

(b) for the purposes of research that has received the approval of an ethics committee; or

(c) for the purposes of 1 or more of the following activities, being activities that are each undertaken to assure or improve the quality of services:

(i) a professionally recognised quality assurance programme:

(ii) an external audit of services:

(iii) an external evaluation of services.

**Right 8**

*Right to support*

Every consumer has the right to have one or more support persons of his or her choice present, except where safety may be compromised or another consumer's rights may be unreasonably infringed.

**Right 9**

*Rights in respect of teaching or research*

The rights in this Code extend to those occasions when a consumer is participating in, or it is proposed that a consumer participate in, teaching or research.

**Right 10**

*Right to complain*

(1) Every consumer has the right to complain about a provider in any form appropriate to the consumer.

(2) Every consumer may make a complaint to—

(a) the individual or individuals who provided the services complained of; and

(b) any person authorised to receive complaints about that provider; and

(c) any other appropriate person, including—

(i) an independent advocate provided under the Health and Disability Commissioner Act 1994; and

(ii) the Health and Disability Commissioner.

(3) Every provider must facilitate the fair, simple, speedy, and efficient resolution of complaints.

(4) Every provider must inform a consumer about progress on the consumer's complaint at intervals of not more than 1 month.

(5) Every provider must comply with all the other relevant rights in this Code when dealing with complaints.

(6) Every provider, unless an employee of a provider, must have a complaints procedure that ensures that—

(a) the complaint is acknowledged in writing within 5 working days of receipt, unless it has been resolved to the satisfaction of the consumer within that period; and

(b) the consumer is informed of any relevant internal and external complaints procedures, including the availability of—

(i) independent advocates provided under the Health and Disability Commissioner Act 1994; and

(ii) the Health and Disability Commissioner; and

(c) the consumer's complaint and the actions of the provider regarding that complaint are documented; and

(d) the consumer receives all information held by the provider that is or may be relevant to the complaint.

(7) Within 10 working days of giving written acknowledgement of a complaint, the provider must,—

(a) decide whether the provider—

(i) accepts that the complaint is justified; or

(ii) does not accept that the complaint is justified; or

(b) if it decides that more time is needed to investigate the complaint,—

(i) determine how much additional time is needed; and

(ii) if that additional time is more than 20 working days, inform the consumer of that determination and of the reasons for it.

(8) As soon as practicable after a provider decides whether or not it accepts that a complaint is justified, the provider must inform the consumer of—

(a) the reasons for the decision; and

(b) any actions the provider proposes to take; and

(c) any appeal procedure the provider has in place.

## 3. Provider compliance

(1) A provider is not in breach of this Code if the provider has taken reasonable actions in the circumstances to give effect to the rights, and comply with the duties, in this Code.

(2) The onus is on the provider to prove it took reasonable actions.

(3) For the purposes of this clause, the circumstances means all the relevant circumstances, including the consumer's clinical circumstances and the provider's resource constraints.

## 4. Definitions

In this Code, unless the context otherwise requires,—

**Advance directive** means a written or oral directive—

(a) by which a consumer makes a choice about a possible future health care procedure; and

(b) that is intended to be effective only when he or she is not competent:

**Choice** means a decision—

(a) to receive services:

(b) to refuse services:

(c) to withdraw consent to services:

**Consumer** means a health consumer or a disability services consumer; and, for the purposes of rights 5, 6, 7(1), 7(7) to 7(10), and 10, includes a person entitled to give consent on behalf of that consumer.

**Discrimination** means discrimination that is unlawful by virtue of Part II of the Human Rights Act 1993.

**Duties** includes duties and obligations corresponding to the rights in this Code.

**Ethics committee** means an ethics committee—

(a) established by, or appointed under, an enactment; or

(b) approved by the Director-General of Health.

**Exploitation** includes any abuse of a position of trust, breach of a fiduciary duty, or exercise of undue influence.

**Optimise the quality of life** means to take a holistic view of the needs of the consumer in order to achieve the best possible outcome in the circumstances.

**Privacy** means all matters of privacy in respect of a consumer, other than matters of privacy that may be the subject of a complaint under Part 5 of the Privacy Act 2020 or matters to which subpart 4 of Part 7 of that Act relates

**Provider** means a health care provider or disability services provider.

**Research** means health research or disability research.

**Rights** includes rights corresponding to the duties in this Code.

**Services** means health services, or disability services, or both; and includes health care procedures.

**Teaching** includes training of providers.

## 5. Other enactments

Nothing in this Code shall require a provider to act in breach of any duty or obligation imposed by any enactment or prevents a provider doing an act authorised by any enactment.

## 6. Other rights

An existing right is not overridden or restricted simply because the right is not included in this Code or is included only in part.

# II. Empirical Research Components

A. Human ethics approval for empirical research (test research conducted on LAWS383 Students prior to main thesis survey)

HUMAN ETHICS COMMITTEE

Secretary, Rebecca
Robinson Telephone: +64 03
369 4588, Extn 94588Email:
human-
ethics@canterbury.ac.nz

UNIVERSITY OF CANTERBURY
*Te Whare Wānanga o Waitaha*
CHRISTCHURCH NEW ZEALAND

Ref: 2019/08/ERHEC-LR

23 July 2019

Christopher
Boniface
School of
Law
UNIVERSITY OF CANTERBURY

Dear Christopher

Thank you for submitting your low risk application to the Educational Research HumanEthics Committee for your research proposal titled "Understanding Trust in Artificial Intelligence and Health Care".

I am pleased to advise that this application has been reviewed and I confirm support of theSchool's approval for this project.

With best wishes for your project.

Yours sincerely

*pp* R. Robinson

Dr Patrick Shepherd

***Chair***

***Educational Research Human Ethics Committee***

*Please note that ethical approval relates only to the ethical elements of the relationship between the researcher, research participants and other stakeholders. The granting of approval by the Educational Research Human Ethics Committee shouldnot be interpreted as comment on the methodology, legality, value or any other matters relating to this research*

B. Human ethics approval for empirical research (main thesis survey)

HUMAN ETHICS COMMITTEE

Secretary, Rebecca
Robinson Telephone: +64 03
369 4588, Extn 94588Email:
human-
ethics@canterbury.ac.nz

UNIVERSITY OF CANTERBURY
*Te Whare Wānanga o Waitaha*
CHRISTCHURCH NEW ZEALAND

Ref: 2020/01/ERHEC-LR

30 April 2020

Christopher
Boniface
School of
Law
UNIVERSITY OF CANTERBURY

Dear Christopher

Thank you for submitting your low risk application to the Educational Research HumanEthics Committee for your research proposal titled "Understanding Trust in Artificial Intelligence and Health Care".

I am pleased to advise that this application has been reviewed and I confirm support of theSchool's approval for this project.

With best wishes for your project.

Yours sincerely

*pp* R. Robinson

Dr Patrick Shepherd

***Chair***

***Educational Research Human Ethics Committee***

*Please note that ethical approval relates only to the ethical elements of the relationship between the researcher, research participants and other stakeholders. The granting of approval by the Educational Research Human Ethics Committee shouldnot be interpreted as comment on the methodology, legality, value or any other matters relating to this research.*

F E S

C. Cover sheet given to survey participants on Qualtrics:

Understanding Trust in Artificial Intelligence and Healthcare
**Information**

My name is Chris Boniface and I am a PhD candidate at the University of Canterbury, School Of Law. My research is into the impact on the New Zealand healthcare system by emergent artificial intelligence (AI) technologies. This survey aims to identify the participant's perceptions and opinions on artificial intelligence within medicine. To do this, the survey will ask you a series of questions about how you would react to the use of an AI system in your healthcare treatment, and your preferences for treatment.

If you choose to take part in this study, your involvement in this project will be answering questions in an online survey. These questions are designed for short answers, 2-5 sentences in length, and should take no more than 25-30 minutes.

In the performance of this involvement, there are no obvious risks to you. However, should you feel uncomfortable with any of the questions, you can choose not to answer specific questions or you can exit the survey at any time.

Participation is voluntary, however once you begin to enter answers into the online survey tool, the anonymized settings of the tool means I will be unable to remove your answers at a later point.

The results of the project may be published, but you may be assured of the complete anonymity and confidentiality of data gathered in this investigation. You will not be asked for your name, nor will I have any way of tracing who has supplied specific answers to questions. Only myself, and my supervisor Dr. Debra Wilson, will have access to the information given in answers. This information will be stored in a password protected electronic file for later use in relation to my PhD (and potentially for journal articles or conference presentations), however as has been explained, no identifying information about participants will exist. A thesis is a public document and will be available through the UC Library. As per university policy, this data will be destroyed 10 years following the submission of my thesis.

Participation in the survey will be taken as informed consent and that you consent to the use of this data for the above listed purposes.

This project has been reviewed and approved by the University of Canterbury Educational Research Human Ethics Committee, and participants should address any complaints to The Chair, Educational Research Human Ethics Committee, University of Canterbury, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz).

For any questions or concerns, I can be contacted at chris.boniface@pg.canterbury.ac.nz. ERHEC Ref: 2020/01/ERHEC-LR

D. Survey given to participants:

These first questions are intended to gauge your understanding and personal interest in the subject of artificial intelligence (AI). This will provide better understanding of your later answers and may help illustrate the approach you took in answering them.

1. **How would you evaluate your personal understanding of artificial intelligence (AI)?**

Strong ☐

Average ☐

Weak ☐

I do not know about AI ☐

2. **Do you have any background or personal interest in the development, use, or possibilities of AI? For example: are you studying a computer science degree or associated engineering degree, or do you have a personal interest (i.e. in science fiction or future technologies in general)?**

Yes – through study ☐

Yes – through personal interest ☐

No ☐

Please briefly detail below:

3. **Which area(s) of daily life do you think AI could provide the most benefit to? Select up to 3.**

Military and Defence ☐

Education ☐

Finance and Banking ☐

Management and Logistics ☐

Medicine ☐

Security ☐

Briefly explain why you have chosen these:

4. **What area(s) of daily life do you think are at the most risk from the implementation of AI?**

Military and Defence ☐

Education ☐

Finance and Banking ☐

Management and Logistics ☐

Medicine ☐

Security ☐

Briefly explain why you have chosen these:

5. **Do you believe you have interacted with an AI in your daily life?**

Definitely yes ☐

Possibly yes ☐

Might or might not ☐

Probably not ☐

Definitely not ☐

☐

**Please provided examples of technologies you have encountered, or experiences you have had, that you believe involved AI:**

The following questions focus on the role and use of AI within the healthcare system. These questions will focus on the hypothetical scenario of you, the participant, as a patient in a healthcare system utilising AI technology.

6. **Rank which of these concepts you consider important in healthcare from most (1) to least (5) important.**

The ability for a doctor to explain their process/decisions

The ability for a doctor to verify their results

Statistically better outcomes

Institutional or systemic efficiency

Accountability and professional discipline

If you wish to detail why you have selected the above order, you can do so below:

7. **Would you prefer your diagnosis and treatment was performed by an AI system or a human doctor? Why?**

AI System ☐

Human Doctor ☐

Depends on the situation/ I am unsure ☐

Prefer not to answer ☐

Briefly explain why…

8. **What information would you consider most important to know when determining whether to allow your treatment to be performed by an AI system?**

9. **What parties do you believe should have access to the medical data collected and analysed by an AI system? Select as many as you like.**

The developer of the system's software (for example, Google)

The developer of the system's hardware components

The hospital where the system is located

The medical professionals relevant to the process and diagnosis (for example, all cancer doctors have access to data relevant to cancer only)

The NZ healthcare system at large

Patients within the data set, or those reliant on it for treatment

The public at large

Private industry

10. **If a doctor made use of a tool that incorporated AI technology, would you like this to be specified to you?**

Yes ☐

No ☐

Depends on the situation/ I am unsure ☐

**11. What medical roles would you feel comfortable with being performed by an AI system? Select all that apply:**

Consultation ☐

Diagnosis ☐

Communication of test results ☐

Treatment Planning ☐

Treatment Provision ☐

Surgery ☐

Other (please specify below) ☐

**12. Are there any situations in which your answers above would change? For example, with a particular illness you would be comfortable with the AI performing different roles than another illness.**

**13. Would you be comfortable with the holder of medical data, collected and used by AI systems, being made available to other parties?**

Yes ☐

No ☐

For specific purposes only ☐

**If you answered "specific purposes only" above, please select which kind of uses you would approve of. Select as many options as you like.**

Medical or health research

Other non-medical research (such as social science research)

Development of AI systems

Training or skills development

Financial gain (for example, selling the data to other corporations)

If there are any other uses you would approve of, please list them below:

**14. Would you expect third party corporations (for example Google) to be subject to the same, or unique, legal responsibilities in respect of your healthcare as the healthcare system?**

| | |
|---|---|
| The same rules | ☐ |
| Different rules | ☐ |
| No rules | ☐ |

**15. Would you be comfortable with diagnostic AI taking on an anticipatory role?**

Would you be comfortable with diagnostic AI taking on an anticipatory role?

This would mean that it makes judgements or identifies connections beyond what you have necessarily gone to the hospital for at this time to identify your needs without consciously consulting you prior.

For example: you go see a doctor for a common cold, and the AI determines your risk for early onset dementia. Would you want to know this information? Would you like this information to be recorded or discarded?

**16. State any right(s) you believe you are entitled to in medicine that might be affected by treatment being done by an AI system, and why? For example: "my right to privacy because…"**

Some common key words often discussed alongside this issue that you may wish to include in your answer:

Discrimination, consent, privacy, bias (racial, cultural, gender), informed.

17. **Discuss below any of the above rights or concepts you believe might be incompatible with the widespread use of AI in healthcare. Please explain whether you think this incompatibility is remediable, or a worthwhile concession.**

For example: some argue that AI systems are incompatible with the concepts of data privacy and racial bias due to the way they utilise information and reach their conclusions. Is this a worthwhile loss when balanced with the benefits of the technology?

18. **Rank the below concepts in order of priority for you. 1 is the most important, and 7 being the least important to you:**

| | |
|---|---|
| The right to privacy | ☐ |
| Superior medical outcomes | ☐ |
| Communication | ☐ |
| Compassion or empathy | ☐ |
| Codes of Conduct (Professional Obligations) | ☒ |
| Patient understanding | ☐ |
| Other (please specify) _____ | ☐ |

19. **Describe briefly how you would prefer an AI system that performs your medical treatment to look?** (images will be provided on the online survey for this component)

- For example: humanoid, realistic, gendered, non-human, does it have a physical body or interact through a screen etc. Some images have been provided to give some guidance.

A: Robotic

B: Humanoid

C: Realistic

D: Machine

Detail your answer below (feel free to provide examples of AI systems or robots you may know of from pop culture, if it makes the description easier):

A: Robotic

B: Humanoid

C: Realistic

D: Machine